Embedding patient-reported outcomes at the heart of artificial intelligence health-care technologies

Samantha Cruz Rivera, Xiaoxuan Liu, Sarah E Hughes, Helen Dunster, Elaine Manna, Alastair K Denniston, Melanie J Calvert**

Integration of patient-reported outcome measures (PROMs) in artificial intelligence (AI) studies is a critical part of the humanisation of AI for health. It allows AI technologies to incorporate patients' own views of their symptoms and predict outcomes, reflecting a more holistic picture of health and wellbeing and ultimately helping patients and clinicians to make the best health-care decisions together. By positioning patient-reported outcomes (PROs) as a model input or output we propose a framework to embed PROMs within the function and evaluation of AI health care. However, the integration of PROs in AI systems presents several challenges. These challenges include (1) fragmentation of PRO data collection; (2) validation of AI systems trained and validated against clinician performance, rather than outcome data; (3) scarcity of large-scale PRO datasets; (4) inadequate selection of PROMs for the target population and inadequate infrastructure for collecting PROs; and (5) clinicians might not recognise the value of PROs and therefore not prioritise their adoption; and (6) studies involving PRO or AI frequently present suboptimal design. Notwithstanding these challenges, we propose considerations for the inclusion of PROs in AI health-care technologies to avoid promoting survival at the expense of wellbeing.

Introduction

Artificial intelligence (AI) is an area of enormous interest that is transforming health care and biomedical research. AI systems have shown the potential to support patients, clinicians, and health-care infrastructure. AI systems could provide rapid and accurate image interpretation, disease diagnosis and prognosis, improved workflow, reduced medical errors, and lead to more efficient and accessible care.[1,2] Incorporation of patient-reported outcome measures (PROMs), could advance AI systems by helping to incorporate the patient voice alongside clinical data. Validated PROMs provide a patient perspective on the impact that a disease (and its treatment) has on their physical, functional, and psychological status without interpretation from anyone else.[3]

Patient-reported outcomes (PROs) are commonly used in clinical trials to provide evidence regarding the safety and effectiveness of health interventions.[4] Beyond clinical trials, PRO data can be used in prognostic models and in routine care settings to monitor and manage disease progression and response to treatment, and to triage patients (figure 1).

A patient's health-care journey from symptom screening, assessment, and treatment planning to symptom and disease management can be facilitated using PROs in routine care. PROs can be used to tailor care to individual patient needs. Electronic capture of PROs in clinic and between appointments allows for real-time monitoring of symptoms and flexible scheduling of appointments in response to patient need, and aids early detection and management of health problems.[4] The integration of AI in PRO data collection in trials and routine care has the potential to reduce symptom burden for people by tailoring questions to their needs.[5] When considering the inclusion of PROs as parameters in AI modelling, it is important to assess the acceptability of the measures among end users (ie, health-care professionals and patients). This process of assessment helps to determine whether the PRO is applicable and has practical relevance to the end users.

PRO data can provide valuable information alongside clinical data to inform clinical decision making. However, scepticism of PROs persists despite evidence suggesting that the fidelity of PRO measurement compares favourably with many objective clinical measures.[6] The severity or nature of patients' experience with their disease (eg, symptom-event reporting and severity ratings) often differs from clinical outcomes.[7] Therefore, the value of inclusion of PROs in AI systems for conditions with high symptom burden should be considered, for example, in the context of mental health, pain management, or cancer treatment.[6,8,9]

Current use of AI in PRO measurement

The application of AI in PRO measurement for predicting clinical outcomes and building prognostic models was examined in 2021.[10] The literature review identified 15 studies demonstrating that PRO data are increasingly being used in clinical analysis research, modelling, and building decision support systems for practitioners.[10] Six articles (40%) focused on post-surgical outcomes, such as achieving minimal clinically important difference

