

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37

Is artificial intelligence for medical professionals serving the patients?

Protocol for a mixed method systematic review on patient-relevant benefits and harms of algorithmic decision-making.

Christoph Wilhelm^{1,2}, christoph.wilhelm@fgw-brandenburg.de, 1st author (corresponding)
Prof. Dr. Anke Steckelberg², anke.steckelberg@medizin.uni-halle.de, co-author
Dr. Felix G. Rebitschek^{1,3}, rebitschek@uni-potsdam.de, co-author

¹Harding Center for Risk Literacy, Faculty of Health Sciences Brandenburg, University of Potsdam
Virchowstr. 2, 14482 Potsdam, Germany

²Institute of Health and Nursing Science, Martin Luther University Halle-Wittenberg, Magdeburger
Str. 8, 06112 Halle (Saale), Germany

³Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany

Contributions

Christoph Wilhelm conceptualised, developed the methodology and wrote this protocol. He is the guarantor of the Manuscript.

Prof. Dr. Anke Steckelberg validated the methodology of this protocol.

Dr. Felix G. Rebitschek conceptualised, reviewed, edited and supervised this protocol. He provides clinical practice.

38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75

Abstract

Background

Algorithmic decision making (ADM) utilizes algorithms to collect and process data and develop models to make or support decisions. Advances in artificial intelligence (AI) have led to the development of support systems that can be superior to medical professionals without AI support in certain tasks. However, whether patients can benefit from this remains unclear. The aim of this systematic review is to assess the current evidence on patient-relevant benefits and harms when healthcare professionals use ADM systems (developed using or working with AI) compared to healthcare professionals without AI-related ADM (standard care) - regardless of the clinical issues. Furthermore, for interpreting collected evidence and analysing preconditions for the implementation of AI-related ADM in healthcare, experts from research, practice, and regulation will be interviewed.

Methods

Following the PRISMA statement and the MECIR standards for reporting systematic reviews, MEDLINE and PubMed (via PubMed), EMBASE (via Elsevier), IEEE Xplore, CENTRAL will be searched using English free text terms in title/abstract, Medical Subject Headings (MeSH) terms and Embase Subject Headings (Emtree) fields. Additional studies will be identified by contacting authors of included studies and through reference lists of included studies. Grey literature searches will be conducted in Google Scholar. Risk of bias will be assessed by using Cochrane’s RoB 2 for randomised trials and ROBINS-I for non-randomised trials. Transparent reporting of the included studies will be assessed using the CONSORT-AI extension statement. Following the SRQR statement, semi-structured interviews will be conducted and analysed with the help of a qualitative content analysis according to Mayring. Based on the research questions and the findings of the systematic review, the study and interview guide will be developed a priori.

Discussion

It is expected that there will be a substantial shortage of suitable studies that compare healthcare professionals with and without ADM systems concerning patient-relevant endpoints. This can be attributed to the prioritization of technical quality criteria and, in some cases, clinical parameters over patient-relevant endpoints in the development of study designs. Furthermore, it is anticipated that a significant portion of the identified studies will exhibit relatively poor methodological quality and provide only limited generalizable results.

Systematic review registration

This study is registered within Prospero (CRD42023412156).

Keywords: Algorithmic decision making, ADM, artificial intelligence, patient relevant, healthcare professionals, decision support

76 Background

77 Artificial Intelligence (AI) is a broad term referring to the field of computer science that develops
78 algorithms mimicking human cognitive functions such as learning, perception, problem-solving, and
79 decision-making (McCarthy et al. 2006: 5-7). AI encompasses various approaches, including rule-based
80 systems, machine learning (ML), deep learning, and many others. It comprises a range of technologies
81 and techniques, including algorithmic decision-making (ADM) (Graili et al. 2021: 1). ADM refers to the
82 process of using algorithms to gather, process and model input data to inform or make automated
83 decisions. Feedback from these decisions can then be used by the system to improve itself (Araujo et
84 al. 2020: 612). An ADM can take various forms depending on how it is framed and presented to the
85 user or decision subject. It can be a simple algorithm that has been known and used for decades, such
86 as classification trees (von Winterfeldt & Edwards 1986), or a more complex system like a
87 recommender or AI that can provide recommendations to human decision-makers, nudge its users in
88 a certain direction, or perform fully automated decision-making processes without human
89 involvement (Araujo et al. 2020: 613). To sum up, artificial intelligent related algorithmic decision-
90 making systems (AI-related ADM) are decision support systems that either apply AI (e.g., relying on ML
91 models) or have been developed with the help of AI.

92 Recent advances in AI have resulted in the development of more complex and sophisticated systems
93 that can outperform humans in certain tasks. One such example is AlphaZero, a deep learning
94 algorithm developed by DeepMind, which uses reinforcement learning to learn how to play games like
95 chess, shogi, and Go without being explicitly programmed. In fact, AlphaZero has convincingly beaten
96 both human world champions and world champion computer applications in these games, simply by
97 inputting the rules of the game (Silver et al. 2018: 3-4). Another notable AI system is ChatGPT,
98 developed by OpenAI, which is a prototype text-based dialogue system that can generate human-like
99 text and imitate writing essays and business plans (Hughes 2023). It can also analyse and write code in
100 various programming languages, making it a valuable tool for debugging and code improvement
101 (OpenAI 2023). Recently, ChatGPT was evaluated for its clinical reasoning ability by testing its
102 performance on questions from the United States Medical Licensing Examination, where it scored at
103 or near the passing threshold on all three exams without any special training or reinforcement (Kung
104 et al. 2023).

105 These advances in AI seem to have enormous potential to transform many different fields and
106 industries, which begs the question: will AI do so in healthcare?

107 In clinical trials, AI systems have already shown potential to help clinicians make better diagnoses (Bahl
108 et al. 2018, Li et al. 2021), help personalise medicine and monitor patient care (Jiang et al. 2017, Ciervo
109 et al. 2019), and contribute to drug development (Ekins et al. 2019). However, successful application
110 in practice is limited (Panch et al. 2019: 77) and potential issues that may be responsible for this gap
111 between research and practice should be revealed by our work.

