

IDEAL approach to the evaluation of machine learning technology in epilepsy surgery: protocol for the MAST trial

Aswin Chari ^{1,2}, Sophie Adler,^{1,2} Konrad Wagstyl,³ Kiran Seunarine,^{1,2} Hani Marcus,^{4,5} Torsten Baldeweg,² Martin Tisdall^{1,2}

To cite: Chari A, Adler S, Wagstyl K, *et al.* IDEAL approach to the evaluation of machine learning technology in epilepsy surgery: protocol for the MAST trial. *BMJ Surg Interv Health Technologies* 2022;**4**:e000109. doi:10.1136/bmjisit-2021-000109

Received 10 August 2021
Accepted 10 January 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of Neurosurgery, Great Ormond Street Hospital, London, UK

²Developmental Neuroscience, Great Ormond Street Institute of Child Health, University College London, London, UK

³Wellcome Centre for Human Neuroimaging, University College London, London, UK

⁴Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, London, UK

⁵Wellcome EPSRC Centre for Interventional and Surgical Sciences, University College London, London, UK

Correspondence to

Dr Aswin Chari;
aswin.chari.18@ucl.ac.uk

ABSTRACT

Epilepsy and epilepsy surgery lend themselves well to the application of machine learning (ML) and artificial intelligence (AI) technologies. This is evidenced by the plethora of tools developed for applications such as seizure detection and analysis of imaging and electrophysiological data. However, few of these tools have been directly used to guide patient management. In recent years, the Idea, Development, Exploration, Assessment, Long-Term Follow-Up (IDEAL) collaboration has formalised stages for the evaluation of surgical innovation and medical devices, and, in many ways, this pragmatic framework is also applicable to ML/AI technology, balancing innovation and safety. In this protocol paper, we outline the preclinical (IDEAL stage 0) evaluation and the protocol for a prospective (IDEAL stage 1/2a) study to evaluate the utility of an ML lesion detection algorithm designed to detect focal cortical dysplasia from structural MRI, as an adjunct in the planning of stereoelectroencephalography trajectories in children undergoing intracranial evaluation for drug-resistant epilepsy.

INTRODUCTION

It is clear that the fields of epilepsy and epilepsy surgery lend themselves ideally to the application of machine learning (ML) and artificial intelligence (AI) technologies to directly benefit patient care in areas such as: seizure detection; analysis of clinical, electrophysiological and imaging data; seizure localisation and prediction of medical and surgical treatment outcomes.^{1 2} Although many of these applications have already been explored, as yet, there have been few reports of ML algorithms directly being used to guide management decisions in patients.^{1 3-5} The difficulty in proceeding from the development of algorithms to prospective evaluation in clinical care is attributable to issues surrounding transparency, reproducibility, ethics and effectiveness and changes required in traditional clinical trial designs.⁶⁻⁸ In recent years, the Idea, Development, Exploration,

Assessment, Long-Term Follow-Up (IDEAL) collaboration has formalised stages for the evaluation of surgical innovation and medical devices, and, in many ways, this pragmatic framework is also applicable to ML/AI technology, balancing innovation and safety (figure 1).⁹⁻¹² This has been recently supplemented with a stage 0 that centres around a proportionate preclinical evaluation that balances safety while facilitating innovative first-in-human studies.¹²

In this paper, we present the IDEAL stage 0 evaluation and protocol for a combined IDEAL stage 1/2a study for a lesion detection algorithm designed to detect focal cortical dysplasia as part of the planning of stereoelectroencephalography (SEEG) electrode trajectories in patients with drug-resistant epilepsy undergoing invasive presurgical evaluation.^{13 14}

Algorithm background

The multicentre epilepsy lesion detection (MELD) algorithm was developed to detect focal cortical dysplasia using volumetric T1-weighted and fluid-attenuated inversion recovery (FLAIR) MRI sequences. Its utility is in aiding in the presurgical evaluation of drug-resistant epilepsy - to localise MRI lesions that may be responsible for epilepsy that may be amenable to surgical resection.