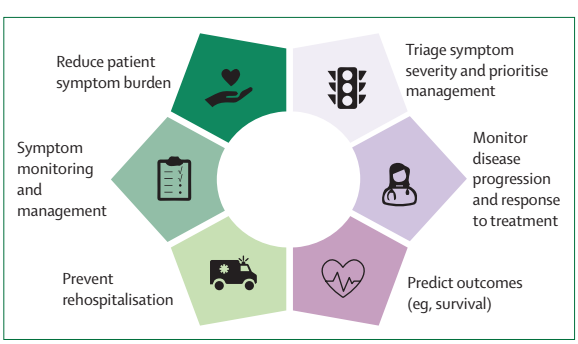

Figure 1: **Use of patient-reported outcomes in routine clinical care**

Lancet Digit Health 2023; 5: e168–73

*Joint last authors

Centre for Patient Reported Outcomes Research, Institute of Applied Health Research (S Cruz Rivera PhD, S E Hughes PhD, E Manna BSc, Prof A K Denniston PhD, Prof M J Calvert PhD), **Birmingham Health Partners Centre for Regulatory Science and Innovation** (S Cruz Rivera, S E Hughes, Prof A K Denniston, Prof M J Calvert), **Data-Enabled Medical Technologies and Devices Hub** (S Cruz Rivera, Prof A K Denniston, Prof M J Calvert), **Academic Unit of Ophthalmology, Institute of Inflammation and Ageing** (X Liu MBChB PhD, Prof A K Denniston), **University of Birmingham Enterprise** (H Dunster BSc), **and National Institute for Health and Care Research Applied Research Collaboration West Midlands** (Prof M J Calvert), **University of Birmingham, Birmingham, UK; Health Data Research UK, London, UK** (Prof A K Denniston, Prof M J Calvert); **National Institute for Health and Care Research Biomedical Research Centre for Ophthalmology, Moorfields Hospital London NHS Foundation Trust and Institute of Ophthalmology, University College London, London, UK** (Prof A K Denniston, Prof M J Calvert); **National Institute of Health Research Applied Research Collaborative West Midlands, Birmingham, UK** (S E Hughes); **National Institute for Health and Care Research Birmingham–Oxford Blood and Transplant Research Unit in Precision Transplant and Cellular Therapeutics, Birmingham, UK** (Prof M J Calvert); **National Institute for Health and Care Research Birmingham Biomedical Research Centre** (Prof M J Calvert), **and National Institute for Health and Care Research Surgical Reconstruction and Microbiology Centre** (Prof M J Calvert), **University Hospitals Birmingham NHS**

Foundation Trust (X Liu, Prof A K Denniston), **Birmingham, UK**

Correspondence to: Dr Samantha Cruz Rivera, Centre for Patient Reported Outcomes Research, Institute of Applied Health Research, University of Birmingham, Birmingham B15 2TT, UK
s.rivera@bham.ac.uk

(MCID) after total joint arthroplasty. MCIDs are patient-derived scores that reflect the minimal amount of change in a clinical intervention that is perceived as beneficial by the patient in terms of magnitude of improvement and the value given to that change.[11,12] Four articles (27%) were aimed at identifying patients with depression and prognosis of anti-depression treatment. Three articles (20%) focused on the prediction of pain, one article (7%) on the risk of hospital re-admission, and one article (7%) on children's oral health outcomes. The review highlighted an absence of external validation and non-availability of datasets as major challenges shared by most articles. Eight (53%) of the 15 studies used a dataset larger than 1000 patients, highlighting the scarcity of large datasets. Six articles (40%) reported essential elements of machine learning-model development, whereas nine articles (60%) failed to adequately do this, potentially reducing the reproducibility of the results. The review did not provide information on the included studies use of PRO data.