112 By searching PubMed for the term "artificial intelligence", we found over 2,000 systematic reviews and
113 meta-analyses published in the last 10 years, with a yearly increasing trend. These include several
114 reviews conducted in the area of AI in healthcare that provide an overview of the current state of AI
115 technologies in specific clinical areas, including AI systems for breast cancer diagnosis in screening
116 programmes (Freeman et al. 2021), ovarian cancer (Xu et al. 2022), early detection of skin cancer (Jones
117 et al. 2022), COVID-19 and other pneumonia (Jia et al. 2022), prediction of preterm birth (Akazawa &
118 Hashimoto 2022), or diabetes management (Kamel et al. 2022). Other reviews have focused on
119 comparing clinicians and AI systems in terms of their performance to show their capabilities in a clinical
120 setting (Shen et al. 2019, Nagendran et al. 2020, Liu et al. 2019).

121 Although these reviews are crucial to the further development of AI systems, they offer little insight
122 into whether patients actually benefit from their use by medical professionals. Indeed, these studies
123 focus on the analytical performance of these systems, rather than on healthcare-related metrics. In
124 most of the studies mentioned here, the underlying algorithms have been evaluated using a variety of
125 parameters, such as the F1 score for error classification, balanced accuracy, false positive rate, and
126 area under the receiver operating characteristic curve (AUROC). However, measures of a system's
127 accuracy often provide non-replicable results (McDermot et al. 2021: 4), do not necessarily indicate
128 clinical efficiency (Keane & Topol 2018: 1), AUROC does not necessarily indicate clinical applicability
129 (Halligan et al. 2015: 935) and in fact, none of these measures reflect beneficial change in patient care
130 (Brocklehurst et al. 2017: 1727, Shah et al. 2019: 1).

131 To summarise, as with any other new technology introduced into healthcare, the clinical effectiveness
132 and safety of AI compared to the standard of care must be evaluated through properly designed
133 studies to ensure patient safety and maximise benefits while minimising any unintended harm (Park
134 et al. 2020: 328). Therefore, **a critical analysis of patient-relevant outcomes is needed, especially the
135 benefits and harms of decisions informed by or made by AI systems.**

136 To this end, this review goes beyond previous studies in several ways. First, we study clinical AI systems
137 that enable algorithmic decision making (AI-related ADM) in general and therefore do not limit
138 ourselves to selected clinical problems. In particular, we focus on machine learning systems that infer
139 rules from observations. Although we omit rule-based systems, we apply the term AI throughout our
140 work because it is often incorrectly and redundantly used for ML and deep learning in the literature
141 we study. Second, we focus on studies that report patient-relevant outcomes that, according to
142 German Institute for Quality and Efficiency in Healthcare (IQWiG 2022: 44) describe, how patients feel,
143 how they can perform their functions and activities or if they survive. These may include, for example,
144 mortality, morbidity (with regard to complaints and complications), length of hospital stay,
145 readmission, time to intervention and health-related quality of life. Third, we focus only on studies
146 that compare medical professionals supported by AI-related ADM systems with medical professionals
147 without AI-related ADM systems (standard care). By doing so, this review provides an overview of the
148 current literature on clinical AI-related ADM systems, summarises the empirical evidence on their

149 benefits and harms for patients and highlights research gaps that need to be addressed in future
150 studies.

151 If clinical AI systems are found to have a well-balanced benefit-to-harm ratio, their implementation
152 into practice can be considered. However, effectively deploying these systems presents a separate
153 frontier. Specific challenges that need to be addressed include, data privacy and security, usability and
154 lack of algorithm transparency (e.g., in the case of proprietary systems), regulations and policies, the
155 changing nature of healthcare work, the risk of harm from system errors, and at the individual level,
156 physician and patient awareness, education, and trust in these systems (Choudhury et al. 2020: 3, He
157 et al. 2019: 33, Yu et al. 2018: 720).

158 To outline these challenges and framework conditions, particularly in the German healthcare system,
159 and to consider the empirical data collected in this review, **it is necessary to consult experts from**
160 **various disciplines to use their experiences, knowledge, and perceptions of the benefits and risks of**
161 **implementing AI-related ADM systems.** Therefore, we will bring together experts from institutions
162 responsible for auditing the safety of AI systems, audit bodies that support implementation processes
163 under scientific and regulatory conditions in the healthcare sector, and bodies that combine practical
164 healthcare work and research with AI systems.

165

166 **Objectives**

167

168 The aim of this review is to systematically assess the current evidence on patient-relevant benefits and
169 harms of ADM systems which are developed or used with AI (AI-related ADM) to support medical
170 professionals compared to medical professionals without this support (standard care) (Study 1).
171 Furthermore, we will elicit expert assessments of our findings and the benefits and risks of the
172 implementation of AI-related ADM in healthcare, which they perceive (Study 2).

173 **Study 1, Systematic review:**

- 174 1. Are there studies that compare patient-relevant effectiveness of AI-related ADM for medical
175 professionals compared to medical professionals without AI-related ADM?
176 2. Do these studies show adequate methodological quality and are their findings generalisable?
177 3. Can AI-related ADM systems help medical professionals to make better decisions in terms of
178 benefits and harms for patients?

179 **Study 2, expert interviews:**

- 180 1. How do experts assess the overall benefit-to-harm ratio of identified AI-related ADM
181 systems?
182 2. In which medical area of competences (diagnosis, treatment, prognosis, prevention) would
183 patients in Western healthcare systems particularly benefit from AI-related ADM systems?
184 3. What are the consequences (e.g., side effects) of implementing AI-related ADM systems in
185 healthcare?

- 186 4. What are the main challenges that hinder the implementation of AI-related ADM systems in
187 clinical practice and what improvements could be suggested?
188 5. What are the regulatory requirements for implementation of AI-related ADM systems in
189 clinical practice?

190

191 **Methods/Design**

192

193 In accordance with the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols
194 (PRISMA-P) statement (Moher et al. 2015), the study protocol for this systematic is registered on the
195 International prospective register of systematic reviews (PROSPERO) database (CRD42023412156). If
196 necessary, post-registration changes to the protocol, will be detailed under the PROSPERO record with
197 an accompanying rationale.

