

The Clinical Practice Integration of Artificial Intelligence (CPI-AI) framework

a proposed application of IDEAL principles to artificial intelligence applications in trauma and orthopaedics

From University of Aberdeen,
Aberdeen, UK

L. Farrow,^{1,2} D. Meek,³ G. Leontidis,¹ M. Campbell,¹ E. Harrison,⁴ L. Anderson¹

¹Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, UK

²Grampian Orthopaedics, Aberdeen Royal Infirmary, Aberdeen, UK

³Department of Orthopaedics, Queen Elizabeth University Hospital, Glasgow, UK

⁴Centre of Medical Informatics, University of Edinburgh, Edinburgh, UK

Cite this article:

Bone Joint Res 2024;13(9):
507–512.

DOI: 10.1302/2046-3758.
139.BJR-2024-0135.R1

Correspondence should be
sent to Luke Farrow luke.farrow@abdn.ac.uk

Despite the vast quantities of published artificial intelligence (AI) algorithms that target trauma and orthopaedic applications, very few progress to inform clinical practice. One key reason for this is the lack of a clear pathway from development to deployment. In order to assist with this process, we have developed the Clinical Practice Integration of Artificial Intelligence (CPI-AI) framework – a five-stage approach to the clinical practice adoption of AI in the setting of trauma and orthopaedics, based on the IDEAL principles (<https://www.ideal-collaboration.net/>). Adherence to the framework would provide a robust evidence-based mechanism for developing trust in AI applications, where the underlying algorithms are unlikely to be fully understood by clinical teams.

Article focus

- Safety, reliability, and transparency concerns are some of the major barriers to clinical practice integration of artificial intelligence (AI) in trauma and orthopaedics.
- This study provides a robust evidence framework for new AI applications, following a similar pathway to that for the integration of new surgical devices or drugs.
- Adherence to the pathway should help to provide strong support for the safe and effective future integration of AI into all aspects of trauma and orthopaedics.

Key messages

- This study sets out a five-stage framework which spans from concept outline to post-deployment model surveillance.
- Key checkpoints to progression through each of the stages are highlighted, along with associated reporting guidelines.

