SUPPLEMENTARY MATERIALS

Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI

Baptiste Vasey, MMed; Myura Nagendran, FFICM; Bruce Campbell, MS; David A. Clifton, DPhil; Gary S. Collins, PhD; Spiros Denaxas, PhD; Alastair K. Denniston, PhD; Livia Faes, MD; Bart Geerts, PhD; Mudathir Ibrahim, MD; Xiaoxuan Liu, PhD; Bilal A. Mateen, MBBS; Piyush Mathur, MD; Melissa D. McCradden, PhD; Lauren Morgan, PhD; Johan Ordish, MA; Campbell Rogers, MD; Suchi Saria, PhD; Daniel SW Ting, MD PhD; Peter Watkinson, MD; Wim Weber, MD PhD; Peter Wheatstone; Peter McCulloch, MD; and the DECIDE-AI expert group.

Table of Contents

SUPPLEMENTARY FIGURES **2**

Supplementary figure 1 – Item list(s) evolution (overview) 2
Supplementary figure 2 – Item list(s) evolution (per item) 3

SUPPLEMENTARY TABLES **4**

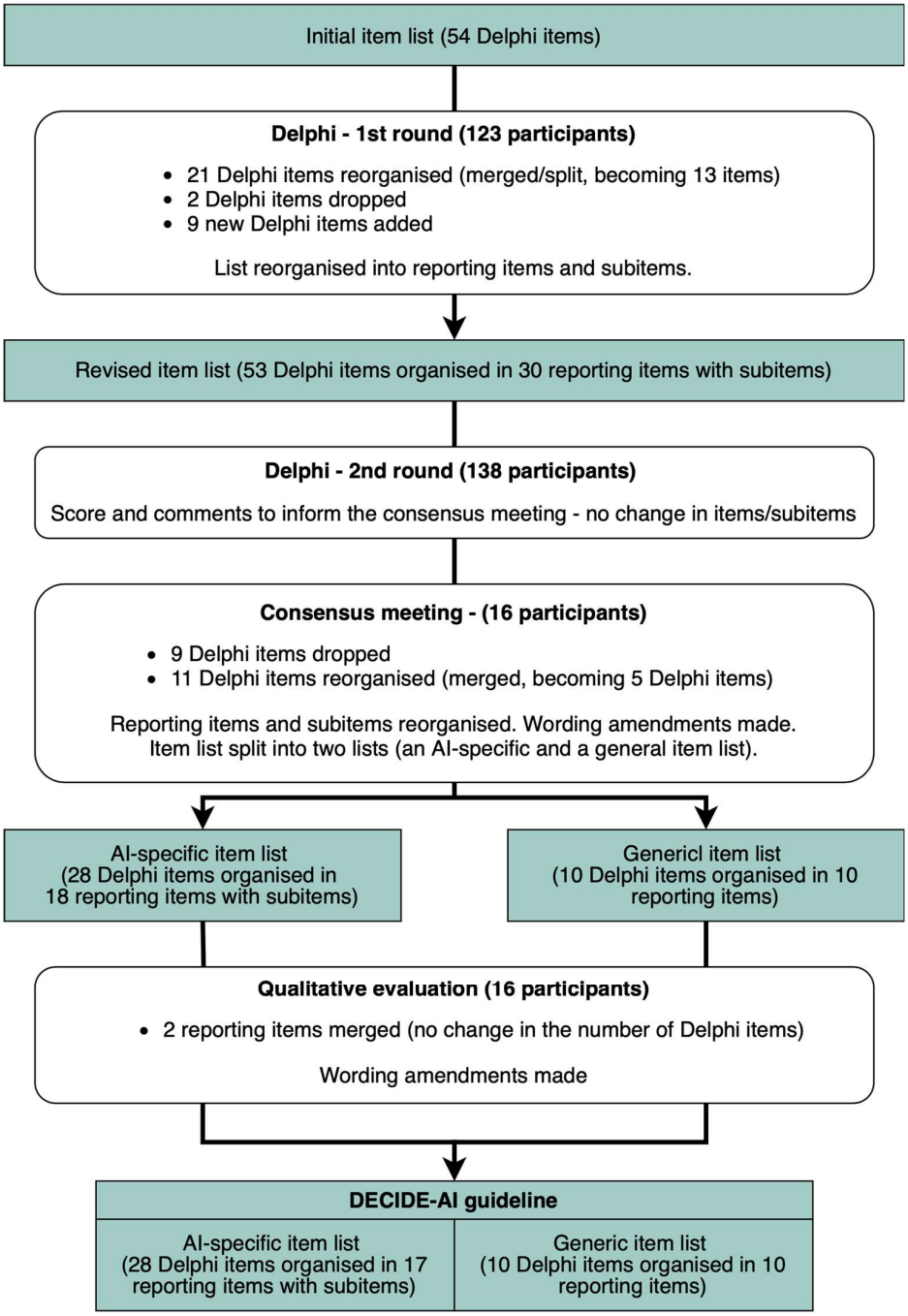
Suppl. table 1a – geographical distribution of the Delphi process participants 4
Suppl. table 1b – Stakeholder group affiliation of the Delphi process participants 4
Suppl. table 2a – geographical distribution of the Consensus Group members 5
Suppl. table 2b – Stakeholder group affiliation of the Consensus Group members 5
Suppl. table 3 – Summary of the consensus meeting votes 6

