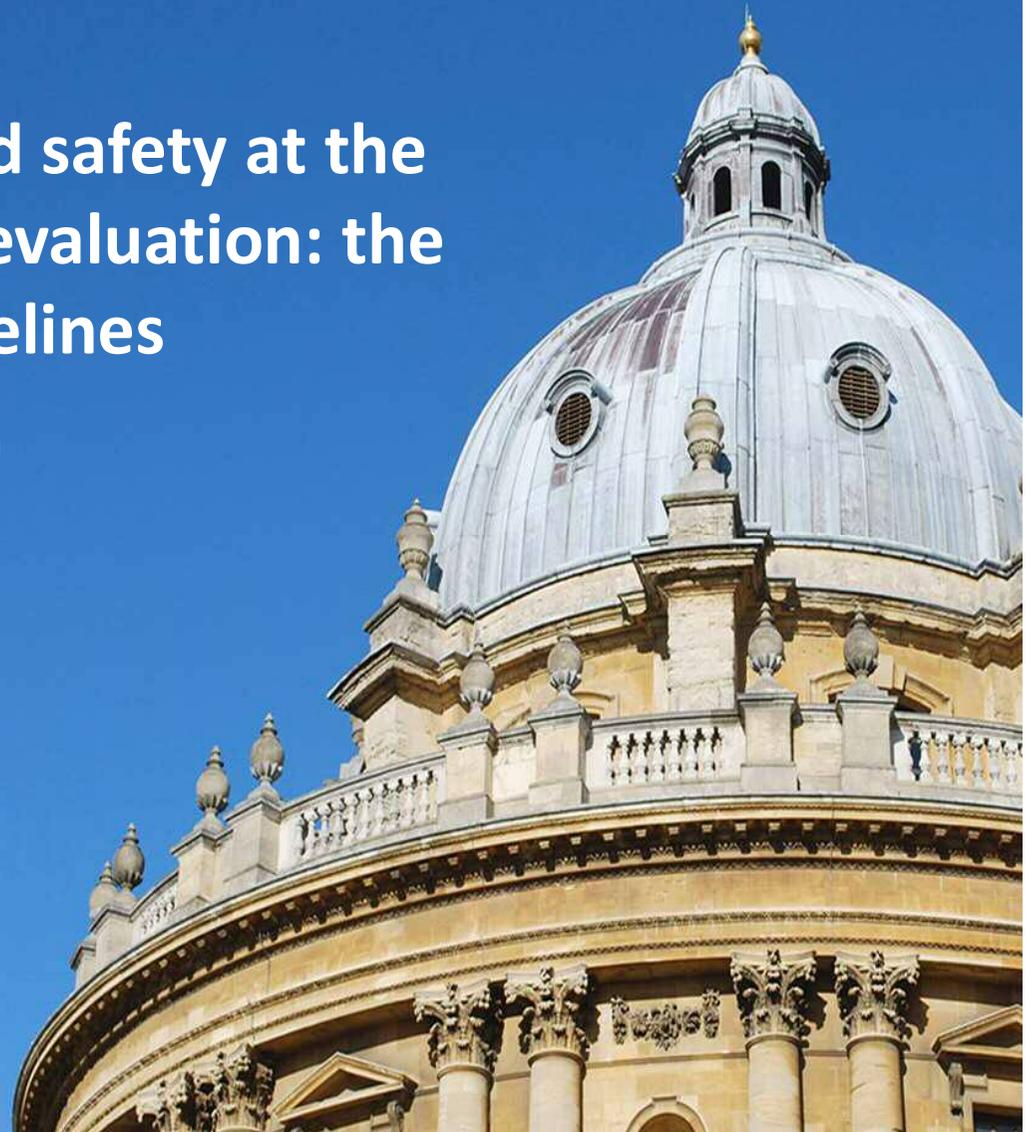


Bringing humans and safety at the centre of clinical AI evaluation: the new DECIDE-AI guidelines

Baptiste Vasey – 04.07.22 – Future Health



How can we assure the safety of AI in healthcare?

Through robust evaluation and
transparent reporting

Need for more guidance and consideration of human users

> [BMJ](#). 2020 Mar 25;368:m689. doi: 10.1136/bmj.m689.

Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies

[Myura Nagendran](#)¹, [Y](#)
[Matthieu Komorowski](#)
[Gary S Collins](#)¹⁰ ¹¹, [M](#)

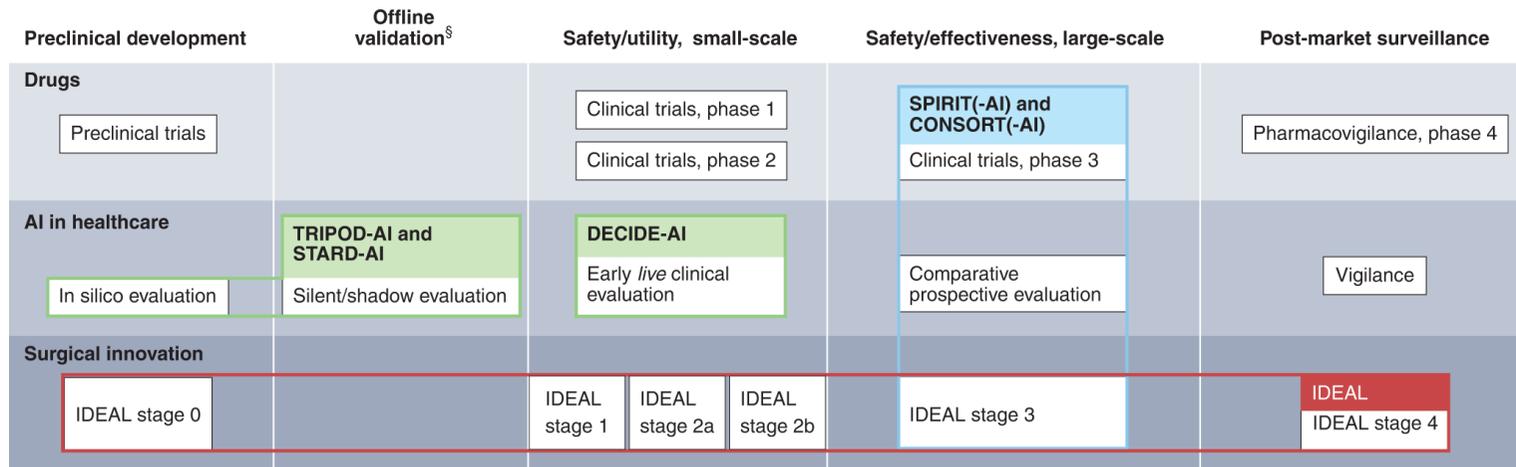
[Comparative Study](#) > [JAMA Netw Open](#). 2021 Mar 1;4(3):e211276.

doi: 10.1001/jamanetworkopen.2021.1276.

Association of Clinician Diagnostic Performance With Machine Learning–Based Decision Support Systems: A Systematic Review

[Baptiste Vasey](#)¹, [Stephan Ursprung](#)², [Benjamin Beddoe](#)³, [Elliott H Taylor](#)¹, [Neale Marlow](#)¹ ⁴,
[Nicole Bilbro](#)⁵, [Peter Watkinson](#)⁶, [Peter McCulloch](#)¹

Healthcare innovation development pathways



Vasey, B., Nagendran, M., Campbell, B. *et al.* Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat Med* **28**, 924–933 (2022). <https://doi.org/10.1038/s41591-022-01772-9>