Its development and retrospective evaluation have been previously presented.^{13 14} Cortical surface-based features such as cortical thickness, grey-white matter interface contrast and FLAIR intensity from confirmed lesions were used to train a neural network classifier. The classifier was initially validated using a leave-one-out cross-validation approach but has since been validated on external cohorts with high sensitivity (73.7%)

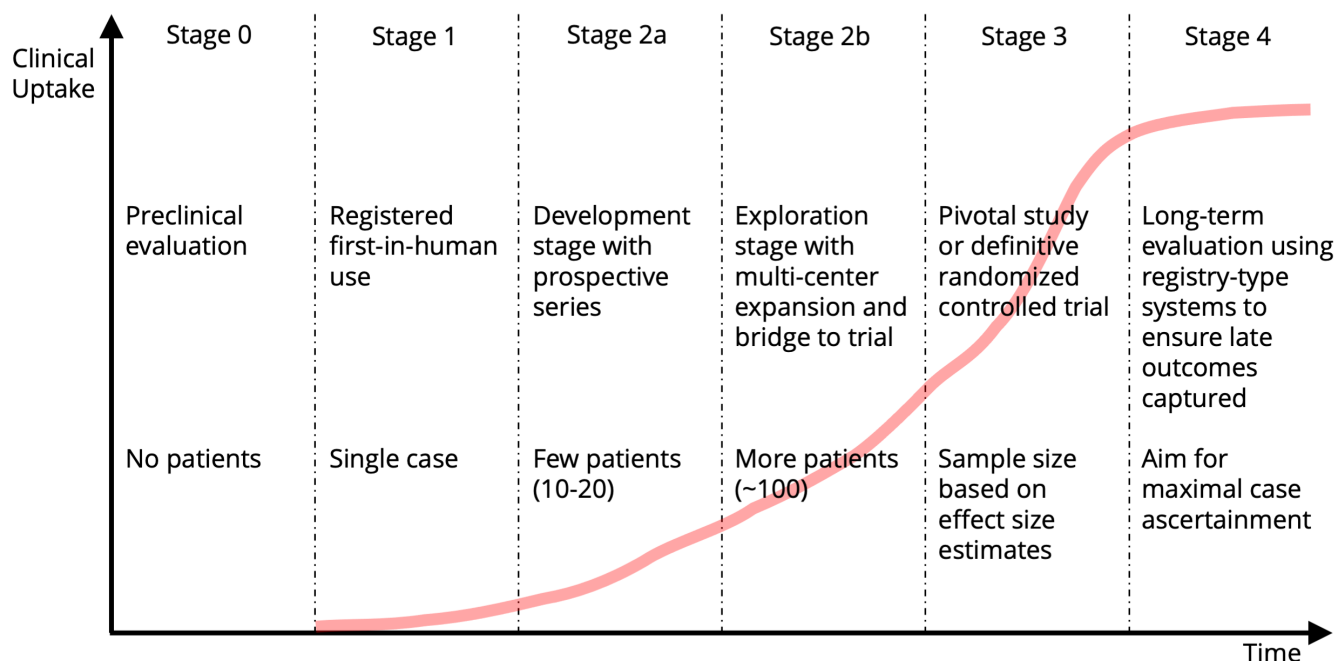


Figure 1 Schematic of the IDEAL stages of evaluation for surgical/medical device innovation. IDEA, Idea, Development, Exploration, Assessment, Long-Term Follow-Up.

specificity (90.0%) and an area under the ROC curve of 0.75.^{14 15}

IDEAL stage 0 evaluation

The IDEAL-D stage 0 evaluation involves classifying the device, assessing the risks using a failure modes and effect analyses (FMEA) and conducting an evaluation proportionate to the potential risks involved.¹² Software is classified according to the associated device. Therefore, in the context of being used to plan SEEG trajectories, the MELD algorithm can be viewed as associated with a non-absorbable implant, putting it in tier 3 of the device categorisation. An FMEA shows that the main risk is the additional electrodes causing bleeding, which would be classified as occasional frequency (<1/100) and potentially of serious severity (could result in injury or impairment).¹⁶ Proportionately, the preclinical evaluation includes device, clinician and patient perspective studies, with a systems perspective gleaned from the existing literature. The cost of inserting additional electrodes is minimal compared with the cost of the SEEG procedure as a whole and within the normal clinical variation and, therefore, health-economic analyses are not required at this stage (table 1).¹²

