Clinical Evaluation of Collaborative Artificial Intelligence Systems: Lessons from the Case of Robot-Assisted Surgery

Alexandre Coste[1][a], Frédéric Barbot[2][b] and Thierry Chevalier[3,4,5][c]
[1]EuroMov Digital Health in Motion, Univ. Montpellier, IMT Mines Ales, Montpellier, France
[2]INSERM CIC 1429, Raymond Poincaré Hospital APHP, France
[3]CHU Nîmes, Department of Biostatistics, Epidemiology, Public Health and Innovation in Methodology, 30029 Nîmes, France
[4]Univ. Montpellier, INSERM, UMR 1302, Institute Desbrest of Epidemiology and Public Health, Montpellier, France
[5]Tech4Health-FCRIN, France

Keywords: Clinical Evaluation, Robotic Surgery, Human-Machine Collaboration, Artificial Intelligence, Medical Devices.

Abstract: Collaborative AI systems, which combine both forms of intelligence (i.e., human and machine), are attracting increasing interest from the scientific and medical communities, with various applications in radiology (clinical decision support systems) and surgery (robot-assisted surgery). However, despite their promise, these systems face significant challenges in integrating into clinical practice due to a lack of transparency, trust, and clinical validation. Drawing on the case of robotic surgery, the aim of this work was to analyse the scientific evidence for ten surgical robots currently on the market (i.e., CE-marked or FDA-cleared/approved) that meet the definition of a collaborative AI system. We found a low number of peer-reviewed publications and a lack of transparency from authors and manufacturers, particularly regarding the functioning of their devices, which are often considered as 'black boxes'. Furthermore, the term 'artificial intelligence' is under-utilised in scientific publications, regulatory submissions, and commercial materials. Based on these findings, we propose three recommendations to promote the integration of these medical devices: 1) promote the transparency, explainability, and comprehensibility of AI devices by encouraging manufacturers to provide more detailed information about their systems and their functioning, including the interrelationship with the user; 2) promote randomised controlled multicentre trials to provide stronger evidence on the performance and safety of these devices; 3) encourage the publication of scientific results in peer-reviewed journals to expose them to scientific scrutiny and improve transparency. These recommendations have been carefully formulated to cover a wide range of AI/ML-enabled medical devices, beyond the case of surgical robots reviewed here.

1 INTRODUCTION

Artificial intelligence (AI) is expanding rapidly, particularly in the healthcare sector. Technological advances, particularly in computer science, have led to increasingly powerful AI systems, but paradoxically only a limited number of these systems have been integrated into clinical practice, a phenomenon known as the 'AI chasm' (e.g., Aristidou et al. 2022, Reyna et al. 2022). Key limiting factors include a lack of transparency, trust, interpretability, adaptability and scientific evidence. In particular, many concerns have been raised in recent years about the fact that certain AI systems have been tested and validated using retrospective, *in silico* data, which does not reflect real-world clinical practice. Moreover, few studies have taken into account the specificities of so-called 'collaborative' AI systems. These systems, which are based on the close collaboration between two forms of intelligence, human and artificial, (Vasey et al. 2022),

[a] https://orcid.org/0000-0002-4497-6473
[b] https://orcid.org/0000-0002-4648-9134
[c] https://orcid.org/0000-0002-5110-6273

Coste, A., Barbot, F. and Chevalier, T.
Clinical Evaluation of Collaborative Artificial Intelligence Systems: Lessons from the Case of Robot-Assisted Surgery.
DOI: 10.5220/0012598500003657

In *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2024) - Volume 1*, pages 852-857
ISBN: 978-989-758-688-0; ISSN: 2184-4305

present significant methodological challenges due to their inherent complexity. This complexity arises primarily from the ongoing interplay between human-related factors, such as the learning curve, the level of expertise or the physical and mental fitness of the operators, and AI-model factors, including algorithmic specificities, the evolutionary nature for continuous learning models, and the quality of the learning data that shapes the model and its performance.