198 To answer the research questions presented, a systematic review in mixed-method design will be
199 conducted, which will be presented into two parts.

200 **Study 1: Systematic review**

201 We will follow the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)
202 statement (Page et al. 2021) and the Methodological Expectations of Cochrane Intervention Reviews
203 (MECIR) standards (Higgins et al. 2021).

204 **Searches**

205 We will search systematically using English free text terms in title/abstract, Medical Subject Headings
206 (MeSH) terms and Embase Subject Headings (Emtree) fields for various forms of keywords related to
207 'artificial intelligence' and relevant subcategories of computer generated and processed decision-
208 making algorithms, 'medical professionals' and keywords describing effectiveness parameters and
209 outcomes as well as preferred study types. Based on the block building approach, keywords and terms
210 are combined using the Boolean operators AND and OR and progressively checked for relevant hits.

211 **Databases to be used for searches**

212 MEDLINE and PubMed (via PubMed), EMBASE (via Elsevier), Institute of Electrical and Electronics
213 Engineers (IEEE) Xplore, will be searched for peer-reviewed articles as well as ClinicalTrials.gov and
214 ICTRP (via CENTRAL) for ongoing trials and protocols.

215 To reduce potential publication bias, additional studies will be identified by contacting authors of
216 included studies, contacting experts in the field, and through reference lists of relevant studies. Grey
217 literature searches will be conducted in Google Scholar. For this purpose, the keywords used in the
218 systematic search will be used in different combinations, as well as their German equivalents. Google
219 Scholar will be searched up to the 10th hit page. The detailed search strategy for each database will
220 be reported under the PROSPERO record once the searches have been conducted.

221

222 **Search strategy**

223 We developed our search strategy using the PICOS scheme (Table 1).

224 **Table 1:** PICOS scheme

Participants	Intervention	Control	Outcome	Study Type
Human patients without restriction in age or sex	Medical professionals supported by an AI-related ADM system applied to a clinical problem	Medical professionals applied to a clinical problem, without support by an AI-related ADM system (standard care)	Patient relevant benefits and harms	Interventional and observational studies

225 While doing preliminary searches for basic literature in Medline and PubMed (via PubMed), we noticed
226 that study conductors from different scientific fields (e.g., computer scientists) used different terms
227 for the intervention outcomes we were looking for. In addition, some studies were not indexed
228 appropriately in PubMed, which complicated our initial search strategy. To carry out the search
229 strategy, we have created and tested the blocks consecutively to gather the best results from each
230 block, expanding and narrowing the search strategy. To assess the right direction of the search
231 strategy, we have used fundamental literature, such as Choudhury & Asan 2020, Park et al. 2020 and
232 Nagendran et al. 2020 as test sets, making sure the results of our search had common ground with
233 these studies.

234 The resulting search string for Medline and PubMed in the individual blocks can be found in table 2
235 and describes the basis for other databases.

236

237

238

239

240

241

242

243

244

245

246

247 **Table 2:** Search string blocks for Medline and PubMed (via PubMed)

<p>Block 1, artificial intelligence:</p>	<p>((("artificial intelligence"[MeSH Terms] OR "artificial intelligence"[Title/Abstract] OR "artificial-intelligence"[Title/Abstract] OR "machine learning"[Title/Abstract] OR "machine-learning"[Title/Abstract] OR "hierarchical learning"[Title/Abstract] OR "computational intelligence"[Title/Abstract] OR "machine intelligence"[Title/Abstract] OR "computer reasoning"[Title/Abstract] OR "deep learning"[Title/Abstract] OR "supervised learning"[Title/Abstract] OR "unsupervised learning"[Title/Abstract] OR "reinforcement learning"[Title/Abstract] OR "representation learning"[Title/Abstract] OR "natural language processing"[Title/Abstract] OR "large language model*" [Title/Abstract] OR "generative model*" [Title/Abstract] OR "representation learning"[Title/Abstract] OR ("knowledge acquisition"[Title/Abstract] AND "computer"[Title/Abstract]) OR ("knowledge representation"[Title/Abstract] AND "computer"[Title/Abstract]) OR "image recognition"[Title/Abstract] OR "machine vision"[Title/Abstract] OR "computer vision"[Title/Abstract] OR "algorithmic decision"[Title/Abstract]))</p>
<p>Block 2, medical professionals:</p>	<p>AND ("expert"[Title/Abstract] OR "experts"[Title/Abstract] OR "medical professional"[Title/Abstract] OR "medical professionals"[Title/Abstract] OR "medical doctor*" [Title/Abstract] OR "physician*" [Title/Abstract] OR "clinician*" [Title/Abstract] OR "general practitioner*" [Title/Abstract] OR "health care professional"[Title/Abstract] OR "health care professionals"[Title/Abstract] OR "healthcare professional"[Title/Abstract] OR "healthcare professionals"[Title/Abstract] OR "nurse"[Title/Abstract] OR "nurses"[Title/Abstract] OR ("therapist"[Title/Abstract] OR "therapists"[Title/Abstract]) OR ("health"[Title/Abstract] AND "alert system"[Title/Abstract]) OR ("medical"[Title/Abstract] AND "alert system"[Title/Abstract]) OR ("practice"[Title/Abstract] AND "alert system"[Title/Abstract]) OR ("hospital"[Title/Abstract] AND "alert system"[Title/Abstract]) OR ("clinic*" [Title/Abstract] AND "alert system"[Title/Abstract]) OR ("health"[Title/Abstract] AND "decision support"[Title/Abstract]) OR ("medical"[Title/Abstract] AND "decision support"[Title/Abstract]) OR ("practice"[Title/Abstract] AND "decision support"[Title/Abstract]) OR ("hospital"[Title/Abstract] AND "decision support"[Title/Abstract]) OR ("clinic*" [Title/Abstract] AND "decision support"[Title/Abstract]) OR ("health"[Title/Abstract] AND "warning system"[Title/Abstract]) OR ("medical"[Title/Abstract] AND "warning system"[Title/Abstract]) OR ("practice"[Title/Abstract] AND "warning system"[Title/Abstract]) OR ("clinic*" [Title/Abstract] AND "warning system"[Title/Abstract]))</p>
<p>Block 3, outcomes:</p>	<p>AND ("effectiveness"[Title/Abstract] OR "effectivity"[Title/Abstract] OR "benefit"[Title/Abstract] OR "benefits"[Title/Abstract] OR "harm"[Title/Abstract] OR "harms"[Title/Abstract] OR "adverse event*" [Title/Abstract] OR "mortality"[Title/Abstract] OR "morbidity"[Title/Abstract] OR "length of hospital stay"[Title/Abstract] OR "readmission"[Title/Abstract] OR "time to intervention"[Title/Abstract] OR "health-related quality of life"[Title/Abstract] OR "endpoint*" [Title/Abstract] OR "outcome*" [Title/Abstract])</p>
<p>Block 4, study types:</p>	<p>AND ("randomised" OR "randomized" OR "RCT" OR "clinical trial*" OR "cohort" OR "observational study" OR "observational design*" OR "case-control" OR "experiment*" OR "retrospective study" OR "retrospective design*" OR "prospective study" OR "prospective design*" OR "non-inferiority" OR "phase* study" OR "intervention study" OR "diagnostic study" OR "pre-post study" OR "pre post study" OR "pre-post design" OR "pre post design")</p>
<p>Filter:</p>	<p>AND (humans[filter]) AND (y_10[filter])</p>