SUPPLEMENTARY NOTES **9**

Suppl. note 1 – Consensus process participants 9

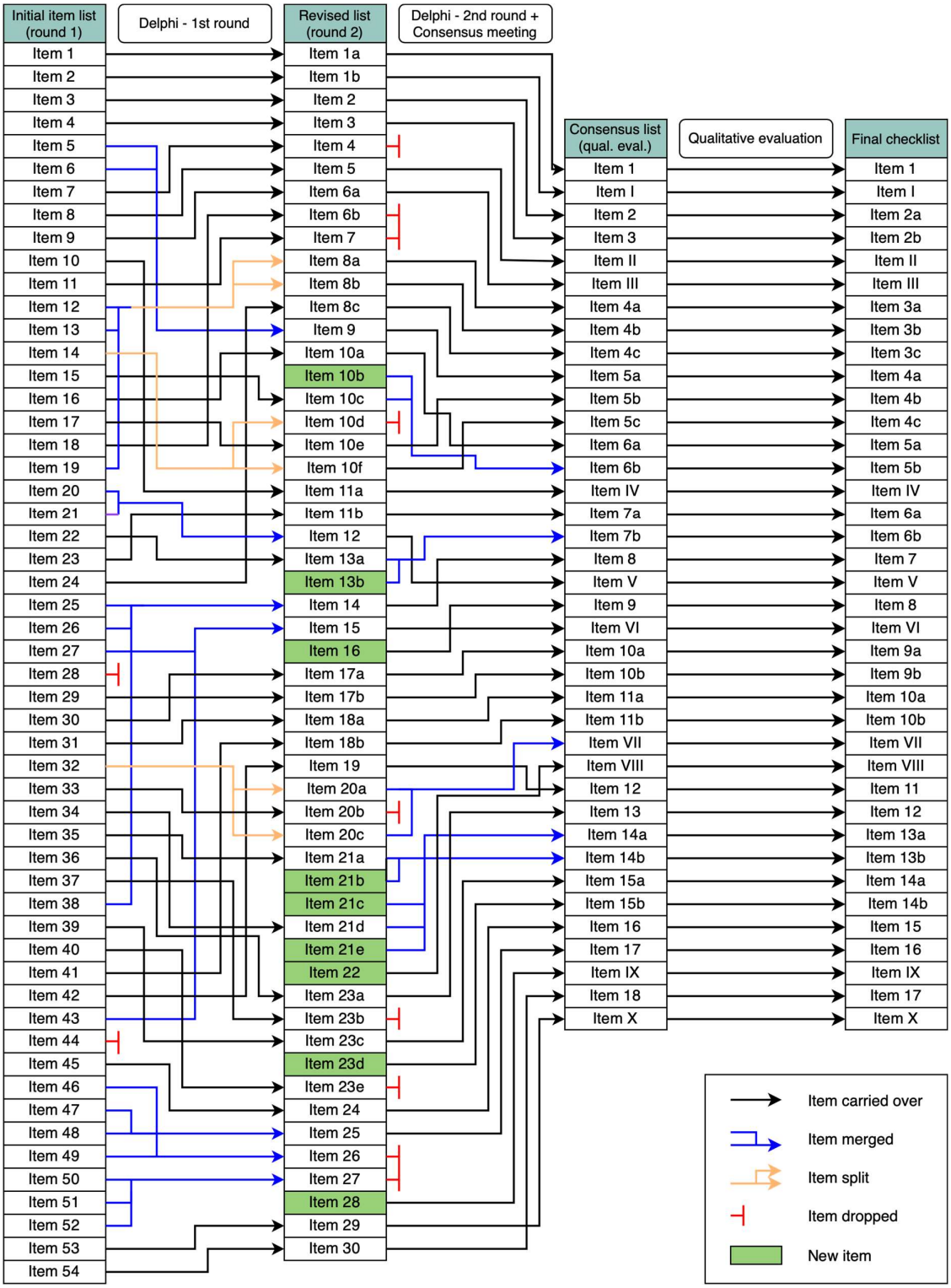
SUPPLEMENTARY FIGURES

Supplementary figure 1 – Item list(s) evolution (overview)



Suppl. figure 1. Graphical representation of the item list(s) evolution during the guideline development process. Delphi items are the recommendations voted on during the Delphi and discussed during the consensus meeting; reporting items are made of one or more Delphi items, organised thematically.

Supplementary figure 2 – Item list(s) evolution (per item)



Suppl. figure 2. Detail of the item list(s) evolution, with individual item follow up. For a better graphical representation, the AI-specific item list (Arab numerals) and generic item list (Roman numerals) are merged, as they appear in the checklist.

SUPPLEMENTARY TABLES

Suppl. table 1a – geographical distribution of the Delphi process participants

Geographical locations of participants' main working place					
Country	Number of participants	Country	Number of participants	Country	Number of participants
United Kingdom (UK)	79 (52%)	Australia (AU)	3 (2%)	Finland (FI)	1 (<1%)
United States of America (USA)	24 (16%)	France (FR)	3 (2%)	Kenya (KE)	1 (<1%)
The Netherlands (NL)	13 (9%)	Italy (IT)	3 (2%)	Portugal (PT)	1 (<1%)
Germany (DE)	7 (5%)	Austria (AT)	1 (<1%)	Singapore (SG)	1 (<1%)
Canada (CA)	6 (4%)	Belgium (BE)	1 (<1%)	South Africa (SA)	1 (<1%)
Republic of Korea (KR)	4 (3%)	Brazil (BR)	1 (<1%)	Spain (ES)	1 (<1%)

Suppl. table 1a: Geographical distribution of the Delphi participants' main working place. Total number of participants = 151.

Suppl. table 1b – Stakeholder group affiliation of the Delphi process participants