Electronic patient-reported outcomes (ePROs)

Individual and clustered patient symptoms can be mapped using ePROs, in combination with clinical data (eg, biomarkers, clinical history, and social determinants of health) from electronic health records. Mapping individual and clustered patient symptoms over time has the potential to provide new insights into disease trajectories, offering earlier diagnosis, patient risk stratification, earlier clinical intervention, and improved patient outcomes.[13,14] Currently, there is scarce evidence on the cost and benefits of including PROs in AI systems. However, in 2016 an oncology clinical trial used web-based symptom reporting with automated clinician email alerts for routine care symptom monitoring. Use of the web-based system was cost-effective and resulted in improved patient survival, reduced hospitalisation, reduced emergency room visits, and improved patient quality of life.[13,14] Although the system did not use AI, there is an opportunity to automate the system. Further work will be required to evaluate the effectiveness and cost-effectiveness of AI interventions, including those with PROs and other parameters, as noted in 2022 by the UK's health-technology assessor, the National Institute for Health and Care Excellence.[15]

The transition to ePROs opens the door to more meaningful integration of PROs into AI-enabled health care, helping to ensure that the patient perspective is not lost in the rush to advance digital health care. A recent review of trials of AI interventions between 1991 and 2021 showed that the main clinical areas that include PROMs in AI trials are mental and behavioural, musculoskeletal, endocrine, and nutritional and metabolic health.[16] The review identified 627 trials of AI health technologies. Of these, 152 (24%) trials included at least one PROM, visual analogue scale (VAS), patient-reported experience measure, or usability measure.

Of the 152 AI trials of health interventions, 44 (29%) focused on testing AI-enabled smart-device applications. These applications were mainly designed for the treatment of mental and behavioural health (n=15) and were intended for patient use. In addition, 23 (15%) of the trials tested clinical decision support systems designed to assist patients or health-care providers in making health-care decisions. Of the 23 trials, six (26%) tested clinical decision support systems that informed decisions on mental and behavioural health, and five (22%) informed musculoskeletal health.[16]

Embedding PROs in AI-enabled health care

We outline a brief framework of three ways that PROs can be embedded within the development and delivery of AI health-care systems: (1) PROs as a model input. The PRO is one of potentially several inputs into an AI model that is trained to predict an outcome. (2) PROs as a model output. The PRO itself is the output that the model is trained to predict. (3) PROs as the outcome by which an AI intervention is evaluated. The PRO is not necessarily an input or an output of the model, but rather it is the outcome used to assess the AI intervention (incorporating the model and the arising action; figure 2).

PROs as an input to an AI model

The use of PRO data as a model input in combination with other clinical parameters can determine the likelihood of a particular diagnosis. For example, machine learning algorithms have been trained to predict the presence and onset of immune-related adverse events in people with cancer treated with immune-checkpoint inhibitors. By use of PRO data related to advanced cancer symptoms, prospectively collected immune-related adverse event data, Common Terminology Criteria for Adverse Events class, and nature and severity of 26 immune-related adverse events (eg, colitis, diarrhoea, and arthritis),[17] an AI prediction model was able to accurately predict the presence and severity of immune-related adverse events, albeit not externally validated in that study.[17] Predicting the future onset of immune-related adverse events, such as fever, chest, and stomach pain, has the potential to improve symptom control and quality of life and, in rare cases, avoid potentially fatal immune-related adverse events, such as colon perforation, pneumonitis, or myocarditis. Since these same symptoms often indicate progressive disease in people with lung cancer or people with lung metastases, timely diagnostic measures are important to evaluate whether the symptoms arising relate to progression of the disease or to the treatment.[17]

PROs as an output from an AI model

In addition, PRO data can be used as an output in AI models, whereby the model is designed to predict a PRO state such as a change in symptoms or health-related quality of life. In this approach, the value of the AI model