The surgical field is undoubtedly one of the most representative areas for collaborative AI systems (Mayor et al. 2022), where the integration of human expertise with AI capabilities shows remarkable potential for advancing surgical practices. In particular, the proposed benefits include: i) enhancing the surgeon's perceptual abilities through three-dimensional imaging; ii) improving the precision of surgical gestures, particularly in minimally invasive procedures, by filtering out tremors and reducing differences associated with laterality preferences.

While autonomous surgery was the main motivation for the pioneers (e.g., the PROBOT for prostate resection), it is the robots that assist the surgeon (i.e., teleoperated or co-manipulated), not intended to replace him, that have become widespread over the last twenty years. There are now hundreds of surgical robots (on the market or under development), covering various medical indications, from general surgery, to gynaecology, orthopaedics and even cardiac surgery. The Da Vinci surgical system, developed by Intuitive Surgical, currently dominates the market with more than 6,000 units sold worldwide and more than 7 million procedures carried out with the robot (figures given by the manufacturer on its website https://www.intuitive.com/). However, little is known about the clinical evaluation of these medical devices required for both European (Medical Device Regulation, MDR) and American (FDA) compliance. In this context, the objective of the present research is to provide an overview of commercially available collaborative AI systems in robotic surgery and to review the associated scientific evidence.

2 METHODS

Following a similar methodology to Wu et al. (2021), Benjamens et al. (2020), and van Leeuwen et al. (2021), we identified ten robotic surgical systems currently available in the market (i.e., compliant with European regulations or FDA approved/cleared, see Figure 1).

2.1 Search Strategy and Selection Criteria

The surgical robots selection process was carried out in two phases. First, we used the following resources/databases:

i) FDA's database:"AI/ML-Enabled Medical Devices," listing FDA approved AI/ML-based medical devices.
ii) The recent review by Muehlematter, Daniore & Vokinger (2021) listing 462 AI-based devices approved in Europe and the U.S. from 2015 to 2020.
iii) The new European Medical Devices Database (Eudamed).
iv) The list of communications for Class IIa, IIb, and III medical devices and implantable medical devices from the ANSM (French National Agency for Medicines and Health Products Safety), covering devices in the market from 2010 to 01/12/2021 (n = 83129).
v) PubMed® and Google Scholar® databases.

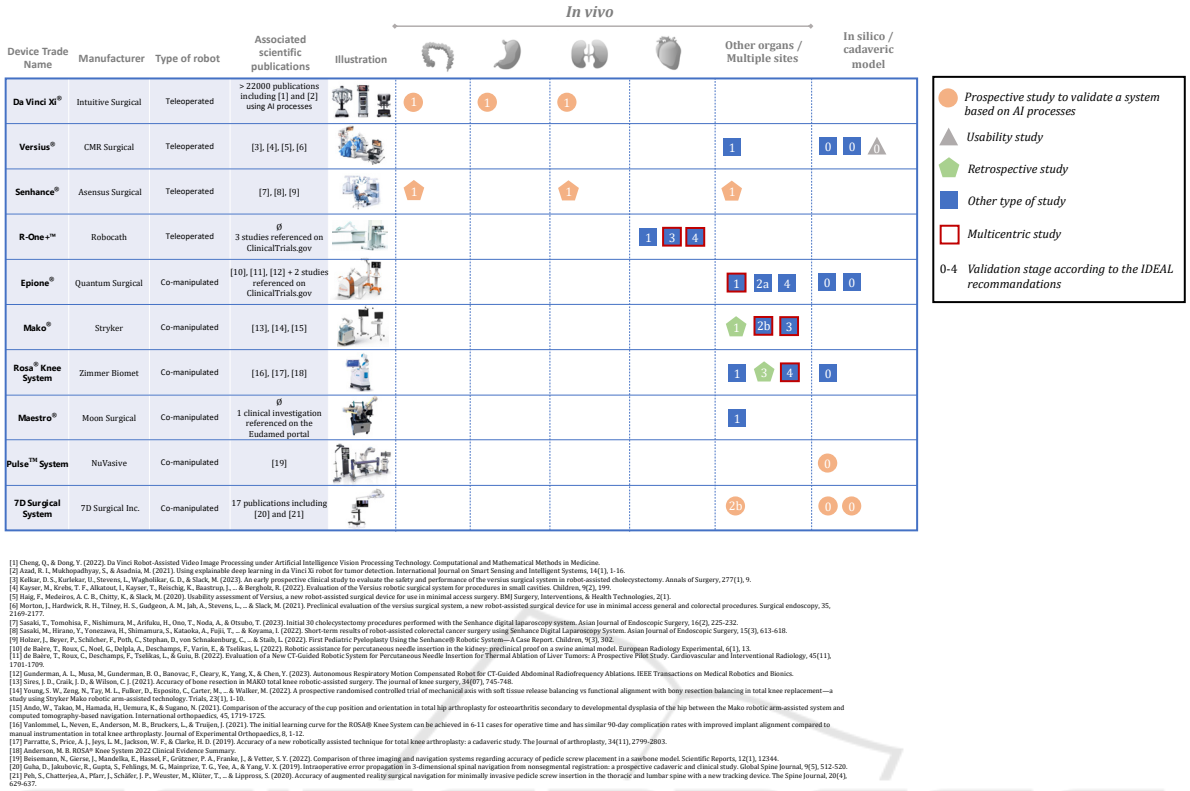
To ensure relevant results, precise keywords were identified using the bilingual version of the INSERM (French National Institute of Health and Medical Research) MeSH (Medical Subject Headings) lexicon. Keywords included terms related to robotic surgery, artificial intelligence and machine learning.