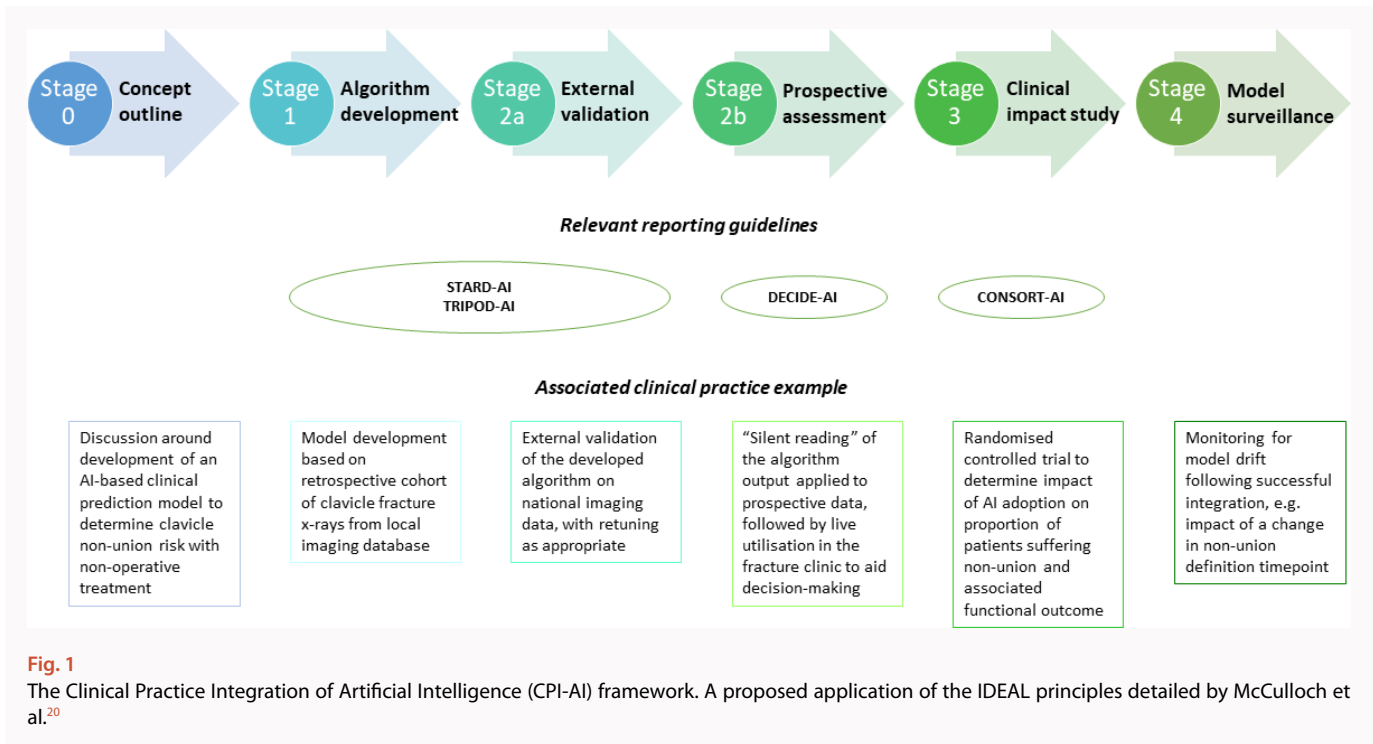
Strengths and limitations

- Strength: Based on the established IDEAL principles governing the introduction of surgical innovation.
- Limitation: Widespread dissemination and uptake of the framework within AI research is required to realistically effect any potential clinical impact.

Introduction

Artificial intelligence (AI) has previously been highlighted as a potential innovation to embrace in trauma and orthopaedics, although caution behind the scale and scope of adoption has been recommended.¹ With the increasing availability and use of large-scale data repositories, it is likely that the use of AI, with its associated abilities to analyze substantial quantities of information and include previously unavailable data sources (for example, digital images and unstructured clinical text information), will play a large role in future healthcare advancements.²

As potential applications of AI to healthcare expand, there is a clear need for key stakeholders (researchers, clinical staff,



industry partners, policymakers, and patients) to understand the pipeline from development of AI algorithms to integration into clinical practice. It is important to have clear guidance on what good evidence looks like at each stage of development, to ensure informed appraisal of emerging evidence and associated claims of benefit. Currently there are hundreds, if not thousands, of developed AI tools that focus on clinical applications, but very few that progress to inform clinical practice.^{3,4}

A number of AI reporting guidelines have been developed (e.g. CONSORT-AI, DECIDE-AI, and TRIPOD-AI), and provide a useful framework to support researchers to produce manuscripts that are of sufficient rigour for stakeholders to understand the potential application of these models.⁵⁻⁸ The vast majority of AI algorithms for use in healthcare are, however, produced without reference to any clear framework to inform potential users of their readiness (or not) for deployment into a useful clinical application. This is perhaps one of the major reasons why we continue to see limited clinical benefit from AI, despite its potential to transform the way in which healthcare operates at all points in the patient journey.^{9,10}

Such guidance exists around other related areas, such as the Medical Research Council (MRC) framework for complex interventions¹¹ and the mobile health (mHealth) evidence reporting and assessment (mERA)¹² checklists, but these are not specific to AI and therefore potentially miss key aspects unique to healthcare innovation in this field. Other published advice relates to specific fields of AI application,¹³ or quality assurance,¹⁴ and do not depict a clear pipeline of the necessary journey from model development to safe use in clinical practice that are applicable to all domains of AI implementation within a healthcare setting. Governmental bodies, including the USA Food and Drug Administration (FDA) and NHS England,^{15,16} have both highlighted the clear need for further development around regulations for AI and

digital health. A robust evidence pipeline is likely to play an integral part of that regulatory process, similar to current medical device and drug developments.

Having a recognized structured pathway for AI deployment (with clear stages of development and associated evidence of benefit required at each stage) would also provide key indicators for safety, and reduce the real danger associated with the potential use of improperly tested AI models within a clinical setting. Perhaps one of the best examples of this is the AI early sepsis prediction model embedded within the EPIC electronic health record system. Researchers found that even though the model was widely accepted, it still failed to identify a significant number of patients with sepsis during external validation, suggesting a lack of robustness and generalization capabilities of this (or a similar) model. The algorithm also set off a high number of false positive warnings for potential sepsis identification that may have led to "alert fatigue".¹⁷ AI models have historically lacked proper cohort representation during development, leading to real-world issues with generalizability and subsequently raising ethical concerns related to poor model performance across different ethnic, racial, and sex categories.^{18,19}