Importance of early-stage evaluation

Analyse performance of
the AI-human pair

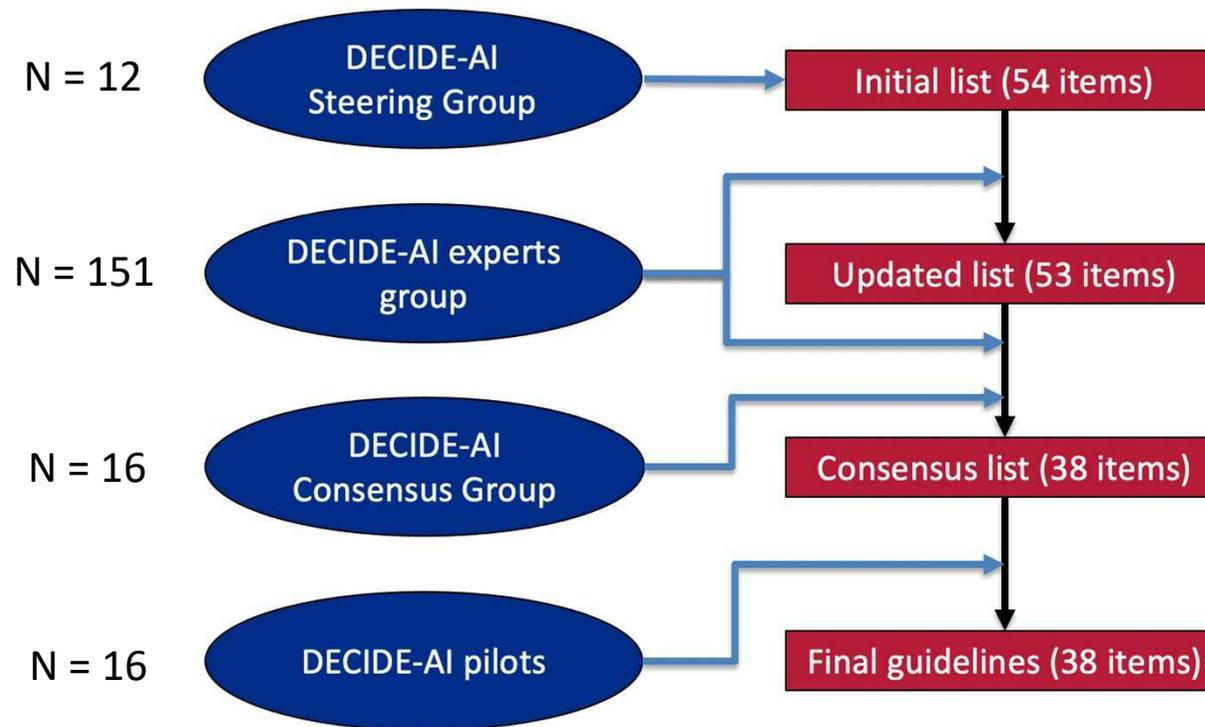
Ensure safety

Evaluate Human Factors
under live conditions

Prepare for large-scale
summative evaluation

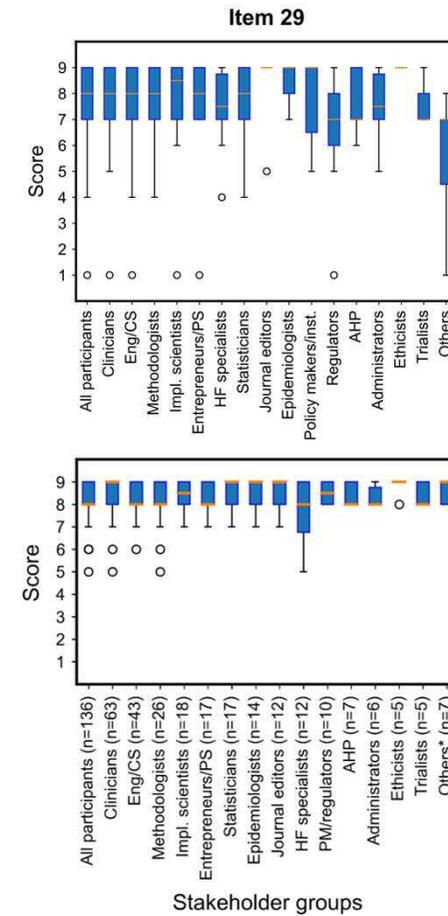
DECIDE  AI

DECIDE-AI – modified Delphi process



A wide range of stakeholder involved

Stakeholder group affiliation			
Stakeholder group	Number of participants	Stakeholder group	Number of participants
Clinicians	68 (45%)	Policy makers/official institutions staff	8 (5%)
Engineers/Computer scientists	50 (33%)	Administrators/hospital management	7 (5%)
Methodologists	30 (20%)	Regulators	7 (5%)
Statisticians	21 (14%)	Trialists	7 (5%)
Implementation scientists	19 (13%)	Ethicists	6 (4%)
Entrepreneurs*	15 (10%)	Private sector representatives*	6 (4%)
Epidemiologists	14 (9%)	Patient representatives	5 (3%)
Human factors specialists*	14 (9%)	Funders	3 (2%)
Journal editors	14 (9%)	Payers/Commissioners	1 (<1%)
Allied health professional	10 (7%)	Psychologists	1 (<1%)