A search of these keywords against those in the above databases identified approximately one hundred potential devices. A detailed analysis involving cross-referencing with various sources, including manufacturers' websites and commercial documentation, led to the selection of devices that met the following criteria:

vi) Surgical robots commercially available in the European or American markets (i.e., EU-MDR or FDA compliant).
vii) Collaborative surgical robots involving human-machine interaction (i.e., co-manipulated or teleoperated).
viii) Surgical robots incorporating AI, machine learning, or deep learning processes.

2.2 Analysis of Scientific Evidence and Clinical Evaluation Methodology

The level of scientific evidence and clinical evaluation methodology for the ten selected devices were examined using two methods. Firstly, a systematic search of PubMed, Google Scholar and



[1] Cheng, Q., & Dong, Y. (2022). Da Vinci Robot-Assisted Video Image Processing under Artificial Intelligence Vision Processing Technology. Computational and Mathematical Methods in Medicine.
[2] Azad, R. L., Mukhopadhyay, S., & Asadnia, M. (2021). Using explainable deep learning in da Vinci Xi robot for tumor detection. International Journal on Smart Sensing and Intelligent Systems, 14(1), 1-16.
[3] Kelkar, D. S., Kurlekar, U., Stevens, L., Waghollikar, G. D., & Slack, M. (2023). An early prospective clinical study to evaluate the safety and performance of the versius surgical system in robot-assisted cholecystectomy. Annals of Surgery, 277(1), 9.
[4] Kayser, M., Krebs, T. F., Alkatout, I., Kayser, T., Reischig, K., Baastrup, J., ... & Bergholz, R. (2022). Evaluation of the Versius robotic surgical system for procedures in small cavities. Children, 9(2), 199.
[5] Haig, F., Medeiros, A. C. R., Chitty, K., & Slack, M. (2020). Usability assessment of Versius, a new robot-assisted surgical device for use in minimal access surgery. BMJ Surgery, Interventions, & Health Technologies, 2(1).
[6] Morton, J., Hardwick, R. H., Tilney, H. S., Gudgeon, A. M., Jah, A., Stevens, L., ... & Slack, M. (2021). Preclinical evaluation of the versius surgical system, a new robot-assisted surgical device for use in minimal access general and colorectal procedures. Surgical endoscopy, 35, 2169-2177.
[7] Sasaki, T., Tomohisa, F., Nishimura, M., Arifuku, H., Ono, T., Noda, A., & Otsubo, T. (2023). Initial 30 cholecystectomy procedures performed with the Senhance digital laparoscopy system. Asian Journal of Endoscopic Surgery, 16(2), 225-232.
[8] Sasaki, M., Hirano, Y., Yonezawa, H., Shimamura, S., Kataoka, A., Fujii, T., ... & Koyama, I. (2022). Short-term results of robot-assisted colorectal cancer surgery using Senhance Digital Laparoscopy System. Asian Journal of Endoscopic Surgery, 15(3), 613-618.
[9] Holzer, J., Beyer, P., Schlicher, F., Poth, C., Stephan, D., von Schnakenburg, C., ... & Staab, L. (2022). First Pediatric Pyeloplasty Using the Senhance® Robotic System—A Case Report. Children, 9(3), 302.