Protocol for IDEAL stage 1/2a evaluation

The aim of the MELD as an Adjunct for SEEG Trajectories (MAST) Trial, a single-arm single-centre prospective pilot (IDEAL stage 1/2a) study, is to assess whether MELD is a helpful adjunct in the planning of SEEG electrode trajectories. Patients undergoing SEEG at our centre will be eligible for inclusion. Exclusion criteria include tuberous sclerosis or other large structural abnormality

and prior resective surgery. The pilot sample size is 20, chosen based on the principles of the IDEAL stage 2a.⁹ Fully informed consent will be required for participation, either from the parent or the child, in line with UK law. The patient pathway is shown in figure 2 and summarised below.

Following routine planning of the SEEG electrode trajectories by the multidisciplinary team (MDT) using the non-invasive evaluation (clinical semiology, EEG videotelemetry and MRI±other adjuncts, figure 3A), the MRI scans, which are collected as part of the routine clinical care, will be run through the MELD algorithm. The putative lesion clusters identified by the algorithm will be reviewed by an expert neuroradiologist with expertise in presurgical evaluation of epilepsy to ensure that they are not driven by artefact. Once artefacts have been excluded, the top three putative lesion clusters will then be identified (on the ipsilateral side only if the planned implantation is unilateral). If these clusters are not already sampled by existing electrodes, up to three extra electrodes may be added to sample from each of those top three lesion clusters (figure 3B).

The primary objective is to assess the proportion of patients that had additional electrode (ie, extra electrodes implanted into unsampled identified clusters) contacts in the clinically defined seizure onset zone (SOZ). Given the small sample size, no statistical testing will be conducted, and the results will be descriptive statistics only. For each patient, we will also assess a number of secondary objectives, including blinded neurophysiological and neurosurgical assessments of whether or not the data from the additional electrodes would have affected

Table 1 Preclinical evaluation of the MELD algorithm, classified by study type according to the IDEAL stage 0 framework

| Study classification | Evidence |
|-----------------------|---|
| Device perspective | Safety and efficacy have been formally assessed in the peer-reviewed literature. ^{13–15} From the perspective of this difficult-to-localise cohort of patients undergoing SEEG, the sensitivity of the classifier was 74% with 100% specificity (no lesions detected in controls). In 3 of these 34 patients, a focal SOZ was not identified and the classifier identified additional lesions that were not implanted. ¹³ Safety relates to identification of false positive clusters. The risk profile has been limited by sampling a maximum of three clusters per patient. |
| System perspective | There is clearly an established need for such algorithms as evidenced by the increasing evaluation of ‘MRI negative’ cases for epilepsy surgery, especially through SEEG. Not all patients undergoing epilepsy surgery have a SOZ identified and, even in those who do, a third do not achieve seizure freedom following resective surgery. ²⁰ |
| Patient perspective | As part of the MAST Trial planning, a group of parents and children with epilepsy were surveyed at to assess whether the addition of up to three extra electrodes would be acceptable. Of 15 respondents (14 parents and one adolescent with epilepsy), all 15 (100%) agreed that the risk would be acceptable when balanced against the potential benefits and all 15 (100%) would enrol in the trial were they/their child to undergo SEEG implantation at our centre. |
| Clinician perspective | Clinicians from multiple specialties (neurosurgery, neurology, neurophysiology, neuroradiology) were involved in the initial development and retrospective evaluation studies, that showed promising performance. Specifically, these same clinicians will be involved in this prospective trial, attesting to the clinical acceptability and utility. The main users of the output will be neurosurgeons and the integration of the output into the planning software has been tested (figure 2B), although more formal learning curves will need to be assessed prior to multicentre studies. |

IDEA, Idea, Development, Exploration, Assessment, Long-Term Follow-Up.; MELD, multicentre epilepsy lesion detection; SEEG, stereoelectroencephalography; SOZ, Seizure Onset Zone.

subsequent management decisions (table 2). To gain an understanding of the situations in which the algorithm might be particularly helpful, we will also collect data on the indication for SEEG and the MDT members’ confidence in the ability of the SEEG implantation to identify a seizure onset zone. A transparent account of harms, including the number of electrodes implanted into ‘false positive’ lesions, will be reported, allowing a balanced assessment of risks and benefits prior to consideration of further stage testing. Steps taken to optimise the workflow will also be reported.