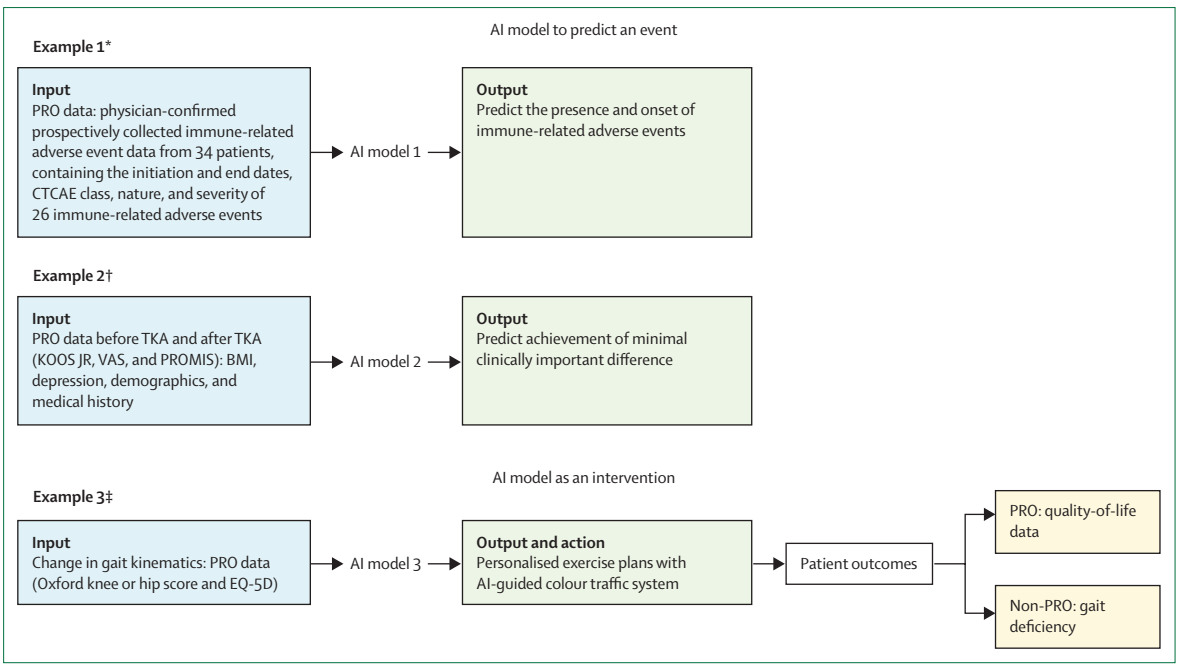

Figure 2: **A framework for the use of PROs in the development and delivery of AI health-care inverventions**
PRO=patient-reported outcome. CTCAE=Common Terminology Criteria for Adverse Events. TKA=total knee arthroplasty. KOOS JR=Knee Injury and Osteoarthritis Outcome Score for Joint Replacement. VAS=visual analog scale. PROMIS=Patient-Reported Outcomes Measurement Information System. *PROs only used as an input. †PROs used as an output and as one of the different inputs to the model. ‡PROs used as an outcome and as one of the input variables to the model.

is that it predicts some aspect of the lived experience of patients. For example, an orthopaedic pilot study showed that machine learning methods have the potential to help to inform the decision to proceed to total knee arthroplasty (TKA) by determining the likelihood of patients achieving MCID after TKA surgery.[18] The algorithm used the Knee Injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS JR),[19] the VAS,[20] and the Patient-Reported Outcomes Measurement Information System (PROMIS) to measure PRO data before and after surgery. These PRO data were combined with patient characteristics (eg, demographics and medical history) to provide input data. The pilot showed that the algorithm has the potential to predict better outcomes for patients undergoing surgery by carefully selecting patients who are likely to achieve MCID in KOOS JR, VAS, and PROMIS after surgery (output data). The PRO data were used to predict whether a MCID could be attained for patients with similar conditions and comorbidities.[18] Such a tool can support shared decision making by providing a personalised prediction of clinical change that is meaningful to the patient (ie, percentage of individuals like you who see significant improvement, stay about the same, or find that their symptoms are worse after surgery).