The IDEAL framework provides a well-established and comprehensive pathway from innovation to potential clinical adoption that has been used for several years to guide surgical innovation.²⁰ This has included applications in the field of trauma and orthopaedics, such as development and clinical testing of the X-Bolt Dynamic Hip Plating System culminating in a large-scale randomized controlled trial (RCT) assessment.²¹⁻²³ IDEAL forms a five-stage (six including the pre-clinical stage 0) concept of appropriate clinical integration and underpinning research for surgical therapy innovation: Stage 0 – Theoretical or in vitro testing of clinical utility and risk; Stage 1 – Proof of concept (In-vivo testing); Stage 2a – Development (Prospective Case-series); Stage 2b – Exploration (Prospective Cohort or feasibility RCT); Stage 3 – Assessment

Stage	Item number	Item	Completed
<i>Stage 0 - Concept outline</i>			
	0a	Discussion of the potential concept including key stakeholders that provide relevant methodological and clinical expertise	<input type="checkbox"/>
	0b	Assessment of feasibility and scope, including review of associated literature	<input type="checkbox"/>
	0c	Development and publication of study protocol (Open Access)	<input type="checkbox"/>
<i>Stage 1 - Algorithm development</i>			
	1a	Development and testing of the model within a representative cohort	<input type="checkbox"/>
	1b	Study details published according to relevant reporting checklist (STARD-AI and TRIPOD-AI)	<input type="checkbox"/>
<i>Stage 2a - External validation</i>			
	2a	Testing and reporting upon application of the model within a wider, or separate, clinical setting, which mimics the situation in which clinical practice integration will be considered	<input type="checkbox"/>
	2b	Assessment of the impact on different patient groups (for example variations in age, sex, race/ethnicity and socioeconomic status) to ensure no systematic bias is present	<input type="checkbox"/>
	2c	Reporting of model updates required as a result of external validation, including details of any mitigations introduced and the subsequent impact on model performance*	<input type="checkbox"/>
<i>Stage 2b - Prospective assessment</i>			
	2d	Testing and assessment of the AI intervention within a "live" setting, including consideration of how this relates to current care pathways or gold standard treatments	<input type="checkbox"/>
	2e	Reporting of assessment should be performed according to the DECIDE-AI guidance, focusing on three key areas: Clinical Utility, Safety and Human Factors*	<input type="checkbox"/>
	2f	Consideration of how results from this stage would be utilised to inform design of future clinical impact assessment	<input type="checkbox"/>
<i>Stage 3 - Clinical impact assessment</i>			
	3a	Formal large-scale testing and assessment should be performed in the form of a Randomised Controlled Trial that evaluates both clinical and cost-effectiveness	<input type="checkbox"/>
	3b	Reporting performed in accordance with the CONSORT-AI statement*	<input type="checkbox"/>
<i>Stage 4 - Implementation and model surveillance</i>			
	4a	Demonstrate robust evidence-based support for adoption of the developed model into clinical practice	<input type="checkbox"/>
	4b	The developers should work with clinical and information technology (IT) staff to assist with the integration process in the context of local healthcare IT infrastructure	<input type="checkbox"/>
	4c	A plan should be agreed for model surveillance, which should consider the nature of the developed model, the target population, and the likelihood of potential model or data drift	<input type="checkbox"/>

STARD = Standards for Reporting of Diagnostic Accuracy Study; TRIPOD = Transparent Reporting of a multivariable prediction model for individual Prognosis Or Diagnosis; DECIDE-AI = Developmental and Exploratory Clinical Investigations of DEcision support systems driven by Artificial Intelligence; CONSORT = Consolidated Standards of Reporting Trials. * Reporting should include open access publication of the model code and data utilised for development and testing where feasible within the constraints of Intellectual Property and data protection policies.

Fig. 2

Clinical Practice Integration of Artificial Intelligence (CPI-AI) progression criteria. A guide to aid determination of research integrity and suitability for stage progression.

(Randomized Controlled Trial); and Stage 4 – Long term study (Surveillance to detect rare / late outcomes).