The risks and benefits of such an approach were carefully considered during the design phase. In order to

evaluate the incremental benefit of the algorithm, the clinical standard of care could not be altered and, therefore, moving existing electrodes was not considered an option. The risk of complications per additional electrode is small.¹⁶ In addition, a group of 14 parents of children with epilepsy and one adolescent with epilepsy were surveyed at the Epilepsy in Childhood: Carers Uniting with Researchers Information Day. All 15 (100%) agreed that the risk would be acceptable when balanced against the potential benefits, and all 15 (100%) would enrol in the trial were they/their child to undergo SEEG implantation.

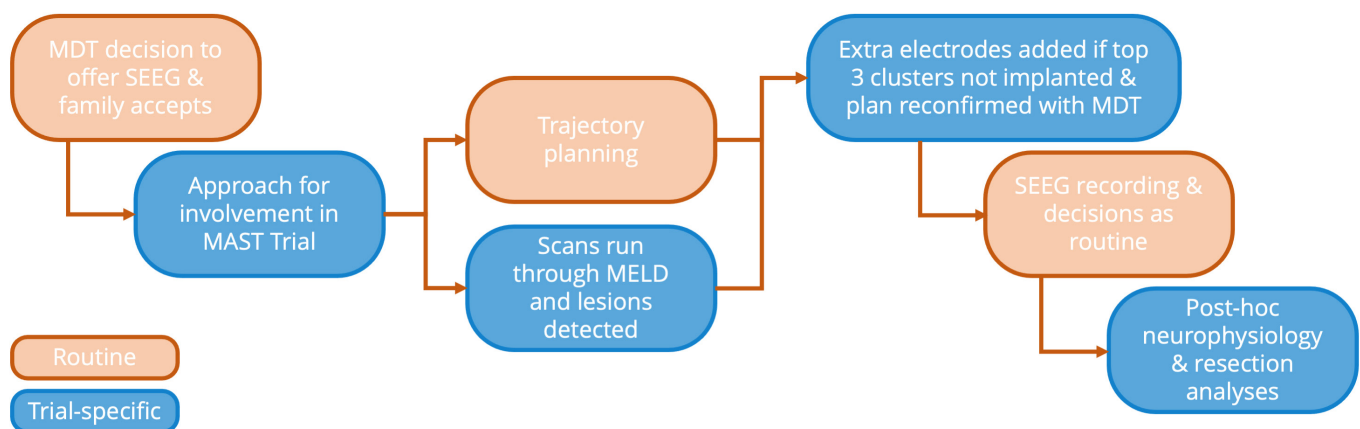


Figure 2 Study flowchart in the MAST trial, with routine clinical care pathway (orange) and trial-specific (blue) elements shown. MDT, multidisciplinary team; MELD, multicentre epilepsy lesion detection; SEEG, stereoelectroencephalography; MAST, MELD as an Adjunct for SEEG Trajectories.