[10] de Baère, T., Roux, C., Noel, G., Delpla, A., Deschamps, F., Varin, E., & Tselikas, L. (2022). Robotic assistance for percutaneous needle insertion in the kidney: preclinical proof on a swine animal model. European Radiology Experimental, 6(1), 13.
[11] de Baère, T., Roux, C., Deschamps, F., Tselikas, L., & Guiu, B. (2022). Evaluation of a New CT-Guided Robotic System for Percutaneous Needle Insertion for Thermal Ablation of Liver Tumors: A Prospective Pilot Study. Cardiovascular and Interventional Radiology, 45(11), 1701-1709.
[12] Gunderman, A. L., Musa, M., Gunderman, B. O., Banovac, F., Cleary, K., Yang, X., & Chen, Y. (2023). Autonomous Respiratory Motion Compensated Robot for CT-Guided Abdominal Radiofrequency Ablations. IEEE Transactions on Medical Robotics and Bionics.
[13] Sires, J. D., Craik, J. D., & Wilson, C. J. (2021). Accuracy of bone resection in MAKO total knee robotic-assisted surgery. The journal of knee surgery, 34(07), 745-748.
[14] Young, S. W., Zeng, N., Tay, M. L., Fulton, D., Esposito, C., Carter, M., ... & Walker, M. (2022). A prospective randomised controlled trial of mechanical axis with soft tissue release balancing vs functional alignment with bony resection balancing in total knee replacement—a study using Stryker Mako robotic arm-assisted technology. Trials, 23(1), 1-10.
[15] Ando, W., Takao, M., Hamada, H., Uemura, K., & Sugano, N. (2021). Comparison of the accuracy of the cup position and orientation in total hip arthroplasty for osteoarthritis secondary to developmental dysplasia of the hip between the Mako robotic arm-assisted system and computed tomography-based navigation. International orthopaedics, 45, 1719-1725.
[16] Vanlommel, L., Neven, E., Anderson, M. B., Bruckers, L., & Truijen, J. (2021). The initial learning curve for the ROSA® Knee System can be achieved in 6-11 cases for operative time and has similar 90-day complication rates with improved implant alignment compared to manual instrumentation in total knee arthroplasty. Journal of Experimental Orthopaedics, 8, 1-12.
[17] Parratte, S., Price, A. J., Jeys, L. M., Jackson, W. F., & Clarke, H. D. (2019). Accuracy of a new robotically assisted technique for total knee arthroplasty: a cadaveric study. The Journal of arthroplasty, 34(11), 2799-2803.
[18] Anderson, M. B. ROSA® Knee System 2022 Clinical Evidence Summary.
[19] Beisemann, N., Gierse, J., Mandelka, E., Hassel, F., Grützner, P. A., Franke, J., & Vetter, S. Y. (2022). Comparison of three imaging and navigation systems regarding accuracy of pedicle screw placement in a sawbone model. Scientific Reports, 12(1), 12344.
[20] Guha, D., Jakubovic, R., Gupta, S., Fehlings, M. G., Maingruise, T. G., Yee, A., & Yang, V. X. (2019). Intraoperative error propagation in 3-dimensional spinal navigation from nonsegmental registration: a prospective cadaveric and clinical study. Global Spine Journal, 9(5), 512-520.
[21] Poh, S., Chatterjea, A., Pfarr, J., Schäfer, J. P., Wesuster, M., Klüter, T., ... & Lapresse, S. (2020). Accuracy of augmented reality surgical navigation for minimally invasive pedicle screw insertion in the thoracic and lumbar spine with a new tracking device. The Spine Journal, 20(4), 629-637.