DECIDE AI

Reporting item checklist

Item n°	Theme	Recommendation	Reported on page
1-17	AI-specific reporting items		
I-X	Generic reporting items		
Title and abstract			
1	Title	Identify the study as early clinical evaluation of a decision support system based on AI or machine learning, specifying the problem addressed.	
I	Abstract	Provide a structured summary of the study. Consider including: intended use of the AI system, type of underlying algorithm, study setting, number of patients and users included, primary and secondary outcomes, key safety endpoints, human factors evaluated, main results, conclusions.	
Introduction			
2	Intended use	a) Describe the targeted medical condition(s) and problem(s), including the current standard practice, and the intended patient population(s). b) Describe the intended users of the AI system, its planned integration in the care pathway, and the potential impact, including patient outcomes, it is intended to have.	
II	Objectives	State the study objectives.	
Methods			
III	Research governance	Provide a reference to any study protocol, study registration number, and ethics approval.	
3	Participants	a) Describe how patients were recruited, stating the inclusion and exclusion criteria at both patient and data level, and how the number of recruited patients was decided. b) Describe how users were recruited, stating the inclusion and exclusion criteria, and how the intended number of recruited users was decided. c) Describe steps taken to familiarise the users with the AI system, including any training received prior to the study.	
4	AI system	a) Briefly describe the AI system, specifying its version and type of underlying algorithm used. Describe, or provide a direct reference to, the characteristics of the patient population on which the algorithm was trained and its performance in preclinical development/validation studies. b) Identify the data used as inputs. Describe how the data were acquired, the process needed to enter the input data, the pre-processing applied, and how missing/low-quality data were handled. c) Describe the AI system outputs and how they were presented to the users (an image may be useful).	
5	Implementation	a) Describe the settings in which the AI system was evaluated. b) Describe the clinical workflow/care pathway in which the AI system was evaluated, the timing of its use, and how the final supported decision was reached and by whom.	
IV	Outcomes	Specify the primary and secondary outcomes measured.	
6	Safety and errors	a) Provide a description of how significant errors/malfunctions were defined and identified. b) Describe how any risks to patient safety or instances of harm were identified, analysed, and minimised.	

7	Human factors	Describe the human factors tools, methods or frameworks used, the use cases considered, and the users involved.	
V	Analysis	Describe the statistical methods by which the primary and secondary outcomes were analysed, as well as any prespecified additional analyses, including subgroup analyses and their rationale.	
8	Ethics	Describe whether specific methodologies were utilised to fulfil an ethics-related goal (such as algorithmic fairness) and their rationale.	
VI	Patient Involvement	State how patients were involved in any aspect of: the development of the research question, the study design, and the conduct of the study.	
Results			
9	Participants	a) Describe the baseline characteristics of the patients included in the study, and report on input data missingness. b) Describe the baseline characteristics of the users included in the study.	
10	Implementation	a) Report on the user exposure to the AI system, on the number of instances the AI system was used, and on the users' adherence to the intended implementation. b) Report any significant changes to the clinical workflow or care pathway caused by the AI system.	
VII	Main results	Report on the prespecified outcomes, including outcomes for any comparison group if applicable.	
VIII	Subgroups analysis	Report on the differences in the main outcomes according to the prespecified subgroups.	
11	Modifications	Report any changes made to the AI system or its hardware platform during the study. Report the timing of these modifications, the rationale for each, and any changes in outcomes observed after each of them.	
12	Human-computer agreement	Report on the user agreement with the AI system. Describe any instances of and reasons for user variation from the AI system's recommendations and, if applicable, users changing their mind based on the AI system's recommendations.	
13	Safety and errors	a) List any significant errors/malfunctions related to: AI system recommendations, supporting software/hardware, or users. Include details of: (i) rate of occurrence, (ii) apparent causes, (iii) whether they could be corrected, and (iv) any significant potential impacts on patient care. b) Report on any risks to patient safety or observed instances of harm (including indirect harm) identified during the study.	
14	Human factors	a) Report on the usability evaluation, according to recognised standards or frameworks. b) Report on the user learning curves evaluation.	
Discussion			
15	Support for intended use	Discuss whether the results obtained support the intended use of the AI system in clinical settings.	
16	Safety and errors	Discuss what the results indicate about the safety profile of the AI system. Discuss any observed errors/malfunctions and instances of harm, their implications for patient care, and whether/how they can be mitigated.	
IX	Strengths and limitations	Discuss the strengths and limitations of the study.	
Statements			
17	Data availability	Disclose if and how data and relevant code are available.	
X	Conflicts of interest	Disclose any relevant conflicts of interest, including the source of funding for the study, the role of funders, any other roles played by commercial companies, and personal conflicts of interest for each author.	