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

Suppl. table 1b: Self-reported stakeholder group affiliation of the Delphi participants. Total number of participants = 151, each participant could select multiple stakeholder group affiliation. *In total, 28 private sector entities of various sizes were represented.

Suppl. table 2a – geographical distribution of the Consensus Group members

Geographical locations of participants' main working place					
Country	Number of participants	Country	Number of participants	Country	Number of participants
United Kingdom (UK)	9 (56%)	The Netherlands (NL)	2 (13%)	Singapore (SG)	1 (6%)
United States of America (USA)	3 (19%)	Canada (CA)	1 (6%)		

Suppl. table 2a: Geographical distribution of the Consensus Group members' main working place. Total number of participants = 16.

Suppl. table 2b – Stakeholder group affiliation of the Consensus Group members

Stakeholder group affiliation			
Stakeholder group	Number of participants	Stakeholder group	Number of participants
Clinicians	8 (50%)	Policy makers/official institutions staff	2 (13%)
Engineers/Computer scientists	4 (25%)	Administrators/hospital management	2 (13%)
Methodologists	6 (38%)	Regulators	1 (6%)
Statisticians	2 (13%)	Trialists	1 (6%)
Implementation scientists	2 (13%)	Ethicists	1 (6%)
Entrepreneurs	3 (19%)	Patient representatives	1 (6%)
Epidemiologists	0 (0%)	Private sector representatives	1 (6%)
Human factors specialists	2 (13%)	Funders	2 (13%)
Journal editors	1 (6%)	Payers/Commissioners	0 (0%)
Allied health professional	0 (0%)	Psychologists	0 (0%)

Suppl. table 2b: Stakeholder group affiliation of the Consensus Group members. Total number of participants = 16, each member could be affiliated to more than one stakeholder group.