Although specific to surgical innovation, the IDEAL framework provides a valid and rigorous structure which provides key principles to inform the development of frameworks for other forms of innovation across the wider clinical spectrum, including AI, and which could be used in conjunction with other standards such as the BS 30440 Validation framework for the use of AI within healthcare.²⁴

Successful use of the IDEAL framework as a tool for separate subscale development has already been demonstrated in the area of device innovation, culminating in the proposed IDEAL-D framework that provides guidance for the evaluation and regulation of medical devices.²⁵

A proposed framework for AI

A framework to guide the development and evaluation of AI applications, which builds on the IDEAL principles together with insights from other validated AI quality assurance standards, has been proposed – the Clinical Practice Integration of Artificial Intelligence (CPI-AI) framework. As highlighted in Figure 1, it is important to understand that development is an iterative process, and that there may be natural flow forwards and backwards between stages to manage potential bias or model drift. Figure 2 provides a list of progression criteria to aid further understanding of suitability for progression of research through the CPI-AI stages.

Stage 0 – Concept outline

As with the IDEAL framework, this stage involves conceptualizing an AI algorithm with input from a diverse team of AI experts, clinicians, patients, and policymakers. This would involve evaluating relevant background literature and assessing the theoretical risks and benefits of the idea, including technological considerations such as infrastructure, interoperability, and data security. Consideration would also be given to feasibility and scope, potential user groups and output format, as well as ethics, transparency, and interpretability, as these will greatly influence the later stages of development. Generative AI may be used to aid concept development, but should include appropriate consideration to any later potential copyright implications.

Stage 1 – Algorithm development

This is currently by far the most frequently examined area of AI application to healthcare. In this stage, the AI algorithm is developed and initially tested at a basic level as a proof of concept. This will typically involve a single/multicentre study whereby a single large, or multiple smaller, dataset(s) is/are used to develop the AI tool.

AI-related reporting checklists for IDEAL Stage 1 are available and would be used to ensure that all key relevant information is contained within any publication of such work, for example STARD-AI (for diagnostic studies)²⁶ or the TRIPOD+AI guidance (for prediction modelling).⁸ At this stage, no strong references would be made to the potential clinical applicability of any developed AI algorithms without more detailed evaluation conducted as with the later stages of the

IDEAL framework. Given the high-stakes environment of AI healthcare applications, it would be anticipated that substantial attention would be paid to explainability and interpretability of the developed model. Active efforts would be made to confirm the results and performance of those early AI models to other available datasets.

An example of a CPI-AI stage 1 concept would be the development of an AI-based clinical prediction model, to determine risk of nonunion with nonoperative management of a clavicle fracture based on radiograph images at the time of injury. It would be anticipated that this would use local/national imaging databases with a ground truth available to determine who achieved union or not. Sample characteristics, as well as key metrics such as accuracy, recall, precision, f1 score, and specificity, would be detailed. Consideration would be given to how best to move towards CPI-AI Stage 2a.

Stage 2a – External validation

The next stage of algorithm development requires testing outside the dataset where the algorithm was originally trained, a process known as external validation. This is currently lacking in many studies, but is an essential part of appropriate use to ensure that the algorithm remains suitably effective within different populations.²⁷ Multiple external validations are likely needed to account for demographic, clinical, and healthcare infrastructure variability. This external validation is in addition to the common AI development practice that involves splitting a dataset into train, validation, and test sets. The external validation will involve a completely separate dataset that will describe a different population and/or clinical setting.

This stage would likely highlight the need to tune certain parts of the model to better calibrate it within this wider set of patients and/or a new clinical infrastructure. This work is essential to ensuring fairness across patient groups and minimizing any bias present in the initial training data. Further testing would then be performed to demonstrate improved performance in the wider dataset. Again, reporting checklists such as those highlighted above (e.g. TRIPOD-AI and STARD-AI) would be used to ensure that this process is accurately documented and transparent. Reporting at this stage should also include calibration across different patient groups where appropriate, for example sex, age, or disease severity. The level of similarities and differences between the original and external validation cohorts should be highlighted.

An example of a CPI-AI Stage 2a for our clavicle fracture project would be testing of the developed algorithm on a different (multiple) local/national imaging database, with suitable tuning to ensure that the algorithm performs well on a wider subset of patients.

Stage 2b – Prospective assessment

Once confirmed that the algorithm is suitable for wider application, the project would move forwards to assess the feasibility of potential clinical practice implementation. This would be done in a prospective fashion with assessment of the intervention in a 'live' setting. This would include consideration of how use of the algorithm impacts on clinical decision-making, with initial evaluation performed as a 'silent reading' phase, so that impact on patient care is not realised while this process occurs. Attention would be placed on

improving upon the explainability of the AI models (incorporating both local (individual AI outputs) and global (general population level) explainability methods), while tuning it for enhanced performance.