Key components

Performance:

- Intended use
- Patient population
- User population
- Implementation settings
- Modifications
- Ethics

Safety:

- Detailed error analysis
- Unexpected risk
- Mitigation strategy

Human Factors:

- Usability (incl. display)
- Learning curves
- Integration in pathways
- Human-computer agreement

Stepping stone:

- Training needs
- Learning curves
- Effect size
- Patient involvement

Methods - items 6a and 6b

- 6a. Provide a description of how significant errors/malfunctions were defined and identified.

- 6b. Describe how any risks to patient safety or instances of harm were identified, analysed, and minimised.

Results – items 13a and 13b

- 13a. List any significant errors/malfunctions related to: AI system recommendations, supporting software/hardware, or users. Include details of: (i) rate of occurrence, (ii) apparent causes, (iii) whether they could be corrected, and (iv) any significant potential impacts on patient care.
- 13b. Report on any risks to patient safety or observed instances of harm (including indirect harm) identified during the study.

Discussion – item 16

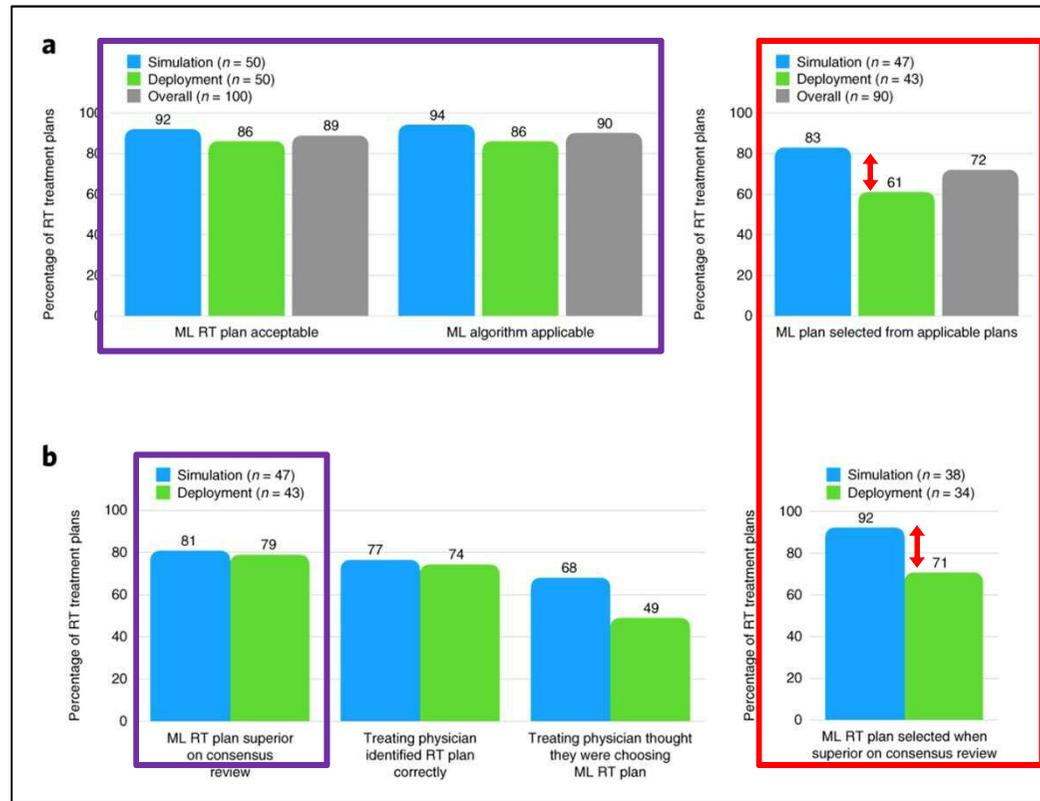
16. Discuss what the results indicate about the safety profile of the AI system. Discuss any observed errors/malfunctions and instances of harm, their implications for patient care, and whether/how they can be mitigated.

Some didn't make it...

- Stand-alone AI system performance
- Interpretability
- Trust curve

... but might join next time.

The trust gap



Equal quality

Trust gap

McIntosh, C., Conroy, L., Tjong, M.C. *et al.* Clinical integration of machine learning for curative-intent radiation treatment of patients with prostate cancer. *Nat Med* **27**, 999–1005 (2021). <https://doi.org/10.1038/s41591-021-01359-w>

How can we assure the safety of AI in healthcare?

By placing humans at the center of AI development and evaluation