Suppl. table 3 – Summary of the consensus meeting votes

	Delphi	Consensus meeting						
	% participants	Votes			% votes[c]		Results	Comments/arguments of the Consensus Group
	Include[a]	include	exclude	blank[b]	include	exclude		
Item 1a	93	15	0	1	100	0	include	
Item 1b	98	13	0	3	100	0	include	
Item 2	96	14	0	2	100	0	include	
Item 3	96	14	0	2	100	0	include	
Item 4	72	9	6	1	60	40	exclude	There is for the time being no widely accepted development stage nomenclature for clinical AI systems. Although DECIDE-AI was developed to align with current regulatory processes and information about the regulatory context of the AI system would be useful, the overall feeling was to keep the focus of the guideline on scientific evaluation.
Item 5	99	12	2	2	86	14	include	
Item 6a	90	12	2	2	86	14	include	
Item 6b	69	1	14	1	7	93	exclude	Issues around patient privacy and data security should already be covered in the ethics application and approval.
Item 7	99	4	11	1	27	73	exclude	Although study design is important, it was felt that the wording of the item was too generic and the important aspects of the study design were covered in other items already.
Item 8a	93	13	0	3	100	0	include	
Item 8b	90	13	1	2	93	7	include	
Item 8c	85	13	0	3	100	0	include	
Item 9	90	13	0	3	100	0	include	
Item 10a	95	14	0	2	100	0	include	
Item 10b+c	91 & 93	15	0	1	100	0	include	
Item 10d	73	8	6	2	57	43	exclude	Details about the technical integration of an AI system is better suited to an instruction manual or audit report than a scientific report. Many research teams might not have integrated their AI system with the hospital infrastructure at the early stage of evaluation. If the developers intend to demonstrate some form of generalisability, the AI system evaluation should not be dependent on local IT integration.
Item 10e	96	14	1	1	93	7	include	
Item 10f	97	15	0	1	100	0	include	
Item 11a	96	14	0	2	100	0	include	
Item 11b	93	13	1	2	93	7	include	

Item 12	85	14	0	2	100	0	include	
Item 13a+b	90.4 & 84.2	14	0	2	100	0	include	
Item 14	76	11	2	3	85	15	include	Human factors are a core component of any new technology evaluation. A description of the human factors methods used is essential to understand the later reporting of the results.
Item 15	57	13	1	2	93	7	include	A wide range of medical journals and regulatory agencies are increasingly putting an emphasis on patient and public involvement. The overall feeling (including from the patient representative) was that including this item was important for ensuring that patients are recognised as important stakeholders in the future of clinical AI.
Item 16	62	12	1	3	92	8	include	The debate around algorithmic fairness and the ethical use of AI systems is now well-established in the literature. The information reported under this item is important to ensure a transparent appraisal of AI systems with regard to their ethical use of AI.
Item 17a	98	12	1	3	92	8	include	
Item 17b	96	13	0	3	100	0	include	
Item 18a	87	13	1	2	93	7	include	
Item 18b	86	11	2	3	85	15	include	
Item 19	89	13	0	3	100	0	include	
Item 20a+c	92 & 91	12	0	4	100	0	include	
Item 20b	74	3	9	4	25	75	exclude	Collecting information about the stand-alone algorithm performance is not always practical during clinical evaluation and may distract attention from the focus on the final supported user decisions, which actually influence clinical outcomes.
Item 21a+b	93 & 91	13	0	3	100	0	include	
Item 21c+d+e	83 & 97 & 85	13	0	3	100	0	include	
Item 22	78	12	1	3	92	8	include	
Item 23a	93	13	0	3	100	0	include	
Item 23b	55	3	10	3	23	77	exclude	Despite the importance of trust (and its evolution) to understand the reaction of users to the AI system outputs (and design large scale evaluation studies), there is currently no commonly accepted way to measure trust in the context of clinical AI.
Item 23c	77	14	0	2	100	0	include	
Item 23d	62	12	2	2	86	14	include	Learning curves are important in surgery and well documented in the surgical literature. There are many examples of hard to interpret trials because authors rushed to randomisation before learning how to perform a procedure to the required standard. AI is also a type of complex intervention for which the evaluation is likely to be impacted by similar considerations.