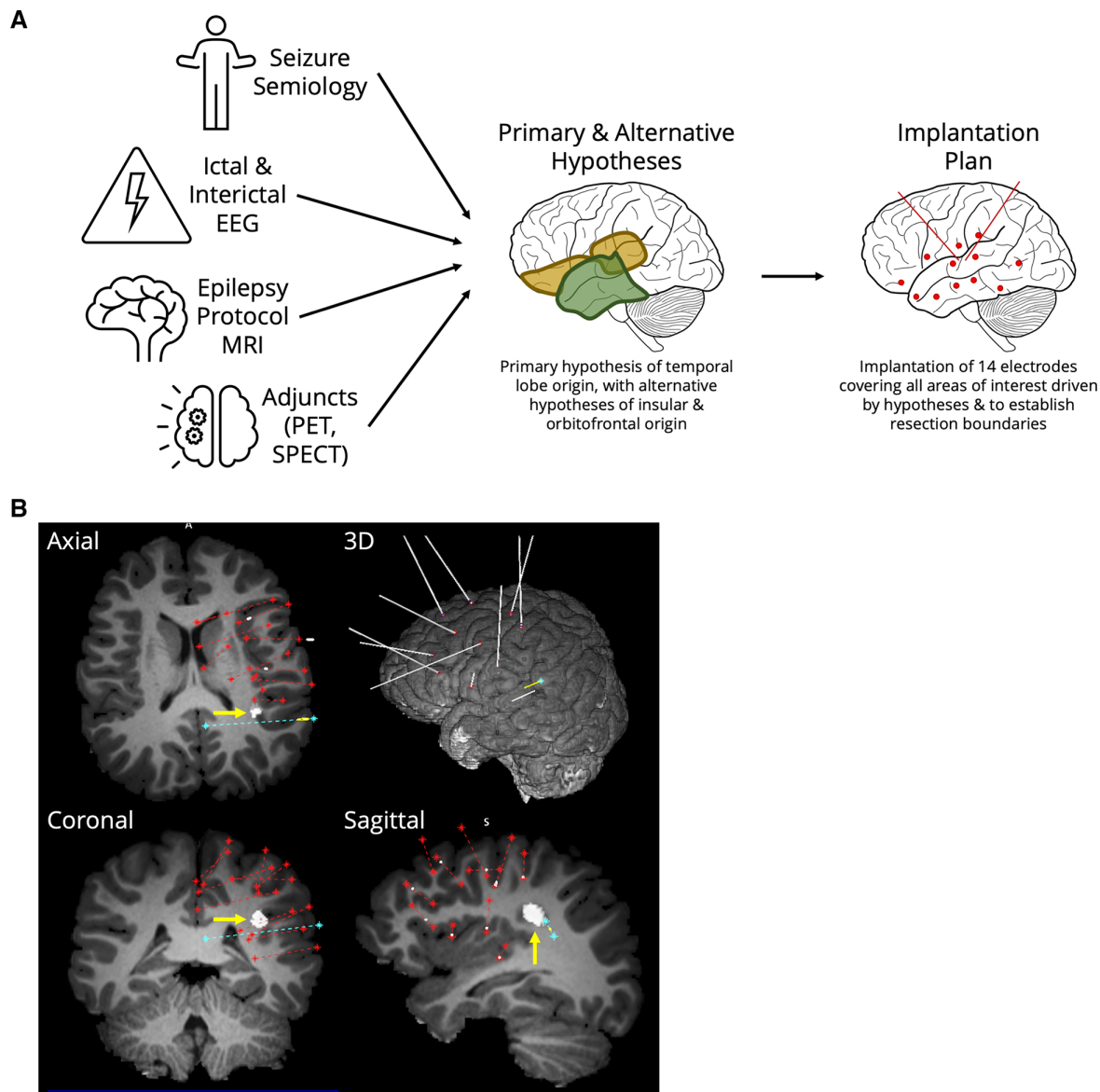


Figure 3 (A) Schematic of current clinical planning of electrode locations using the non-invasive evaluation by the MDT. The data from the pre-surgical evaluation are used to generate hypotheses of the location of the SOZ. The primary hypothesis in this example is the temporal lobe (green) with alternative areas involving orbitofrontal and insular cortices (yellow). Electrode trajectories are then planned to sample these areas of interest and establish important resection boundaries, with red dots and lines showing orthogonal and parallel electrodes respectively. (B) Example of planned electrodes on sagittal, axial, coronal and 3D views of a T1-weighted MRI scan. The MELD-identified lesion (grey, yellow arrow) is not sampled as the nearest electrode (yellow on the 3D panel) traverses postero-inferior to it. In this example, an extra electrode would, therefore, be added to sample the identified lesion in the depths of the supramarginal gyrus. EEG, electroencephalography; MDT, multidisciplinary team; MELD, multicentre epilepsy lesion detection; SOZ, seizure onset zone; PET, positron emission tomography; SPECT, single-photon emission computed tomography.