249 **Types of studies to be included**

250 For the systematic search peer reviewed interventional and observational studies published in German
251 or English, 10 years retrospectively from the date of the search will be considered. For the search of
252 grey literature, scientific reports published in German or English 10 years retrospectively from the date
253 of the search will be considered. To extract potentially relevant studies from (systematic) reviews and
254 meta-analyses, secondary studies will be gathered and screened. However, secondary studies will not
255 be included in the synthesis.

256 In contrast to studies of effectiveness and safety, pure efficacy studies (e.g., focusing on algorithms
257 accuracy) will be excluded as these outcomes are not directly relevant for patients. Patient-relevant
258 outcomes will be defined according to the IQWiG method paper (2022). In addition, studies that used
259 AI systems beyond our scopes, such as robotics (systems that support the implementation of decisions)
260 will be excluded. Editorials, commentaries, letters, and other informal publication types will be
261 excluded as well.

262 We will provide a list of all references screened in full text including exclusion reasons in the appendix
263 of the final study.

264 **Participants**

265 Our study is focusing on human patients without restriction of age or sex. Therefore, the input data
266 for the algorithms must include real human data gathered either during routine care and saved for use
267 in research or generated specifically for the individual study.

268 **Intervention**

269 Our study is focusing on medical professionals utilizing an AI-related ADM system to address a clinical
270 problem.

271 In our working definition, a medical professional is a qualified individual who has the authority to
272 perform necessary medical procedures within their professional scope of practice. Their goal is to
273 improve, maintain, or restore the health of individuals by examining, diagnosing, prognosticating,
274 and/or treating clinical problems. This may include medical doctors, registered nurses, and other
275 medical professionals. Clinical problems can encompass illnesses, injuries, and physical or mental
276 disorders, among other conditions.

277 In our working definition, an AI-related ADM system is a clinical decision support system that either
278 apply AI (e.g., relying on ML models) or that has been developed with the help of AI. Clinical decision
279 support models without any involvement of AI will be excluded.

280 **Control**

281 Medical professionals, as described in the working definition, are addressing a clinical problem without
282 the support of an AI-related ADM system (standard care).

283 **Outcomes**

284 Patient-relevant benefits and harms, according to the IQWiG method paper (2022), are gathered.
285 These may include, for example, mortality, morbidity (with regard to complaints and complications),
286 length of hospital stay, readmission, time to intervention and health-related quality of life.

287 **Study types**

288 We will collect both interventional and observational studies, which may encompass randomised
289 controlled trials, cohort studies, case-control studies, randomised surveys, retrospective and
290 prospective studies, phase studies, as well as non-inferiority or diagnostic studies.

291 **Data extraction**

292 Records arising from the literature search will be stored in the citation manager Citavi 6 (c) Swiss
293 Academic Software. After removing duplicates, two reviewers will independently review all titles and
294 abstracts via the browser application Rayyan (Ouzzani et al. 2016). Studies potentially meeting the
295 inclusion criteria will then be screened in full text independently by two reviewers. Disagreements over
296 eligibility of studies will be discussed and, if necessary, resolved by a third reviewer. Authors of the
297 included studies will be contacted if clarification of their data or study methods is required. The
298 PRISMA 2020 flow diagram (Page et al. 2021) will be used to keep the study selection process
299 transparent.

300 Using a standardised data collection form, two reviewers will extract data independently from the
301 included studies and will compare them for discrepancies. Missing data will be requested from study
302 authors. Extracted data will include the following (Table 3).

303

304

305

306

307

308

309

310

311

312

313 **Table 3:** Study data to be extracted

Table/ Item	Example
Study characteristics	
Reference, registration	Meier, 2022
Country of conduction	Germany
Setting	Hospital
Study design	RCT
Observation duration	January 2017 until September 2018
Medical specialty	Intensive care unit (ICU)
Prediction/ classification goal of AI-related ADM	Sepsis
Patient-relevant outcome	Mortality, length of hospital stay
Intervention procedure/ instrument	ICU bedside monitors with recommender
comparison procedure/ instrument	ICU bedside monitors without recommender
Study funding	No funding
Characteristics of the evaluation population	
Patient population	
Inclusion criteria	Participants age over 18 under 64
Exclusion criteria	Pre-existing septic shock
Mean age (SD)	49.8 (1.55)
Population total (share of sex in %)	N=75 (n=30 females)
Medical professional population	
Inclusion criteria	ICU Physician, trained in used AI-related ADM system
Exclusion criteria	Physician at ICU for less than 2 years
Mean age (SD)	45.0 (3.5)
Population total (share of sex in %)	N=6 (n=3 females)
Characteristics of used algorithm	
Algorithm name	ResNet-18
Algorithm architecture	Convolutional neural network (CNN)
Data source	In house digital medical records, monitoring data
Development	Laboratory and health metrics (HR, RR, SpO2, etc.) of n=677 cases (n=220 females)
Validation	Internal: random split sample, external: no
Risk of bias assessment (RoB 2/ ROBINS-I)	
Reporting assessment (CONSORT-AI)	
Study Results	
Supported decision	Initiation of life-saving measures, hospital discharge
Patient benefit (effect size)	Reduction length of hospital stay: 2.3 days Mortality rate reduction: 12/100 patients Reduction length of stay ICU: 0 days
Patient harm (effect size)	Not reported
Other effects	Not reported
Assessment for clinical use	
Implementation status	Not implemented
Author's restrictions on clinical use	System requires more training and testing
Author's recommendation on clinical use	Not mentioned