Figure 1: Overview of the ten collaborative surgical robots integrating AI/ML processes marketed in the U.S. and/or Europe.

IEEE Xplore was performed using the trade name and/or the manufacturer name of each robot. This allowed us to extract peer-reviewed articles from 1988 to April 2023. The ClinicalTrials.gov registry and the medRxiv biomedical research preprint platform were also consulted to avoid potential publication bias and to obtain a comprehensive view of ongoing research.

Secondly, the FDA and European Commission (Eudamed) databases were consulted to access detailed information on devices, including preclinical and clinical data submitted by manufacturers to regulatory authorities for the conformity assessment.

3 RESULTS

Figure 1 presents the ten collaborative surgical robots selected for the analysis, categorized according to their trade name, their manufacturer, their type (teleoperated or co-manipulated), the associated scientific publications, and the type of validation studies.

3.1 Collaborative Surgical Robots: A Highly Heterogeneous Landscape

Among the ten collaborative surgical robots, four are teleoperated (Da Vinci Xi®, Versius®, Senhance®, R-One+™), and six are co-manipulated (Epione®, Mako®, Rosa Knee System®, Maestro®, Pulse System™, 7D Surgical System®). All teleoperated robots belong to the same risk class, i.e., Class IIb for EU-MDR compliance (4/4) and Class II for FDA compliance when obtained (2/4: Da Vinci Xi® and Senhance®). In contrast, the risk class for co-manipulated devices is more heterogeneous, ranging from Class I to Class IIb for EU-MDR compliance and from Class I to II for FDA compliance. This diversity is partly explained by the variety of technologies used and the range of covered indications, including orthopaedic, cardiac, spinal, and general laparoscopic surgery. Notably, the Maestro® robot stands out by being classified in the lowest risk class (Class I), contrary to the general trend where active devices are typically classified at least in Class IIa according to the MDR. Also, it is important to note that all the analysed surgical robots have obtained U.S. compliance through the 510(k) procedure, a simplified procedure highly coveted by

manufacturers. Indeed, the manufacturers must only demonstrate that their device is as safe and effective, i.e., substantially equivalent, to a legally marketed device.

3.2 A Significant Lack of Transparency

Surprisingly, 70% of robot manufacturers do not explicitly mention the use of artificial intelligence or machine learning processes. Some manufacturers, such as Asensus Surgical®, use instead terms like 'augmented intelligence' without explicit mention of AI in regulatory documents. Nuvasive® is one of the few manufacturers explicitly using the term 'artificial intelligence' on its website, but the term does not appear in any compliance submission documents. AI or not AI: it seems that talking about artificial intelligence can be beneficial in certain cases, less so in others, particularly with regulatory authorities.

3.3 Lack of Scientific Evidence

A detailed examination of publications associated with the devices reveals varied levels of scientific evidence. While some devices have limited or no peer-reviewed articles, others, like the Da Vinci®, have extensive literature due to their longer market presence. Importantly, the number of studies specifically dedicated to evaluating AI algorithm performance and safety is extremely limited, even for the well-established robot like Da Vinci®. Moreover, most of these studies focus on preclinical stages or involve a very small number of patients. Only 20% of the analysed studies (6 out of 30) are multicentric, emphasizing the need for more comprehensive research.