A further example of PROs as an output of AI models is the 2021 study by Pierson and colleagues,[21] in which an AI model was trained to predict knee pain scores (output) based on x-rays (input). Higher ratings of knee pain are reported in underserved populations. It had previously been reported that these pain disparities persisted even after attempting to correct for radiological disease severity. However in the study by Pierson and colleagues,[21] the human radiologist was compared with a deep-learning model trained on racially and socioeconomically diverse data. The algorithm was able to predict severity of pain in underserved groups much more accurately than the human, so accounting for much of the previously unexplained disparity. By choosing a prediction task focused on PROMs, this approach provides a means of discovery for how AI can improve care beyond what clinicians can already do, rather than simply matching it.

Social determinants of health (eg, socioeconomic position, level of education, and employment status) are important factors that should be considered in the development of AI models to reduce the risk that such models may perpetuate and even amplify health inequalities. Many large datasets commonly used for the development of AI models do not consider who and how patients are represented in the data in terms of diversity and incluisvity, with probable consequences for the generalisability of model performance.

For more on **PROMIS** see https://www.healthmeasures.net/explore-measurement-systems/promis

For more on **standards for data diversity** see https://www.datadiversity.org/home

PROs as an outcome measure to evaluate an AI intervention compared with standard care

Use of PROs as outcome measures to evaluate AI interventions is crucial for a better understanding of the effects that AI interventions can have on an individual's quality of life or symptoms, or both. Although the primary purpose of an AI intervention might be to

improve the workflow and support clinicians in decision making, it is important to evaluate whether such an intervention has downstream effects that benefit—or at least do not cause harm—to the lived experience of individual patients. For instance, the MAPREHAB trial (NCT04289025) aims to measure participants' change in gait kinematics and quality of life, as measured with the Oxford Hip Score[22] or Oxford Knee Score,[23] and the EQ-5D,[24] after hip or knee surgery. Participants in the AI-intervention group are given a personalised exercise plan that addresses their gait deficiency, as measured and guided by the GaitSmart device and associated analytics.[25] Participants in the control group will also be tested for gait deficiencies, but will not receive personalised exercises. The inclusion of PROs has been reported to increase satisfaction among users, and reduce the risk of falls and overall frailty.[25]

Outside of the AI research, there are numerous examples of a new intervention being advantageous in one domain (eg, improved survival or more cost-effective) but unacceptable to patients due to poor quality of life, or frequent or severe side-effects, or both.[26] These issues can be identified at an early stage if PROs are considered in the trial's design as a primary or secondary outcome measure. Without formal assessment of PROs, it is possible that these harms would only be recognised at a late stage in the intervention's development, as individual patients respond by taking themselves off the prescribed medication, or practitioners recognise patterns of symptoms and avoid prescribing the medication.

Challenges and recommendations of integrating PROs into AI systems

Notwithstanding the benefits of PROs in the context of AI, there are some specific challenges to consider when integrating PROs into AI systems. First, PRO data collection is fragmented and might not be representative of the diverse target population in which the AI intervention will be delivered. This under-representation of certain groups of people within the datasets used for the training and testing of algorithms can lead to those algorithms having substandard performance (or at least inadequately validated performance) for those groups. This is a form of health data poverty: "the inability for individuals, groups, or populations to benefit from a discovery or innovation due to a scarcity of data that are adequately representative".[27]

Second, a major limitation of many AI systems is that these systems are trained and validated against human performance (ie, what the clinician thought), rather than outcome data (ie, what happened to the patient).[28] This approach to AI systems development risks perpetuating a paternalistic medical model that traditionally privileges objective clinical measurement (ie, clinician observation, diagnostic tests, and imaging) over the patient perspective. We see this repeatedly in the AI for health literature, with the emphasis on clinician-level performance or expert-level accuracy. There is also risk in an uncritical acceptance of the clinician as the provider of ground truth. This idea could lead to algorithms that reflect or even amplify clinician biases, perpetuating inequalities that might not have occurred if the ground truth was derived from (or at least partially derived from) the assessment from the people with symptoms, themselves.