315 **Risk of bias and quality assessment**

316 Risk of bias will be assessed by using the revised Cochrane risk-of-bias tool for randomized trials (RoB
317 2) (Sterne et al. 2019) and the risk-of-bias in non-randomized studies for interventions (ROBINS-I) tool
318 (Sterne et al. 2017). Disagreements between the authors over the risk of bias in the included studies
319 will be resolved by discussion, or with involvement of a third author if necessary. Transparent reporting
320 of the included studies will be assessed through the Consolidated Standards of Reporting Trials
321 interventions involving Artificial Intelligence (CONSORT-AI) extension by Liu et al. (2020). The
322 CONSORT-AI extension includes 14 new items that were considered sufficiently important for AI
323 interventions to be routinely reported in addition to the core CONSORT items by Schulz et al. (2010).
324 CONSORT-AI aims to improve the transparency and completeness in reporting clinical trials for AI
325 interventions. It will assist to understand, interpret, and critically appraise the quality of clinical trial
326 design and risk of bias in the reported outcomes.

327 **Data synthesis**

328 Given the expected likelihood of heterogeneity between studies in the different medical specialties in
329 terms of outcome measures, study designs and interventions, we do not know if performing a meta-
330 analysis will be possible. However, a systematic narrative synthesis will be provided of the results with
331 an overview of the relevant effects for the outcomes, with information presented in the text and tables
332 to summarise and explain the characteristics and findings of the included studies. The measures will
333 be presented in absolute risks and risk differences. Studies of unclear or high risk of bias will not be
334 excluded, as the authors assume that most potentially relevant studies will be of low methodological
335 quality. This is also reported in recent reviews that examine the methodological quality of machine
336 learning systems in the clinical setting (e.g., Nagendran et al. 2020). Nevertheless, the influence of the
337 potential biases on the results will be discussed.

338 In addition to the synthesis through Study 1, the results of this review will be discussed with experts
339 (Study 2) and presented in the text.

340 **Study 2: Expert interviews**

341 In order to meet the study objectives, and considering the findings of Study 1, we elicit experts'
342 experiences, knowledge, and their perceived benefits and risks of implementing AI-related ADM
343 systems. By doing so, we hope to gain a practice-related understanding of the findings of Study 1 and
344 identify barriers and needs regarding future research questions.

345 Based on the research questions and the findings of Study 1, the study and interview guide will be
346 developed a priori. The study design and the interview guide will be submitted to the ethics committee
347 of the University of Potsdam and pre-registered within the PROSPERO protocol for the systematic
348 review. If necessary, post-registration changes to the protocol will be detailed and justified in the
349 PROSPERO protocol. To ensure transparency of the qualitative research part of our review, we will
350 follow the Standards for Reporting Qualitative Research (SRQR) (O'Brien et al. 2014).

351 **Study design**

352 We will conduct semi-structured interviews. To analyse the interviews, we will use qualitative content
353 analysis according to Mayring (2020). This systematic and rule-based approach will ensure that the
354 analysis is understandable, comprehensible, and verifiable by others.

355 **Data collection**

356 As our interview study aims to understand the practical implications of the findings of Study 1, we will
357 use a purposive sampling strategy to select at least 3 interviewees according to their expertise through
358 their function, profession, practice, and research in relation to the research questions (Mey & Mruck
359 2020: 322, Schreier 2020: 24). Therefore, we will combine experts of the following types: Auditing
360 bodies responsible for auditing AI systems for safety (e.g., TÜV NORD IT Secure Communications GmbH
361 & Co. KG), auditing bodies that support implementation procedures under scientific and regulatory
362 conditions in the healthcare sector (e.g., VDI/VDE Innovation + Technik GmbH), and institutions that
363 combine practical healthcare-related work and research with AI systems (e.g., Berlin Institute of Health
364 at Charité Berlin). The experts will be recruited on the basis of publications, via professional networks,
365 and by sending inquiries to the relevant organisations and institutions. Once relevant experts will have
366 been identified, they will be contacted via e-mail or phone, informing about the specific research
367 questions, and showing some exemplary findings from Study 1, to ask for participation and to set the
368 scope of the interview. There is no reimbursement planned for the experts who participate in the
369 interview.

370 Written informed consent will be obtained prior to each interview and participants will be fully
371 informed in writing and verbally about the purpose, risks, and scope of the study.

372 Approximately one week prior to the interview, participants will be presented with the key findings of
373 Study 1 and the interview guide, which they will be asked about in the interview. In addition to the
374 questions on research interest, descriptive data will be collected on field of activity, specific expertise,
375 organisation, length of involvement in the topic, as well as age and gender of the respective
376 interviewee. Interviews should not exceed 45 minutes. All interviews will be audio recorded and
377 automatically transcribed verbatim using Adobe Premiere Pro 2023 (version 23.1.0). The automated
378 transcripts will be corrected manually by the interviewer.

379 Throughout the data collection process, the interviewer will keep a research diary to record initial
380 impressions immediately after each interview and to encourage reflexivity about how their role in the
381 researcher-participant relationship might influence the data collection process (Mayring 2020: 12).

382 **Data analysis and extraction**

383 The transcripts and notes from the research diary will be transferred to the relevant memos of each
384 transcript in MAXQDA Plus 2022 (version 22.2.0). All data analysis will be carried out by two
385 independent researchers to ensure rigour and trustworthiness. In the event of disagreement between

386 the two researchers on coding, we will revisit the data to develop a shared understanding of the
387 consistent interpretations. The interviews will be analysed using qualitative content analysis according
388 to Mayring (2020). The main categories will initially be developed deductively from the interview guide,
389 which will be developed from the findings of Study 1. Additional major categories will be developed
390 inductively from the data. The final code system, including illustrative interview quotes, will be part of
391 the scientific publication. For this purpose, extracts from the transcript will be translated into English
392 after coding, if necessary.