As per the DECIDE-AI reporting checklist,⁶ which has been developed specifically to address this type of early-stage clinical deployment, this stage would focus on three key areas: Clinical Utility, Safety, and Human Factors. Potential cost implications would also be important within financially constrained healthcare settings, which has typically been lacking previously. It would also be important to ensure that the interface is compliant with the necessary regulations regarding the processing and potential storage of patient data. Again, further updating of the model may be required following this prospective assessment and stakeholder feedback with re-testing and fluidity within Stage 2.

An example of a CPI-AI Stage 2b for our clavicle project would be initial 'silent reading' of the algorithm output, followed by live use in the fracture clinic to aid decision-making regarding the potential risks of nonunion and help determine optimal treatment strategy on a case-by-case basis. Saliency maps could be used for local explainability, to help demonstrate the regional importance of included images in algorithmic outcomes to reassure clinical teams and patients of the correct rationale for decision-making.

Stage 3 – Clinical impact assessment (RCT)

Similar to any other healthcare intervention, AI technologies require formal assessment of their impact on clinical practice, utilizing the most robust and unbiased methodology available. Typically, this would be a large-scale intervention in the form of a multicentre RCT, comparing the safety and outcomes of the use of the algorithm against current alternative best practice methodology. The effects on the wider system would also need to be assessed.

Potential outcomes examined would likely differ dependent on the type of the AI algorithm and its intended use, for example with the clavicle project example the primary outcome could be the proportion of patients who successfully avoided a nonunion and their linked patient-reported outcome measure (PROM) data over time (as one would anticipate that successful avoidance of nonunion would improve PROMs, particularly in the earlier stages of evaluation). Cost-effectiveness analyses would likely also be an important part of any assessment, balancing the cost of AI implementation and monitoring versus the potential savings from avoidance of unnecessary surgery.

Reporting checklists for a potential CPI-AI Stage 3 would include the CONSORT-AI statement,⁵ as well as the SPIRIT-AI checklist for RCT protocol design and publication.²⁸

Stage 4 – Implementation and model surveillance

Following successful RCT analysis and confirmation of clinical and cost-effectiveness, implementation of the model is performed. Flexibility of the AI intervention may be required to widen adoption and should be evaluated iteratively. Supporting sustainability has previously been identified as a key driver of long-term behavioural change.²⁹

It is also essential that any fully deployed AI model undergoes serial evaluation (through continued assessment of relative performance metric attainment) to confirm its

execution is maintained over time. AI models are at particular risk of two major issues related to changes in how they function over time: concept drift and data drift.³⁰

Concept drift relates to changes in how one labels data and interprets the algorithm findings. For example, clinicians may change the time at which a nonunion is declared to have occurred, or decide that painless radiological nonunions should be classified in the healed category.

Data drift is the more common concern, where the characteristics of the inputted data change over time. One such example would be an ageing population presenting different population demographics, or perhaps updated X-ray technology that changes the quality of the images assessed by the algorithm.

If model drift is identified, then retraining, and re-tuning, or adapting the model would likely be required to ensure adequate ongoing performance and maintenance of cost-effectiveness. Such techniques are currently under development and could be used in this setting, but further scrutiny is required for the problem in hand.³¹

Conclusion

AI algorithms are complex medical interventions and need to be appropriately evaluated as such. Our suggested CPI-AI framework would allow a clear pathway from development to clinical practice application of AI models in trauma and orthopaedics, with stage gates to ensure appropriate onward development of only effective algorithms – maximizing the potential that AI has to provide a more personalized and precise healthcare service.