Third, the training of an AI system generally requires large-scale data, something that is available for few PROs and for a small number of specific indications. Moving to collect PROs as part of standard care across the breadth of medical conditions should be a priority, both in supporting better care for the individual, but also helping to ensure that future algorithms that will support or drive care in that condition, and incorporate patient experience (eg, symptom burden or quality of life).

Fourth, the selection of PROs requires careful consideration. As with other outcome measures, there is a need for the tools selected to be of sufficient quality, accurately and sensitively reflecting the state that they are seeking to measure in the target population. Generic PROMs allow for comparison across conditions, but vary in their sensitivity to detect clinically meaningful changes for specific conditions, and could therefore fail to detect change within the scope of a clinical trial for that disease. In contrast, disease-specific PROMs cannot be compared between diseases, but should be sensitive to changes within that disease.[29] The first step to successfully incorporate PROMs in AI systems is to consider the psychometric properties of the PROM of interest, and gather patients' and clinicians' input on the relevance of the PROM selected to assess the outcome of interest.

Fifth, the infrastructure for collecting PRO data can be considered a challenge. In terms of technical factors, there is now increasingly availability of ePRO data collection software with smartphone apps and web-based entry options that are facilitating PRO data collection.[30] In terms of human factors, there are residual concerns as to whether clinicians might be resistant to adopting PROMs, for example due to time constraints, lack of confidence in interpreting PRO data, or a lack of recognition of its value. However, it has been shown that clinicians are willing to implement PRO data collection after receiving training and understanding its importance.[31] Finally, the design of studies involving PRO or AI, or both, are often suboptimal.[32] The international reporting guidelines for trial protocols for studies involving AI (SPIRIT-AI)[33] and PROs (SPIRIT-PRO)[34] have highlighted additional elements of good design that are often neglected, but need consideration.

Inclusion of PROs in AI systems has the potential to improve people's health and quality of life by supporting health-care professionals to deliver better and faster patient-centred care. To successfully incorporate PROs in clinical practice, there is an ongoing need for multistakeholder collaboration between regulators, clinicians,

***Panel:* Considerations for the inclusion of patient-reported outcomes (PROs) in AI health-care technologies**

- At the design stage, actively explore with key stakeholders how PROs might be relevant as data input, output, or outcomes for the evaluations of the proposed AI system
- Ensure that datasets used for training and testing the AI system are appropriately inclusive such that they represent the diversity of the target population (including, but not limited to, age, sex, ethnicity, and socioeconomic status)
- Involve patients and members of the public in the co-design of PRO research and including selection of relevant of PRO measures (PROMs)
- Carefully select the PROMs for use, considering measurement properties, interpretation guidelines, patient acceptability, and symptom burden, and ideally selecting measures validated in the target population
- Assess the acceptability and applicability of the included PROMs to reduce and avoid symptom burden
- Report study designs openly and transparently using international standards for clinical trials of AI health interventions (SPIRIT-AI)[33] and patient-reported outcomes (SPIRIT-PRO)[34]

service providers, policy makers, PRO methodologists, patient partners, and technology companies to optimise how PROs are integrated into these AI systems within routine clinical workflows. Several considerations are proposed to address the challenges for the inclusion of PROs in AI health-care technologies (panel).

Conclusion

Careful consideration should be given to the inclusion of PROs within the function and evaluation of AI health care to avoid promoting survival at the expense of wellbeing and being processed at the expense of being heard; "what is the point of simply surviving - not living - if your existence is a painful, feeble and miserable one?"[35] Integrating PROs within AI and its evaluation would support the humanisation of AI for health and ensure that the patient's voice is not lost in a rush to digitise and automate health care.