393 **Data management, data storage and data privacy**

394 After being contacted by the interviewer via phone or e-mail, the experts will receive all relevant study
395 documents via mail or e-mail. Verbal consent to receive the study documents will be obtained by
396 phone or e-mail, and written consent will be obtained in the consent form before the interview takes
397 place. All participants will be assured of confidentiality, anonymity of data, and the right to withdraw
398 from the study or skip questions without negative consequences. Individual contact details will be
399 stored on the interviewer's password-protected computer until the time of the interview. Contact
400 details will be deleted immediately after the interview. The audio recording is used only for quality
401 assurance and will be deleted immediately after it is transcribed and compared. Transcripts will be
402 pseudonymised by changing personal characteristics (such as names and addresses) and personal data,
403 i.e., data that make indirect identification possible (e.g., places or institutions) to pseudonyms.

404

405 **Discussion**

406

407 It is to be expected that there is a significant lack of suitable studies comparing healthcare professionals
408 with and without AI-related ADM systems regarding patient-relevant outcomes. It is assumed that this
409 is due to, first, the lack of approval regulations for AI systems, second, the prioritisation of technical
410 and clinical parameters over patient-relevant outcomes in the development of study designs, and,
411 third, the prioritisation of AI for supporting clinical processes (e.g., administration). In addition, it is to
412 be expected that a large proportion of the studies to be identified are of rather poor methodological
413 quality and provide results that are rather difficult to generalise. Although reporting guidelines such as
414 the Consolidated Standards of Reporting Trials (CONSORT) Statement (Schulz et al. 2010) are well-
415 known and widely used in medical and public health research, they do not necessarily correspond to
416 the novel protocol and study designs that are relevant for the assessment of the research questions
417 relevant here. The extension of the Reporting Guidelines for Clinical Study Reports of Interventions
418 Using Artificial Intelligence (CONSORT-AI) (Liu et al. 2020) may fill the gap but this guideline is relatively
419 new and not necessarily always applied.

420 Engaging in discussions with experts from AI safety audit organizations, healthcare implementation
421 audit bodies, and institutions bridging healthcare practice and AI research, we will gain valuable
422 insights into the implications of study 1's findings (e.g., explanations for limited patient-relevant

423 studies). Additionally, these discussions will illuminate the challenges and opportunities associated
424 with the integration of AI-related ADM systems in healthcare practice and enlighten their benefits and
425 harms for patients.

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451 **List of abbreviations**

452

453	ADM	Algorithmic Decision Making.
454	AI	Artificial Intelligence.
455	AUROC	Area Under the Receiver Operating Characteristic Curve.
456	CENTRAL	Cochrane Central Register of Controlled Trials.
457	CNN	Convolutional Neural Network.
458	CONSORT	Consolidated Standards of Reporting Trials.
459	CONSORT-AI	Consolidated Standards of Reporting Trials for Artificial Intelligence.
460	CRD	Centre for Reviews and Dissemination.
461	EMTREE	Embase subject headings.
462	HR	Heart Rate.
463	ICU	Intensive Care Unit.
464	IEEE	Institute of Electrical and Electronics Engineers
465	IQWiG	German Institute for Quality and Efficiency in Healthcare.
466	MECIR	Methodological Expectations of Cochrane Intervention Reviews.
467	MeSH	Medical Subject Headings.
468	ML	Machine Learning.
469	nRCT	non Randomized Controlled Trial.
470	PICO	Participants, Intervention, Control, Outcome.
471	PRISMA	Preferred Reporting Items for Systematic Review and Meta-Analysis.
472	PRISMA-P	Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols.
473	PROSPERO	International Prospective Register of Systematic Reviews.
474	RCT	Randomized Controlled Trial.
475	ResNet-18	A convolutional neural network that is 18 layers deep.
476	RoB 2	Revised Cochrane Risk-of-Bias Tool for randomized trials.
477	ROBINS-I	Risk-of-Bias in non-randomized studies for interventions.

478	RR	Respiratory Rate.
479	SD	Standard Deviation.
480	SpO2	Oxygen Saturation.
481	SRQR	Standards for Reporting Qualitative Research.
482		
483		
484		
485		
486		
487		
488		
489		
490		
491		
492		
493		
494		
495		
496		
497		
498		
499		
500		
501		
502		
503		
504		
505		
506		
507		
508		

509 **Declarations**

510

511 **Ethics approval and consent to participate**

512 Ethical approval for this study was obtained from the Ethics Committee of the University of Potsdam
513 (103/2023).

514

515 **Consent for publication**

516 Not applicable.

517

518 **Availability of data and materials**

519 Not applicable.

520

521 **Competing interests**

522 The authors declare that they have no competing interests.

523

524 **Funding**

525 This research does not receive any funding.

526

527 **Acknowledgements**

528 Not applicable

529

530

531

532

533

534

535

536

537 References

538

539 Akazawa M, Hashimoto K. Prediction of preterm birth using artificial intelligence: a systematic
540 review. *J Obstet Gynaecol.* 2022 Aug;42(6):1662-1668. doi:
541 10.1080/01443615.2022.2056828.

542 Araujo T, Helberger N, Kruikemeier S, de Vreese CH. In AI we trust? Perceptions about automated
543 decision-making by artificial intelligence. *AI & Soc* 35, 611–623 (2020). doi: 10.1007/s00146-
544 019-00931-w.

545 Bahl M, Barzilay R, Yedidia AB, Locascio NJ, Yu L, Lehman CD. High-Risk Breast Lesions: A Machine
546 Learning Model to Predict Pathologic Upgrade and Reduce Unnecessary Surgical Excision.
547 *Radiology.* 2018 Mar;286(3):810-818. doi: 10.1148/radiol.2017170549.

548 Brocklehurst P, Field D, Greene K, Juszczak E, Keith R, Kenyon S, et al. Computerised interpretation of
549 fetal heart rate during labour (INFANT): a randomised controlled trial. *Lancet.*
550 2017;389:1719–29. doi: 10.1016/s0140-6736(17)30568-8.