Social media

Follow L. Farrow on X @docfarrow

References

1. **Clement ND, Simpson AHRW.** Artificial intelligence in orthopaedics. *Bone Joint Res.* 2023;12(8):494–496.
2. **Kunze KN, Orr M, Krebs V, Bhandari M, Piuze NS.** Potential benefits, unintended consequences, and future roles of artificial intelligence in orthopaedic surgery research: a call to emphasize data quality and indications. *Bone Jt Open.* 2022;3(1):93–97.
3. **Ramkumar PN, Pang M, Polisetty T, Helm JM, Karnuta JM.** Meaningless applications and misguided methodologies in artificial intelligence-related orthopaedic research propagates hype over hope. *Arthroscopy.* 2022;38(9):2761–2766.
4. **Andaur Navarro CL, Damen JAA, Takada T, et al.** Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. *BMJ.* 2021;375:2281.
5. **Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI and CONSORT-AI Working Group.** Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med.* 2020;26(9):1364–1374.
6. **Vasey B, Clifton DA, CollinsGS, DECIDE-AI Steering Group.** DECIDE-AI: new reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence. *Nat Med.* 2021;27(2):186–187.
7. **Farrow L, Zhong M, Ashcroft GP, Anderson L, Meek RMD.** Interpretation and reporting of predictive or diagnostic machine-learning research in Trauma & Orthopaedics. *Bone Joint J.* 2021;103-B(12):1754–1758.
8. **Cohen JF, Bossuyt PMM.** TRIPOD+AI: an updated reporting guideline for clinical prediction models. *BMJ.* 2024;385:q824.
9. **Polisetty TS, Jain S, Pang M, et al.** Concerns surrounding application of artificial intelligence in hip and knee arthroplasty. *Bone Joint J.* 2022;104-B(12):1292–1303.
10. **Varghese J.** Artificial intelligence in medicine: chances and challenges for wide clinical adoption. *Visc Med.* 2020;36(6):443–449.
11. **Skivington K, Matthews L, Simpson SA, et al.** A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *BMJ.* 2021;374:2061.
12. **Agarwal S, LeFevre AE, Lee J, et al.** Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. *BMJ.* 2016;352:i1174.
13. **Bizzo BC, Dasegowda G, Bridge C, et al.** Addressing the challenges of implementing artificial intelligence tools in clinical practice: principles from experience. *J Am Coll Radiol.* 2023;20(3):352–360.
14. **No authors listed.** Blueprint for Trustworthy AI Implementation Guidance and Assurance for Healthcare. Coalition for Health AI. 2023. https://www.coalitionforhealthai.org/papers/blueprint-for-trustworthy-ai_V1.0.pdf (date last accessed 30 August 2024).
15. **No authors listed.** AI Regulation: Improving the regulatory approval process and building trust in robust standards. NHS Transformation Directorate. 2021. <https://transform.england.nhs.uk/ai-lab/ai-lab-programmes/regulating-the-ai-ecosystem/> (date last accessed 30 August 2024).
16. **No authors listed.** Executive Summary for the Patient Engagement Advisory Committee Meeting: Artificial Intelligence (AI) and Machine Learning (ML) in Medical Devices. U.S. Food & Drug Administration. 2021. <https://www.fda.gov/media/151482/download> (date last accessed 10 September 2024).
17. **Wong A, Otles E, Donnelly JP, et al.** External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern Med.* 2021;181(8):1065–1070.
18. **Obermeyer Z, Powers B, Vogeli C, Mullainathan S.** Dissecting racial bias in an algorithm used to manage the health of populations. *Science.* 2019;366(6464):447–453.
19. **Hundt A, Agnew W, Zeng V, Kacianka S, Gombolay M.** Robots Enact Malignant Stereotypes [abstract]. FACCT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency. 2022.
20. **McCulloch P, Altman DG, Campbell WB, et al.** No surgical innovation without evaluation: the IDEAL recommendations. *Lancet.* 2009;374(9695):1105–1112.
21. **Griffin XL, Parsons N, McArthur J, Achten J, Costa ML.** The Warwick Hip Trauma Evaluation One: a randomised pilot trial comparing the X-Bolt Dynamic Hip Plating System with sliding hip screw fixation in complex extracapsular hip fractures: WHiTE (One). *Bone Joint J.* 2016;98-B(5):686–689.
22. **Kahane S, Vaghela KR, Stammers J, Goldberg A, Smitham P.** Biomechanical study comparing cut-out resistance of the X-Bolt® and dynamic hip screw at various tip-apex distances. *Surg Technol Int.* 2019;35:395–401.
23. **Griffin XL, Achten J, O'Connor HM, Cook JA, Costa ML, WHiTE Four Investigators.** Effect on health-related quality of life of the X-Bolt dynamic plating system versus the sliding hip screw for the fixation of trochanteric fractures of the hip in adults: the WHiTE Four randomized clinical trial. *Bone Joint J.* 2021;103-B(2):256–263.
24. **No authors listed.** BS 30440:2023: Validation framework for the use of artificial intelligence (AI) within healthcare. Specification. BSI Group. 2022. <https://landingpage.bsigroup.com/LandingPage/Standard?UPI=00000000030434912> (date last accessed 10 September 2024).
25. **Sedrakyan A, Campbell B, Merino JG, Kuntz R, Hirst A, McCulloch P.** IDEAL-D: a rational framework for evaluating and regulating the use of medical devices. *BMJ.* 2016;353:i2372.
26. **Sounderajah V, Ashrafian H, Aggarwal R, et al.** Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: The STARD-AI Steering Group. *Nat Med.* 2020;26(6):807–808.
27. **de Vries CF, Colosimo SJ, Staff RT, et al.** Impact of different mammography systems on artificial intelligence performance in breast cancer screening. *Radiol Artif Intell.* 2023;5(3):e220146.
28. **Cruz Rivera S, Liu X, Chan A-W, et al.** Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med.* 2020;26(9):1351–1363.
29. **Khalil H, Kynoch K.** Implementation of sustainable complex interventions in health care services: the triple C model. *BMC Health Serv Res.* 2021;21(1):143.