Ethical approval has been received for this study from the UK Health Research Authority (IRAS ID 275480) and the study began recruitment in September 2020. Further details of the protocol are available at <https://clinicaltrials.gov/ct2/show/NCT04383028>. The study has been funded by EANS-Stryker Research Grant 2020 & The Rosetrees Trust (A2665), but the funders had no input into the design or conduct of the study. The study will be reported in according with relevant guidelines including IDEAL stage 1/2a, Transparent reporting of a multivariable prediction model for individual prognosis

or diagnosis (TRIPOD) & Standards for reporting diagnostic accuracy (STARD) guidelines.^{11 17 18}

IDEAL stage 2b and beyond

We see the MELD algorithm as an additional part of the presurgical evaluation, which should be available to clinicians before the SEEG electrode trajectories are planned. The current algorithm was developed using single-centre data, with the clinical study being conducted at the same centre. The first step to further generalise the applicability would be to construct a large multicentre data set

Table 2 Primary and secondary objectives of the MAST trial

| Objectives | Outcome measures/endpoints |
|----------------------|--|
| Primary objective | Assess the proportion of patients that had additional electrode contacts implanted in the SEEG-defined seizure onset zone. |
| Secondary objectives | <p>For each case, we will assess:</p> <ul style="list-style-type: none"> ▶ Pre-implantation confidence of the MDT members in identifying a seizure onset zone (prior to MELD information) as a measure of the ‘difficulty’ of the SEEG exploration ▶ Number of electrodes added ▶ Number of electrodes already in identified lesions ▶ Was a MELD-identified cluster part of the SOZ (and if so, how many clusters?) ▶ Would the SOZ have been identified without MELD? ▶ Blinded neurophysiological assessment of the SOZ contacts with and without additional electrodes ▶ Putative resection boundaries with and without the additional electrodes, to be modelled by a neurosurgeon ▶ Safety of adding additional electrodes |

MDT, multidisciplinary team; MELD, multicentre epilepsy lesion detection; SEEG, stereoelectroencephalography; SOZ, Seizure Onset Zone.

that accounts for variability in subjects, scan strength and protocols across centres. This is currently being constructed as part of the MELD study.¹⁹

If this should progress beyond the current stage 1/2a study, we envision conducting a stage 2b study as a feasibility randomised trial in a small number of centres, which assesses the proportion of patients in whom a SOZ is identified on SEEG with and without the benefit of the algorithm. This would allow us to optimise stability of the algorithm across centres, train other surgeons to integrate it into their clinical workflow, establish estimated effect sizes and assess feasibility of conducting a larger randomised study across multiple centres as an IDEAL stage 3 study.

Analysis of the stage 2b or 3 studies may also result in the additional findings, such as a select group of patients (eg, with interictal positron emission tomography (PET), ictal single-photo emission computed tomography (SPECT) or magnetoencephalography (MEG) concordance with the top identified lesions) in whom it may be debated that SEEG is not necessary prior to resection. If such findings were identified, the subsequent prospective clinical evaluation would have to go through the rigorous process of evaluation from stage 0 again as it would be using the algorithm in a different context to the present evaluation. However, many aspects may overlap with the current evaluation, making the process more streamlined.

CONCLUSION

ML technologies have much to offer to clinicians and patients in the field of neurosurgery and safe and robust evaluation is crucial to realising these opportunities. In this paper, we have outlined the IDEAL stage 0 evaluation, a protocol for an IDEAL stage 1/2a clinical trial and future directions for the use of the MELD algorithm in assisting the planning of SEEG trajectories. The robust framework balances safety and innovation, ultimately benefitting patients by improving outcomes.

Twitter Aswin Chari @aswinchari

Acknowledgements The authors would like to acknowledge the SEEG team at Great Ormond Street Hospital for embracing the study and integrating it into the clinical workflow.

Contributors AC, SA, KW, TB and MT conceived the study and developed the protocol. KS and HM provided methodological input to optimise study design. All authors were involved in the drafting and editing of the manuscript and have reviewed and approved the final version.