551 Choudhury A, Asan O. Role of Artificial Intelligence in Patient Safety Outcomes: Systematic Literature
552 Review. *JMIR Med Inform.* 2020 Jul 24;8(7):e18599. doi: 10.2196/18599.

553 Ciervo J, Shen SC, Stallcup K, Thomas A, Farnum MA, Lobanov VS, Agrafiotis DK. A new risk and issue
554 management system to improve productivity, quality, and compliance in clinical trials. *JAMIA*
555 *Open.* 2019 Mar 19;2(2):216-221. doi: 10.1093/jamiaopen/ooz006.

556 Ekins S, Puhl AC, Zorn KM, Lane TR, Russo DP, Klein JJ, Hickey AJ, Clark AM. Exploiting machine
557 learning for end-to-end drug discovery and development. *Nat Mater.* 2019 May;18(5):435-
558 441. doi: 10.1038/s41563-019-0338-z.

559 EQUATOR. Enhancing the QUALity and Transparency Of health Research. 2021. [http://www.equator-](http://www.equator-network.org)
560 [network.org](http://www.equator-network.org). Accessed 21 November 2023.

561 Freeman K, Geppert J, Stinton C, Todkill D, Johnson S, Clarke A, Taylor-Phillips S. Use of artificial
562 intelligence for image analysis in breast cancer screening programmes: systematic review of
563 test accuracy. *BMJ.* 2021 Sep 1;374:n1872. doi: 10.1136/bmj.n1872.

564 Graili P, Ieraci L, Hosseinkhah N, Argent-Katwala M. Artificial intelligence in outcomes research: a
565 systematic scoping review. *Expert Rev Pharmacoecon Outcomes Res.* 2021 Aug;21(4):601-
566 623. doi: 10.1080/14737167.2021.1886083.

567 Halligan S, Altman DG, Mallett S. Disadvantages of using the area under the receiver operating
568 characteristic curve to assess imaging tests: a discussion and proposal for an alternative
569 approach. *Eur Radiol.* 2015 Apr;25(4):932-9. doi: 10.1007/s00330-014-3487-0.

570 He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence
571 technologies in medicine. *Nat Med.* 2019 Jan;25(1):30-36. doi: 10.1038/s41591-018-0307-0.

572 Higgins J, Lasserson T, Chandler J, Tovey D, Thomas J, Flemyng E, Churchill R. Methodological
573 Expectations of Cochrane Intervention Reviews (MECIR). Standards for the conduct and
574 reporting of new Cochrane Intervention Reviews, reporting of protocols and the planning,
575 conduct and reporting of updates. 2021.
576 <https://community.cochrane.org/sites/default/files/uploads/MECIR-February-2021.pdf>.
577 Accessed 12 November 2023.

578 Hughes A. ChatGPT: Everything you need to know about OpenAI's GPT-3 tool. *BBC Science Focus*
579 *Magazine.* <https://www.sciencefocus.com/future-technology/gpt-3/>. 2023. Accessed 12
580 November 2023.

581 Institute for Quality and Efficiency in Healthcare (IQWiG). *Allgemeine Methoden.* Version 6.1, from
582 24th of January 2022. <https://www.iqwig.de/methoden/allgemeine-methoden-v6-1.pdf>.
583 2022. Accessed 12 November 2023.

- 584 Jia LL, Zhao JX, Pan NN, Shi LY, Zhao LP, Tian JH, Huang G. Artificial intelligence model on chest
585 imaging to diagnose COVID-19 and other pneumonias: A systematic review and meta-
586 analysis. *Eur J Radiol Open*. 2022;9:100438. doi: 10.1016/j.ejro.2022.100438.
- 587 Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y, Dong Q, Shen H, Wang Y. Artificial intelligence in
588 healthcare: past, present and future. *Stroke Vasc Neurol*. 2017 Jun 21;2(4):230-243. doi:
589 10.1136/svn-2017-000101.
- 590 Jones OT, Matin RN, van der Schaar M, Prathivadi Bhayankaram K, Ranmuthu CKI, Islam MS, Behiyat
591 D, Boscott R, Calanzani N, Emery J, Williams HC, Walter FM. Artificial intelligence and
592 machine learning algorithms for early detection of skin cancer in community and primary
593 care settings: a systematic review. *Lancet Digit Health*. 2022 Jun;4(6):e466-e476. doi:
594 10.1016/S2589-7500(22)00023-1.
- 595 Kamel Rahimi A, Canfell OJ, Chan W, Sly B, Pole JD, Sullivan C, Shrapnel S. Machine learning models
596 for diabetes management in acute care using electronic medical records: A systematic
597 review. *Int J Med Inform*. 2022 Apr 2;162:104758. doi: 10.1016/j.ijmedinf.2022.104758.
- 598 Keane PA, Topol EJ. With an eye to AI and autonomous diagnosis. *NPJ Digit Med*. 2018 Aug 28;1:40.
599 doi: 10.1038/s41746-018-0048-y.
- 600 Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, Madriaga M, Aggabao R, Diaz-
601 Candido G, Maningo J, Tseng V. Performance of ChatGPT on USMLE: Potential for AI-assisted
602 medical education using large language models. *PLOS Digit Health*. 2023 Feb
603 9;2(2):e0000198. doi: 10.1371/journal.pdig.0000198.
- 604 Li Q, Zhao K, Bustamante CD, Ma X, Wong WH. Xrare: a machine learning method jointly modeling
605 phenotypes and genetic evidence for rare disease diagnosis. *Genet Med*. 2019
606 Sep;21(9):2126-2134. doi: 10.1038/s41436-019-0439-8.
- 607 Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK; SPIRIT-AI and CONSORT-AI Working Group.
608 Reporting guidelines for clinical trial reports for interventions involving artificial intelligence:
609 the CONSORT-AI extension. *Nat Med*. 2020 Sep;26(9):1364-1374. doi: 10.1038/s41591-020-
610 1034-x.
- 611 Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, Mahendiran T, Moraes G, Shamdas M, Kern C,
612 Ledsam JR, Schmid MK, Balaskas K, Topol EJ, Bachmann LM, Keane PA, Denniston AK. A
613 comparison of deep learning performance against health-care professionals in detecting
614 diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health*.
615 2019 Oct;1(6):e271-e297. doi: 10.1016/S2589-7500(19)30123-2.
- 616 Mayring P. Qualitative Inhaltsanalyse. In: Mey G, Mruck K, editors. *Handbuch Qualitative Forschung*
617 *in der Psychologie*. 2nd ed. Wiesbaden: Springer Fachmedien; 2020. p. 495-511.
- 618 McCarthy J, Minsky M, Rochester N, Shannon C. A proposal for the Dartmouth summer research
619 project on artificial intelligence, August 31, 1955. *AI Magazine* 2006;27(4):12. doi:
620 10.1609/aimag.v27i4.1904.
- 621 Mey G, Mruck K. Qualitative Interviews. In: Mey G, Mruck K, editors. *Handbuch Qualitative Forschung*
622 *in der Psychologie*. 2nd ed. Wiesbaden: Springer Fachmedien; 2020. p. 315-336.
- 623 Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA. Preferred
624 Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015
625 statement. *Syst Rev*. 2015;4(1):1. doi: 10.1609/aimag.v27i4.1904.
- 626 Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, Topol EJ, Ioannidis JPA,
627 Collins GS, Maruthappu M. Artificial intelligence versus clinicians: systematic review of
628 design, reporting standards, and claims of deep learning studies. *BMJ*. 2020 Mar
629 25;368:m689. doi: 10.1136/bmj.m689.
- 630 O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a
631 synthesis of recommendations. *Acad Med*. 2014 Sep;89(9):1245-51. doi:
632 10.1097/ACM.0000000000000388.