Item 23e	67	5	9	2	36	64	exclude	Despite the importance of interpretability to gain user and patient trust in the AI system, as well as to interpret the system's outputs in the context of the broader clinical information environment, there is currently no generally accepted way of quantifying or evaluating interpretability. Moreover, the clinical value of an AI system output may be independent of its interpretability.
Item 24	96	14	0	2	100	0	include	
Item 25	97	12	2	2	86	14	include	
Item 26	84	10	3	3	77	23	exclude	'Deviation' implies algorithm superiority, which is not always the case. The guideline should avoid being too prescriptive in the discussion section and focus on the discussion of the main results in the context of the intended use. Discussion of the human factors results should be integrated to those of the main clinical performance results.
Item 27	78	3	11	2	21	79	exclude	This could be out of scope and at risk of being speculative. Scale up is not necessarily the only logical next step; authors might be interested in generalising the results to other settings first. It is better to focus the discussion on the main outcomes and integrate the discussion of the remaining challenges to those of the main clinical performance results.
Item 28	96	12	2	2	86	14	include	
Item 29	98	13	1	2	93	7	include	
Item 30	93	14	0	2	100	0	include	

Suppl. table 3: Summary of the consensus meeting vote and results. [a]score ≥7 was defined as a recommendation to include; [b]including abstentions; [c]only considering non-blank votes.

SUPPLEMENTARY NOTES

Suppl. notes 1 – Consensus process participants

DECIDE-AI Steering Group (n=12)

Alastair K. Denniston (University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK); Baptiste Vasey (Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK); Bart Geerts (Healthpus.ai BV, Amsterdam, NL); David A. Clifton (Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK); Gary S. Collins (Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, University of Oxford, Oxford, UK); Lauren Morgan (Morgan Human Systems Ltd, Shrewsbury, UK); Livia Faes (Moorfields Eye Hospital, London, UK); Peter McCulloch (Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK); Peter Watkinson (Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK); Piyush Mathur (Cleveland Clinic, Cleveland, Ohio, USA); Suchi Saria (Departments of Computer Sciences, Statistics, and health Policy, and Division of Informatics, Johns Hopkins University, Baltimore, MD, USA); Xiaoxuan Liu (University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK);

Delphi participants (first and/or second round, n=151)

Aaron Y. Lee (Department of Ophthalmology, School of Medicine, University of Washington, Seattle, WA, USA); Alan G. Fraser (School of Medicine, Cardiff University, Cardiff, UK); Alastair K. Denniston (University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK); Ali Connell (Google Health, London, UK); Alykhan Vira (Quantium Health, Johannesburg, SA); Andre Esteva (Artera Research, Artera, Mountain View, CA, USA); Andrew D. Althouse (University of Pittsburgh, Pittsburgh, PA, USA); Andrew L. Beam (Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA); Anne de Hond (CAIRElab, Leiden University Medical Centre, Leiden, NL); Anne-Laure Boulesteix (Institute for Medical Information Processing, Biometry and Epidemiology, Ludwig Maximilian University, Munich, DE); Anthony Bradlow (Rheumatology Department, Royal Berkshire Hospital, Reading, UK); Ari Ercole (Cambridge Centre for AI in Medicine, University of Cambridge, Cambridge, UK); Arsenio Paez (Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK); Athanasios Tsanas (Usher Institute, Edinburgh Medical School, University of Edinburgh, Edinburgh, UK); Baptiste Vasey (Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK); Barry Kirby (K Sharp, Llanelli, UK); Bart Geerts (Healthpus.ai BV, Amsterdam, NL);