Contributors

SCR, MJC, XL, AKD, and HD conceived the idea for the Viewpoint. SCR developed the first draft. MJC, XL, AKD, HD, SEH, and EM revised the manuscript. All authors read and approved the final manuscript.

Declaration of interests

SCR receives funding from UK SPINE and European Regional Development Fund DEMAND Hub, and has received support to attend the Organisation for European Cancer Institutes 2022 conference as an invited speaker. MJC is Director of the Birmingham Health Partners Centre for Regulatory Science and Innovation, Director of the Centre for Patient Reported Outcomes Research, and is a National Institute for Health and Care Research (NIHR) Senior Investigator. MJC receives funding from the NIHR, UK Research and Innovation (UKRI), NIHR Birmingham Biomedical Research Centre, the NIHR Surgical Reconstruction and Microbiology Research Centre, NIHR, Applied Research Collaboration (ARC) West Midlands, UK SPINE, Research England, European Regional Development Fund DEMAND Hub at the University of Birmingham and University Hospitals Birmingham NHS Foundation Trust, and the NIHR Birmingham–Oxford Blood and Transplant Research Unit in Precision Transplant and Cellular Therapeutics. MJC also receives funding from Health Data Research UK, Innovate UK (part of UKRI), Macmillan Cancer Support, UCB Pharma, Janssen, GSK, Gilead Sciences, European Commission, European Federation of Pharmaceutical Industries and Associations, and The Brain Tumor Charity. MJC has received personal fees from Aparito, CIS Oncology, Takeda Pharmaceuticals, Merck, Daiichi Sankyo, Glaukos, GSK, the Patient-Centered Outcomes Research Institute, Genentech, and Vertex Pharmaceuticals outside the submitted work. MJC has received lecture fees from the University of Maastricht, Maastricht, Netherlands. In addition, a family member owns shares in GSK. The views expressed in this Viewpoint are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. SEH receives funding from the NIHRARC West Midlands, UKRI, and declares personal fees from Aparito and Cochlear outside the submitted work. SEH has received honoraria from the Cognitive Hearing Science for Communication 2019 (CHSCOM2019) and Implantable Auditory Prostheses. In addition, she has received support to attend the CHSCOM2019 and British Academy of Audiology Conference. SEH is a member of the trial steering committee of the Both EARS Study. All other authors declare no competing interests.

Acknowledgments

The authors acknowledge the feedback provided by Marcus A Ong, Spectra Analytics.

References

1 Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019; **25:** 44–56.
2 Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019; **17:** 195.
3 US Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health. Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims. 2009.
4 Calvert M, Kyte D, Price G, Valderas JM, Hjollund NH. Maximising the impact of patient reported outcome assessment for patients and society. *BMJ* 2019; **364:** k5267.
5 Aiyegbusi OL, Roydhouse J, Rivera SC, et al. Key considerations to reduce or address respondent burden in patient-reported outcome (PRO) data collection. *Nat Commun* 2022; **13:** 6028
6 Hahn EA, Cella D, Chassany O, Fairclough DL, Wong GY, Hays RD. Precision of health-related quality-of-life data compared with other clinical measures. *Mayo Clin Proc* 2007; **82:** 1244–54.
7 Xiao C, Polomano R, Bruner DW. Comparison between patient-reported and clinician-observed symptoms in oncology. *Cancer Nurs* 2013; **36:** E1–16.
8 Basch E, Jia X, Heller G, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. *J Natl Cancer Inst* 2009; **101:** 1624–32.
9 Veitch ZW, Shepshelovich D, Gallagher C, et al. Underreporting of symptomatic adverse events in phase I clinical trials. *J Natl Cancer Inst* 2021; **113:** 980–88.
10 Verma D, Bach K, Mork PJ. Application of machine learning methods on patient reported outcome measurements for predicting outcomes: a literature review. *Informatics* 2021; **8:** 56.
11 McGlothlin AE, Lewis RJ. Minimal clinically important difference: defining what really matters to patients. *JAMA* 2014; **312:** 1342–43.
12 Cook CE. Clinimetrics corner: the minimal clinically important change score (MCID): a necessary pretense. *J Manual Manip Ther* 2008; **16:** E82–83.
13 Basch E, Deal AM, Kris MG, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. *J Clin Oncol* 2016; **34:** 557–65.
14 Basch E, Deal AM, Dueck AC, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer rreatment. *JAMA* 2017; **318:** 197–98.