Funding This project is supported by the EANS-Stryker Research Grant 2020 and The Rosetrees Trust (A2665). AC is supported by a Great Ormond Street Hospital (GOSH) Children’s Charity Surgeon Scientist Fellowship. This work has been supported by the GOSH-National Institute of Health Research Biomedical Research Centre.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Aswin Chari <http://orcid.org/0000-0003-0053-147X>

REFERENCES

- 1 Abbasi B, Goldenholz DM. Machine learning applications in epilepsy. *Epilepsia* 2019;60:2037–47.
- 2 Larivière S, Bernasconi A, Bernasconi N, *et al.* Connectome biomarkers of drug-resistant epilepsy. *Epilepsia* 2021;62:6–24.
- 3 Pedersen M, Verspoor K, Jenkinson M, *et al.* Artificial intelligence for clinical decision support in neurology. *Brain Commun* 2020;2:fcaa096.
- 4 Goldenholz DM, Goldenholz SR, Romero J, *et al.* Development and validation of forecasting next reported seizure using e-Diaries. *Ann Neurol* 2020;88:588–95.
- 5 Gleichgerricht E, Keller SS, Drane DL, *et al.* Temporal lobe epilepsy surgical outcomes can be inferred based on structural connectome hubs: a machine learning study. *Ann Neurol* 2020;88:970–83.
- 6 Kelly CJ, Karthikesalingam A, Suleyman M, *et al.* Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019;17:195.
- 7 Vollmer S, Mateen BA, Bohner G, *et al.* Machine learning and artificial intelligence research for patient benefit: 20 critical questions



- on transparency, replicability, ethics, and effectiveness. *BMJ* 2020;368:l6927.
- 8 Shah P, Kendall F, Khozin S, *et al.* Artificial intelligence and machine learning in clinical development: a translational perspective. *NPJ Digit Med* 2019;2:1–5.
 - 9 Hirst A, Philippou Y, Blazeby J, *et al.* No surgical innovation without evaluation: evolution and further development of the ideal framework and recommendations. *Ann Surg* 2019;269:211–20.
 - 10 Sedrakyan A, Campbell B, Merino JG, *et al.* IDEAL-D: a rational framework for evaluating and regulating the use of medical devices. *BMJ* 2016;353:i2372.
 - 11 Bilbro NA, Hirst A, Paez A, *et al.* The ideal reporting guidelines: a Delphi consensus statement stage specific recommendations for reporting the evaluation of surgical innovation. *Ann Surg* 2021;273:82–5.
 - 12 Marcus HJ, Bennett A, Chari A, *et al.* IDEAL-D framework for device innovation: a consensus statement on the preclinical stage. *Ann Surg* 2022;275:73–9.
 - 13 Wagstyl K, Adler S, Pimpel B, *et al.* Planning stereoelectroencephalography using automated lesion detection: retrospective feasibility study. *Epilepsia* 2020;61:1406–16.
 - 14 Adler S, Wagstyl K, Gunny R, *et al.* Novel surface features for automated detection of focal cortical dysplasias in paediatric epilepsy. *Neuroimage Clin* 2017;14:18–27.
 - 15 Jin B, Krishnan B, Adler S, *et al.* Automated detection of focal cortical dysplasia type II with surface-based magnetic resonance imaging postprocessing and machine learning. *Epilepsia* 2018;59:982–92.
 - 16 McGovern RA, Ruggieri P, Bulacio J, *et al.* Risk analysis of hemorrhage in stereo-electroencephalography procedures. *Epilepsia* 2019;60:571–80.
 - 17 Collins GS, Reitsma JB, Altman DG, *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015;162:55–63.
 - 18 Bossuyt PM, Reitsma JB, Bruns DE, *et al.* Stard 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;351:h5527.
 - 19 Wagstyl K, Whitaker K, Raznahan A. Multi-Centre epilepsy lesion detection (MELD) project: predictors of lesion location and postsurgical seizure freedom in focal cortical dysplasia. *medRxiv* 2021.
 - 20 Barba C, Cross JH, Braun K, *et al.* Trends in pediatric epilepsy surgery in Europe between 2008 and 2015: Country-, center-, and age-specific variation. *Epilepsia* 2020;61:216–27.