Ben Glocker (Department of Computing, Imperial College London, London, UK); Bilal A. Mateen (The Wellcome Trust, London, UK); Bruce Campbell (University of Exeter Medical School, Exeter, UK); Campbell Rogers (HeartFlow Inc., Redwood City, CA, USA); Carmelo Velardo (Sensyne Health, UK and Department of Engineering Science, University of Oxford, Oxford, UK); Chang Min Park (Seoul National University College of Medicine, Seoul, KR); Charisma Hehakaya (Division of Imaging & Oncology, University Medical Center Utrecht, Utrecht, NL); Chris Baber (University of Birmingham, Birmingham, UK); Chris Paton (Nuffield Department of Medicine, University of Oxford, Oxford, UK); Christian Johner (Johner Institute, Konstanz, DE); Christopher J. Kelly (Google Health, London, UK); Christopher J. Vincent (PDD Group Ltd, London, UK); Christopher Yau (University of Manchester, Manchester, UK); Clare McGenity (Pathology and Data Analytics, University of Leeds, Leeds, UK); Constantine Gatsonis (Department of Biostatistics, Brown University School of Public Health, Providence, RI, USA); Corinne Faivre-Finn (The Christie NHS Foundation Trust, Manchester, UK); Crispin Simon (London School of Economics, London, UK); Cyrus Espinoza (Patient representative); Daniel P. Jenkins (DCA Design, UK); Daniel S.W. Ting (Singapore National Eye Center, Singapore Eye Research Institute, Singapore, SG); Danielle Sent (Department of Medical Informatics, Amsterdam UMC, University of Amsterdam, Amsterdam, NL); Danilo Bzdok (Mila, Quebec Artificial Intelligence Institute, Montreal, CA); Darren Treanor (Leeds Teaching Hospitals NHS Trust, Leeds, UK); David A. Clifton (Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK); David C. Wong (Department of Computer Science and Centre for Health Informatics, University of Manchester, Manchester, UK); David F. Steiner (Google Health, Palo Alto, USA); David Higgins (Berlin Institute of Health, Berlin, DE); Dawn Benson (Patient representative); Deborah Morrison (National Institute for Health and Care Excellence, UK); Declan P. O'Regan (MRC London Institute of Medical Sciences, Imperial College London, London, UK); Dominic Danks (Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK); Emanuele Neri (University of Pisa, Pisa, IT); Evangelia Kyrimi (School of Electronic Engineering and Computer Science (EECS), Queen Mary University of London, London, UK); Falk Schwendicke (Charité Universitätsmedizin Berlin, Berlin, DE); Farah Magrabi (Australian Institute of Health Innovation, Macquarie University, Sydney, AU); Frances Ives (West Midlands Academic Health Science Network, Birmingham, UK); Frank E. Rademakers (Department Cardiovascular sciences, KU Leuven, Leuven, BE); Gary S. Collins (Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, University of Oxford, Oxford, UK); George E. Fowler (Bristol Centre for Surgical Research, Department of Population Health Sciences, Bristol Medical School, Bristol, UK); Giuseppe Frau (Deep Blue, Rome, IT); H. D. Jeffry Hogg (Population Health Science Institute, Newcastle University, Newcastle upon Tyne, UK); Hani J. Marcus (Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, Queen Square, London, UK); Heang-Ping Chan (Department of Radiology, University of Michigan, Ann Arbor, MI, USA); Henry Xiang (The

Abigail Wexner Research Institute, Nationwide Children's Hospital, The Ohio State University, Columbus, OH, USA); Hugh F. McIntyre (Department of Medicine, East Sussex Healthcare Trust, Hastings, UK); Hugh Harvey (Hardian Health, UK); Hyungjin Kim (Department of Radiology, Seoul National University Hospital, Seoul, KR); Ibrahim Habli (Department of Computer Science, University of York, York, UK); James C. Fackler (Department of Anesthesiology and Critical Care Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD, USA); James Shaw (Joint Centre for Bioethics, University of Toronto, Toronto, CA); Janet Higham (University of Oxford, Oxford, UK); Jared M. Wohlgemut (Centre for Trauma Sciences, Blizard Institute, Queen Mary University of London, London, UK); Jaron Chong (Medical Imaging, Western University, London, CA); Jean-Emmanuel Bibault (Radiation Oncology Department, Hôpital Européen Georges Pompidou, AP-HP, Paris, FR); Jérémie F. Cohen (Center of Research in Epidemiology and Statistics (Inserm 1153), Université de Paris, Paris, FR); Jesper Kers (Department of Pathology, Amsterdam UMC, University of Amsterdam, Amsterdam, NL); Jessica Morley (Oxford Internet Institute, University of Oxford, Oxford, UK); Joachim Krois (Oral Diagnostics & Digital Health & Health Services Research, Charité Universitätsmedizin Berlin, Berlin, Germany); Joao Monteiro (Nature Medicine, New York, NY, USA); Joel Horovitz (Department of Surgery, Maimonides Medical Center, Brooklyn, NY, USA); Johan Ordish (The Medicines and Healthcare products Regulatory Agency, London, UK); John Fletcher (The BMJ, London, UK); Jonathan Taylor (Nuclear Medicine / 3DLab, Sheffield Teaching Hospitals, Sheffield, UK); Jung Hyun Yoon (Department of Radiology, Severance Hospital, Yonsei University College of Medicine, Seoul, KR); Karandeep Singh (Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, MI, USA); Karel G.M. Moons (Julius Center, UMC Utrecht, Utrecht University, Utrecht, NL); Kassandra Karpathakis (Harvard TH Chan School of Public Health, Boston, MA, USA); Ken Catchpole (Medical University of South Carolina, Charleston, SC, USA); Kerenza Hood (Centre for Trials Research, Cardiff University, Cardiff, UK); Konstantinos Balaskas (Moorfields Ophthalmic Reading Centre and Clinical AI Hub, Moorfields Eye Hospital, London, UK); Konstantinos Kamnitsas (School of Computer Science, University of Birmingham, Birmingham, UK); Laura Militello (Applied Decision Science LLC, USA); Laure Wynants (Department of Epidemiology, CAPHRI Care and Public Health Research Institute, Maastricht University, Maastricht, NL); Lauren Morgan (Morgan Human Systems Ltd, Shrewsbury, UK); Livia Faes (Moorfields Eye Hospital, London, UK); Luc J.M. Smits (Department of Epidemiology, Maastricht University, Maastricht, NL); Ludwig C. Hinske (Institute for Biomedical Information Processing, Biometry and Epidemiology, Ludwig-Maximilians-University Munich, Munich, DE); Luke Oakden-Rayner (Australian Institute for Machine Learning, University of Adelaide, Adelaide, AU); M. Khair ElZarrad (U.S. Food and Drug Administration)[§]; Maarten van Smeden (Julius Center for Health Sciences and Primary Care,