15 NICE. Evidence standards framework (ESF) for digital health technologies, 2022 Nhttps://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies (accessed Aug 17, 2022).

16 Pearce FJ, Cruz Rivera S, Liu X, Manna E, Denniston AK, Calvert MJ. The role of patient-reported outcome measures in artificial intelligence health technologies: a systematic evaluation of ClinicalTrials.gov records (1997–2022). *Lancet Digit Health* 2023; **5:** e160–67.

17 Iivanainen S, Ekstrom J, Virtanen H, Kataja VV, Koivunen JP. Electronic patient-reported outcomes and machine learning in predicting immune-related adverse events of immune checkpoint inhibitor therapies. *BMC Med Inform Decis Mak* 2021; **21:** 205.

18 Virji AZ, Brennan CW, Skrabonja L, et al. Patients like you: how machine learning can be used as a shared decision-making tool to improve care. *NEJM Catal* 2021; **2.**

19 Lyman S, Lee Y-Y, Franklin PD, Li W, Cross MB, Padgett DE. Validation of the KOOS, JR: a short-form knee arthroplasty outcomes survey. *Clin Orthop Relat Res* 2016; **474:** 1461–71.

20 Delgado DA, Lambert BS, Boutris N, et al. Validation of digital visual analog scale pain scoring with a traditional paper-based visual analog scale in adults. *J Am Acad Orthop Surg Glob Res Rev* 2018; **2:** e088.

21 Pierson E, Cutler DM, Leskovec J, Mullainathan S, Obermeyer Z. An algorithmic approach to reducing unexplained pain disparities in underserved populations. *Nat Med* 2021; **27:** 136–40.

22 Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br* 1996; **78:** 185–90.

23 Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg Br* 1998; **80:** 63–69.

24 Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* 2011; **20:** 1727–36.

25 NICE. GaitSmart assessment and vGym personalised exercise rehabilitation programme for people with gait and mobility issues. Feb 01, 2022. https://www.nice.org.uk/advice/mib283 (accessed March 10, 2022).

26 Haskell H. Cumberlege review exposes stubborn and dangerous flaws in healthcare. *BMJ* 2020; **370:** m3099.

27 Ibrahim H, Liu X, Zariffa N, Morris AD, Denniston AK. Health data poverty: an assailable barrier to equitable digital health care. *Lancet Digit Health* 2021; **3:** e260–65.

28 Nagendran M, Chen Y, Lovejoy CA, et al. Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies. *BMJ* 2020; **368:** m689.

29 Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. *Med Care* 1989; **27** (suppl)**:** S217–32.

30 Greenhalgh J. The applications of PROs in clinical practice: what are they, do they work, and why? *Qual Life Res* 2009; **18:** 115–23.

31 Santana MJ, Haverman L, Absolom K, et al. Training clinicians in how to use patient-reported outcome measures in routine clinical practice. *Qual Life Res* 2015; **24:** 1707–18.

32 Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health* 2019; **1:** e271–97.

33 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:** 1351–63.

34 Calvert M, Kyte D, Mercieca-Bebber R, et al. Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: the spirit-pro extension. *JAMA* 2018; **319:** 483–94.

35 Schopen F. Cancer patients need more than survival. The Guardian, Oct 5, 2017. https://www.theguardian.com/commentisfree/2017/oct/05/cancer-patients-survival-miracle-cures-medical-research (accessed Feb 17, 2022).