4 DISCUSSION

The aim of this work was to delineate the contours and inherent challenges in the clinical validation of collaborative AI systems, particularly those involving close collaboration between human and artificial intelligence, with a particular focus on robotic surgery. As mentioned above, these systems present new methodological challenges in clinical evaluation due to the inherent variability of individual factors and those related to the AI system itself. The introduction of an AI component adds a new dimension and complexity to the existing challenges of validating technological innovation in surgery. Previously, it was known that different levels of expertise could lead to different levels of performance, creating a performance bias in favour of established technologies (Rudicel & Esdaile, 1985). Now, the performance of collaborative AI systems can vary between different user profiles. Furthermore, the capabilities of continuously learning systems can evolve over time, either positively or negatively.

4.1 Current Regulatory Framework and Issues

The current regulatory framework, both in Europe with the MDR and in the U.S., does not adequately address the specificities of AI systems, especially collaborative AI systems. Formulated at a time when devices had limited interactivity and infrequent updates, the existing framework struggles to accommodate the evolving and interactive nature of new technologies, particularly those incorporating AI (Gilbert et al., 2023). However, this is changing, with the imminent arrival of the first regulation on artificial intelligence (i.e., 'EU AI Act') and the FDA's draft guidance 'Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence/Machine Learning (AI/ML) - Enabled Device Software Functions / Draft Guidance for Industry and Food and Drug Administration Staff'. The latter specifically aims to address the evolving nature of these new devices, which are capable of real-time or near real-time learning. These regulatory advances are welcome, particularly in light of the observations made in this work.

4.2 Promoting Transparency, Explainability, and Intelligibility of Devices

A key finding of this study is that leading surgical robots on the market currently lack sufficient detailed information about the technologies used, despite recommendations from the World Health Organization in its core ethical principle number 3 (i.e., ensure transparency, explainability, and intelligibility) and the ISO/IEC TR 24028:2020. While recognising the highly competitive nature of the robotic surgery market, characterised by a constant drive for innovation and the protection of intellectual property, it remains crucial to ensure a minimum level of transparency, particularly in the context of AI. Transparency goes beyond regulatory compliance and is a key factor in building trust among both practitioners and patients. The development of Eudamed represents a real opportunity for greater

transparency on the part of manufacturers, as envisaged by the European Commission in the creation of this unique database, which will provide public access to certain information on marketed medical devices. in Europe (device identification, reported incidents, ongoing clinical investigations, ...). However, it is regrettable that the Eudamed database is not yet fully operational and is not as comprehensive as the FDA databases. It is also regrettable that the Summary of Safety and Clinical Performance (SSCP), required by the Art. 32 of the MDR, is limited to implantable devices and class III devices. As we have observed, most surgical robots fall into the IIa and IIb categories and are therefore not directly subject to this obligation.

4.3 Promoting Randomized Controlled Multicentre Studies and Scientific Publications

Another important point is the lack of robust evidence from rigorous clinical trials. Indeed, most of the reported trials were monocentric and observational, which can lead to significant methodological biases. In particular, monocentric studies may produce results that are not generalisable to geographically diverse patient populations with different economic, educational, social, behavioural, ethnic and cultural characteristics (Kaushal et al, 2020). In addition, randomised controlled trials (RCTs) are considered the gold standard in clinical trials as they provide the highest level of scientific evidence. In this regard, authors/manufacturers can use various published guidelines such as SPIRIT-AI (Rivera et al., 2020), DECIDE-AI (Vasey et al., 2022), STARD-AI (Sounderajah et al., 2021), TRIPOD-AI and PROBAST-AI (Collins et al., 2021) to better develop their research protocols and write their scientific papers.

4.4 Limitations and Future Perspectives

Naturally, this research has some limitations. Firstly, it focuses exclusively on surgical robots, which limits its representativeness in terms of the diversity of AI solutions available on the market. However, these surgical robots illustrate well the concept of collaborative AI systems, and the recommendations formulated herein are intended to be transversal and applicable to a wider range of medical devices, including autonomous or non-surgical devices. Secondly, it is important to note that our analysis exclusively concentrated on robots that are already on the market (i.e., having obtained EU or US conformity), specifically in the context of their clinical validation This approach excludes the pre-approval phases, including the development of the idea into a product. Consequently, there might exist additional barriers not identified within this study. For a more comprehensive insight, future research could expand its purview by examining a wider range of medical devices, including aspects associated with the development of medical devices. Furthermore, it would be interesting to consider an extension of the IDEAL protocol (i.e., IDEAL-AI, see McCulloch et al. 2009) to include specificities related to the validation of surgical technology innovations based on AI/ML processes, such as the collaborative surgical robots studied here.