[§] Participation represents personal views and perspectives that may not necessarily reflect the positions and opinions of the U.S. FDA.

University Medical Center Utrecht, Utrecht University, Utrecht, NL); Mara Giavina-Bianchi (Hospital Israelita Albert Einstein, São Paulo, BR); Mark Daley (The University of Western Ontario, London, CA); Mark P. Sendak (Duke Institute for Health Innovation, Durham, USA); Mark Sujan (Human Factors Everywhere Ltd, Woking, UK); Maroeska Rovers (Department of Operating rooms, Radboudumc, Nijmegen, NL); Matthew DeCamp (University of Colorado, Boulder, CO, USA); Matthieu Komorowski (Dept of Surgery and Cancer, Imperial College London, London, UK); Max Marsden (Centre for Trauma Science, Blizard Institute, Queen Mary University of London, London, UK); Maxine Mackintosh (Genomics England, Queen Mary University of London, London, United Kingdom); Melissa D. McCradden (The Hospital for Sick Children, Toronto, CA); Michael D. Abramoff (University of Iowa, Iowa City, IA, USA); Miguel Ángel Armengol de la Hoz (Big Data Department, FPS, Regional Ministry of Health of Southern Spain, ES); Myura Nagendran (UKRI Centre for Doctoral Training in AI for Healthcare, Imperial College London, London, UK); Neale Hambidge (National Hospital for Neurology and Neurosurgery, Queen Square, London, UK); Neil Daly (Skin Analytics, London, UK); Niels Peek (Division of Informatics, Imaging and Data Science, The University of Manchester, Manchester, UK); Oliver Redfern (Kadoorie Centre for Critical Care Research and Education, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK); Omer F. Ahmad (Wellcome/EPSRC centre for Interventional & Surgical Sciences (WEISS), University College London, London, UK); Patrick M. Bossuyt (Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, NL); Pearse A. Keane (Institute of Ophthalmology, University College London, London, UK); Pedro N.P. Ferreira (CENTEC - IST, University of Lisbon, Lisbon, PT); Peter McCulloch (Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK); Peter Watkinson (Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK); Peter Wheatstone (Patient representative); Petra Schnell-Inderst (Institut of Public Health, Medical Decision Making and HTA, UMIT - University for Health Sciences, Medical Informatics and Technology, Hall i. T., AT); Pietro Mascagni (Gastrointestinal Endoscopic Surgery, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, IT); Piyush Mathur (Cleveland Clinic, Cleveland, Ohio, USA); Prokar Dasgupta (King's Health Partners Academic Surgery, King's College London, London, UK); Pujun Guan (Graduate School of Biomedical Sciences, University of Texas MD Anderson Cancer Center and UTHealth, Houston, USA); Rawen Kader (Division of Surgery and Interventional Sciences, University College London, London, UK); Reena Chopra (Google Health, London, UK); Ritse M. Mann (Department of medical imaging, Radboud University Medical Center, Nijmegen, NL); Rupa Sarkar (The Lancet Digital Health, The Lancet Group, London, UK); Saana M. Mäenpää (Department of Neurosurgery, Helsinki University Hospital, Helsinki, FI); Samuel G. Finlayson (Harvard Medical School, Boston, MA, USA); Sarah Vollam (Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK); Sean P. White (NHS England, UK); Sebastian J. Vollmer (Data Science and its Application, DFKI, Kaiserslautern, DE); Seong Ho Park (Department of Radiology,