30. **Carter RE, Anand V, Harmon DM, Pellikka PA.** Model drift: when it can be a sign of success and when it can be an occult problem. *Intell Based Med.* 2022;6:100058.

31. **Thota M, Yi D, Leontidis G.** LLEDA -- Lifelong Self-Supervised Domain Adaptation. arXiv.org e-Print archive. 2023. <https://arxiv.org/pdf/2211.09027.pdf> (date last accessed 30 August 2024).

Author information

L. Farrow, MBChB, MRCS, SCREDS Clinical Lecturer, Speciality Registrar, Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, UK; Grampian Orthopaedics, Aberdeen Royal Infirmary, Aberdeen, UK.

D. Meek, MBChB, BSc (Hons), MD, FRCS (Tr&Orth), Consultant Orthopaedic Surgeon, Department of Orthopaedics, Queen Elizabeth University Hospital, Glasgow, UK.

G. Leontidis, BSc, MSc, PhD, Director - Interdisciplinary Centre for Data & Artificial Intelligence

M. Campbell, BSc (Hons), MSc, PhD, CStat, FSS, FFPH, FSCT, FRSE, FMedSci, Professor of Health Services Research

L. Anderson, PhD, MPHe, BSc(Hons), PGCHET, FHEA, Chair of Health Data Science, Institute of Applied Health Sciences, University of Aberdeen, Aberdeen, UK.

E. Harrison, OBE, FRSE, FMedSci, Professor of Surgery and Data Science, Centre of Medical Informatics, University of Edinburgh, Edinburgh, UK.

Author contributions

L. Farrow: Conceptualization, Methodology, Writing – original draft, Writing – review & editing.

D. Meek: Conceptualization, Methodology, Writing – review & editing.

G. Leontidis: Conceptualization, Methodology, Writing – review & editing.

M. Campbell: Conceptualization, Methodology, Writing – review & editing.

E. Harrison: Conceptualization, Methodology, Writing – review & editing.

L. Anderson: Conceptualization, Methodology, Supervision, Writing – review & editing.

Funding statement

The authors disclose receipt of the following financial or material support for the research, authorship, and/or publication of this

article: a grant from the Chief Scientist Office Scotland Clinical Academic Fellowship, as reported by L. Farrow.

ICMJE COI statement

L. Farrow reports receipt of a grant from the Chief Scientist Office (CSO) Scotland Clinical Academic Fellowship, of which the submitted work will form part of his associated PhD thesis, related to this study. M. Campbell reports leadership as Chair of the Medical Research Council (MRC)/National Institute for Health and Care Research (NIHR) Better Methods Better Research funding panel, not related to this study. M. Campbell also reports work as an advisor for the IDEAL Collaboration Council, not related to this study.

Data sharing

The data that support the findings for this study are available to other researchers from the corresponding author upon reasonable request.

Acknowledgements

The authors are grateful to the original IDEAL authors for the outline on which this updated framework is based.

Open access funding

The authors report that they received open access funding for their manuscript from Chief Scientist Office Scotland.

© 2024 Farrow et al. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives (CC BY-NC-ND 4.0) licence, which permits the copying and redistribution of the work only, and provided the original author and source are credited. See <https://creativecommons.org/licenses/by-nc-nd/4.0/>