5 CONCLUSIONS

In this work, we have identified several barriers to the implementation of collaborative AI systems in clinical practice, in particular the lack of transparency and scientific publications. We have therefore formulated a set of recommendations aimed at promoting the integration of AI systems into clinical practice, namely: i) promoting transparency, explainability and intelligibility of AI devices, ii) promoting the conduct of randomised controlled multicentre trials, and iii) encouraging the publication of study results in peer-reviewed journals. These recommendations have been formulated to be as transversal as possible and applicable to a wide range of AI/ML-enabled medical devices, not just surgical robots.

ACKNOWLEDGEMENTS

AC thanks the Tech4Health network (https://www.reseau-tech4health.fr) for its financial support.

REFERENCES

Aristidou, A., Jena, R., & Topol, E. J. (2022). Bridging the chasm between AI and clinical implementation. The Lancet, 399(10325), 620.

Benjamens, S., Dhunnoo, P., & Meskó, B. (2020). The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ digital medicine, 3(1), 118.

Collins, G. S., Dhiman, P., Navarro, C. L. A., Ma, J., Hooft, L., Reitsma, J. B., ... & Moons, K. G. (2021). Protocol

for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ open, 11(7), e048008.

Gilbert, S., Fenech, M., Hirsch, M., Upadhyay, S., Biasiucci, A., & Starlinger, J. (2021). Algorithm change protocols in the regulation of adaptive machine learning–based medical devices. Journal of Medical Internet Research, 23(10), e30545

Kaushal, A., Altman, R., & Langlotz, C. (2020). Geographic distribution of US cohorts used to train deep learning algorithms. Jama, 324(12), 1212-1213.

Mayor, N., Coppola, A. S., & Challacombe, B. (2022). Past, present and future of surgical robotics. Trends in Urology & Men's Health, 13(1), 7-10.

McCulloch, P., Altman, D. G., Campbell, W. B., Flum, D. R., Glasziou, P., Marshall, J. C., & Nicholl, J. (2009). No surgical innovation without evaluation: the IDEAL recommendations. The Lancet,

374(9695), 1105-1112.Muehlematter, U. J., Daniore, P., & Vokinger, K. N. (2021). Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. The Lancet Digital Health, 3(3), e195-e203.

Reyna, M. A., Nsoesie, E. O., & Clifford, G. D. (2022). Rethinking algorithm performance metrics for artificial intelligence in diagnostic medicine. JAMA, 328(4), 329-330.

Rivera, S. C., Liu, X., Chan, A. W., Denniston, A. K., Calvert, M. J., Ashrafian, H., ... & Yau, C. (2020). Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. The Lancet Digital Health, 2(10), e549-e560.

Rudicel, S., & Esdaile, J. (1985). The randomized clinical trial in orthopaedics: obligation or option? JBJS, 67(8), 1284-1293.

Sounderajah, V., Ashrafian, H., Golub, R. M., Shetty, S., De Fauw, J., Hooft, L., ... & Liu, X. (2021). Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. BMJ open, 11(6), e047709.

Vasey, B., Nagendran, M., Campbell, B., Clifton, D. A., Collins, G. S., Denaxas, S., ... & McCulloch, P. (2022). Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nature medicine, 28(5), 924-933.

van Leeuwen, K. G., Schalekamp, S., Rutten, M. J., van Ginneken, B., & de Rooij, M. (2021). Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *European radiology*, *31*, 3797-3804.

Wu, E., Wu, K., Daneshjou, R., Ouyang, D., Ho, D. E., & Zou, J. (2021). How medical AI devices are evaluated: limitations and recommendations from an analysis of FDA approvals. Nature Medicine, 27(4), 582-584.