Asan Medical Center, Seoul, KR); Shakir Laher (University of York, York, UK); Shalmali Joshi (SEAS, Harvard University, Cambridge, MA, USA); Spiros Denaxas (Institute of Health Informatics, University College London, London, UK); Suchi Saria (Departments of Computer Sciences, Statistics, and health Policy, and Division of Informatics, Johns Hopkins University, Baltimore, MD, USA); Susan C. Shelmerdine (Department of Clinical Radiology, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK); Tom J.W. Stocker (PUBLIC ltd, Oxford, UK); Valentina Giannini (University of Turin, Turin, IT); Valerie Keston-Hole (Patient representative); Vince I. Madai (QUEST Centre for Responsible Research, Berlin Institute of Health, Charité Universitätsmedizin Berlin, Berlin, Germany); Virginia Newcombe (University Division of Anaesthesia, Department of Medicine, University of Cambridge, Cambridge, UK); Wendy A. Rogers (Philosophy Department and School of Medicine, Macquarie University, Sydney, AU); William Ogallo (IBM Research Africa, Nairobi, KE); Wim Weber (The BMJ, London, UK); Xiaoxuan Liu (University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK); Zane B. Perkins (Centre for Trauma Sciences, Queen Mary University of London, London, UK).

Consensus meeting participants (n=16+3)

With voting rights (n=16):
Baptiste Vasey (Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK); Bart Geerts (Healthpus.ai BV, Amsterdam, NL); Bilal A. Mateen (The Wellcome Trust, London, UK); Campbell Rogers (HeartFlow Inc., Redwood City, CA, USA); Daniel S.W. Ting (Singapore National Eye Center, Singapore Eye Research Institute, Singapore, SG); Gary S. Collins (Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, University of Oxford, Oxford, UK); Johan Ordish (The Medicines and Healthcare products Regulatory Agency, London, UK); Lauren Morgan (Morgan Human Systems Ltd, Shrewsbury, UK); Melissa D. McCradden (The Hospital for Sick Children, Toronto, CA); Peter McCulloch (Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK); Peter Wheatstone (Patient representative); Piyush Mathur (Cleveland Clinic, Cleveland, Ohio, USA); Spiros Denaxas (Institute of Health Informatics, University College London, London, UK); Suchi Saria (Departments of Computer Sciences, Statistics, and health Policy, and Division of Informatics, Johns Hopkins University, Baltimore, MD, USA); Wim Weber (The BMJ, London, UK); Xiaoxuan Liu (University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK);

Chair (n=1):
Bruce Campbell (University of Exeter Medical School, Exeter, UK)

Observers (n=2):
Myura Nagendran (Imperial College London, London, UK), Livia Faes (Moorfields Eye Hospital, London, UK)

Experts participating in the qualitative evaluation (n=16)

Ali Connell (Google Health, London, UK); Dinesh V. Gunasekaran (Singapore National Eye Center, Singapore Eye Research Institute, Singapore, SG); Falk Schwendicke (Charité Universitätsmedizin Berlin, Berlin, DE); Hani J. Marcus (Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, Queen Square, London, UK); Jean-Emmanuel Bibault (Radiation Oncology Department, Hôpital Européen Georges Pompidou, AP-HP, Paris, FR); Laurence B. Lovat (University College London, London, UK); Lisa Baker (Mima, London, UK); Mara Giavina-Bianchi (Hospital Israelita Albert Einstein, São Paulo, BR); Matthew Woodward (THIS Institute, School of Clinical Medicine, University of Cambridge, Cambridge, UK); Oliver Redfern (Kadoorie Centre for Critical Care Research and Education, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK); Omer F. Ahmad (Wellcome/EPSRC centre for Interventional & Surgical Sciences (WEISS), University College London, London, UK); Rachel Barnett (Healthplus.ai B.V., Amsterdam, NL); Siri L. van der Meijden (Healthplus.ai B.V., Amsterdam, NL and Leiden University Medical Center, Leiden, NL); Tien-En Tan (Singapore National Eye Center, Singapore Eye Research Institute, Singapore, SG); Wei Yan Ng (Singapore National Eye Center, Singapore Eye Research Institute, Singapore, SG); Yoonyoung Park (Center for Computational Health, IBM Research, Cambridge, MA USA).