- 633 OpenAI. Introducing ChatGPT. 2023. <https://openai.com/blog/chatgpt>. Accessed 12 November 2023.
- 634 Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid, A. Rayyan - a web and mobile app for systematic
635 reviews. *Syst Rev* 5, 210 (2016). doi: 10.1186/s13643-016-0384-4.
- 636 Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl
637 EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW,
638 Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA,
639 Whiting P, Moher D. The PRISMA 2020 statement: an updated guideline for reporting
640 systematic reviews. *BMJ*. 2021 Mar 29;372:n71. doi: 10.1136/bmj.n71.
- 641 Panch T, Mattie H, Celi LA. The "inconvenient truth" about AI in healthcare. *NPJ Digit Med*. 2019 Aug
642 16;2:77. doi: 10.1038/s41746-019-0155-4.
- 643 Park Y, Jackson GP, Foreman MA, Gruen D, Hu J, Das AK. Evaluating artificial intelligence in medicine:
644 phases of clinical research. *JAMIA Open*. 2020 Sep 8;3(3):326-331. doi:
645 10.1093/jamiaopen/ooaa033.
- 646 Schreier M. Fallauswahl. In: Mey G, Mruck K, editors. *Handbuch Qualitative Forschung in der*
647 *Psychologie*. 2nd ed. Wiesbaden: Springer Fachmedien; 2020. p. 19-39.
- 648 Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for
649 reporting parallel group randomised trials. *BMJ*. 2010 Mar 23;340:c332. doi:
650 10.1136/bmj.c332.
- 651 Shah NH, Milstein A, Bagley PhD SC. Making Machine Learning Models Clinically Useful. *JAMA*. 2019
652 Oct 8;322(14):1351-1352. doi: 10.1001/jama.2019.10306.
- 653 Shen J, Zhang CJP, Jiang B, Chen J, Song J, Liu Z, He Z, Wong SY, Fang PH, Ming WK. Artificial
654 Intelligence Versus Clinicians in Disease Diagnosis: Systematic Review. *JMIR Med Inform*.
655 2019 Aug 16;7(3):e10010. doi: 10.2196/10010.
- 656 Silver D, Hubert T, Schrittwieser J, Antonoglou I, Lai M, Guez A, Lanctot M, Sifre L, Kumaran D,
657 Graepel T, Lillicrap T, Simonyan K, Hassabis D. A general reinforcement learning algorithm
658 that masters chess, shogi, and Go through self-play. *Science*. 2018 Dec 7;362(6419):1140-
659 1144. doi: 10.1126/science.aar6404.
- 660 Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, Henry D, Altman DG,
661 Ansari MT, Boutron I, Carpenter JR, Chan AW, Churchill R, Deeks JJ, Hróbjartsson A, Kirkham
662 J, Jüni P, Loke YK, Pigott TD, Ramsay CR, Regidor D, Rothstein HR, Sandhu L, Santaguida PL,
663 Schünemann HJ, Shea B, Shrier I, Tugwell P, Turner L, Valentine JC, Waddington H, Waters E,
664 Wells GA, Whiting PF, Higgins JP. ROBINS-I: a tool for assessing risk of bias in non-randomised
665 studies of interventions. *BMJ*. 2016 Oct 12;355:i4919. doi: 10.1136/bmj.i4919.
- 666 Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, Cates CJ, Cheng HY, Corbett MS,
667 Eldridge SM, Emberson JR, Hernán MA, Hopewell S, Hróbjartsson A, Junqueira DR, Jüni P,
668 Kirkham JJ, Lasserson T, Li T, McAleenan A, Reeves BC, Shepperd S, Shrier I, Stewart LA, Tilling
669 K, White IR, Whiting PF, Higgins JPT. RoB 2: a revised tool for assessing risk of bias in
670 randomised trials. *BMJ*. 2019 Aug 28;366:l4898. doi: 10.1136/bmj.l4898.
- 671 Xu HL, Gong TT, Liu FH, Chen HY, Xiao Q, Hou Y, Huang Y, Sun HZ, Shi Y, Gao S, Lou Y, Chang Q, Zhao
672 YH, Gao QL, Wu QJ. Artificial intelligence performance in image-based ovarian cancer
673 identification: A systematic review and meta-analysis. *EClinicalMedicine*. 2022 Sep
674 17;53:101662. doi: 10.1016/j.eclinm.2022.101662.
- 675 Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng*. 2018 Oct;2(10):719-
676 731. doi: 10.1038/s41551-018-0305-z.