Designing health research projects that leverage digital technology

Dr Kit Huckvale, Lead, Digital Health Validitron

Dr Olivia Metcalf, Senior Research Fellow, Centre for Digital Transformation of Health & Department of Psychiatry

THE UNIVERSITY OF MELBOURNE

(They suffered so you don’t have to.)

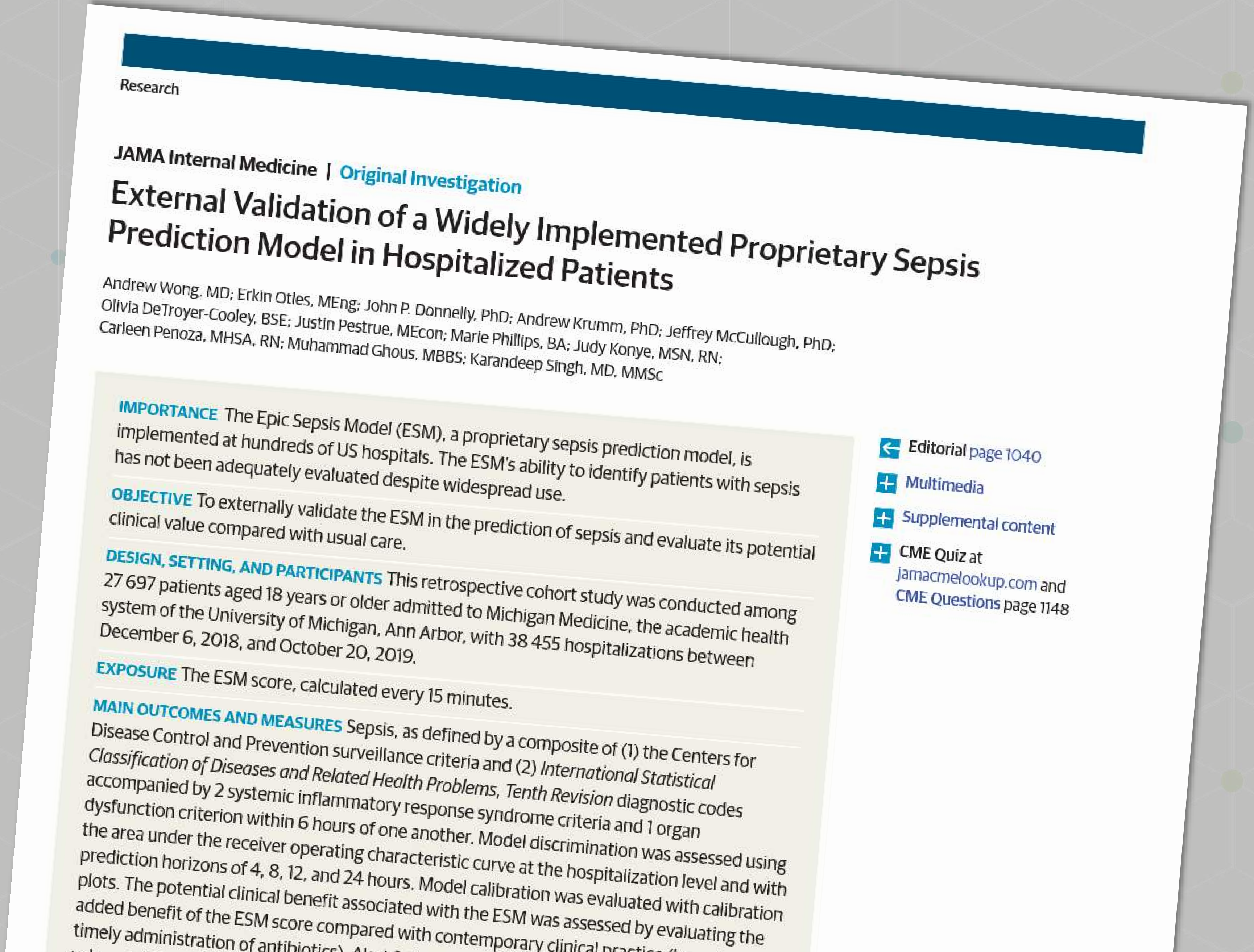
The value reported by this peak flow calculator app changes depending on whether you hold your phone vertically or horizontally.

An evidence-based treatment converted to app format (i.e., the same content but without the human-to-human therapeutic relationship) distressed users and resulted in disengagement.

In its default state, this diabetes management app recommends increasing insulin doses in response to falling blood glucose levels.

>10,000 downloads

Implementation of a hospital PROMs tool failed because no-one was assigned to charge the iPad.


Research

JAMA Internal Medicine | Original Investigation

External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients

Andrew Wong, MD; Erkin Otles, MEng; John P. Donnelly, PhD; Andrew Krumm, PhD; Jeffrey McCullough, PhD; Olivia DeTroyer-Cooley, BSE; Justin Pestrue, MEcon; Marie Phillips, BA; Judy Konye, MSN, RN; Carleen Penoza, MHSA, RN; Muhammad Ghous, MBBS; Karandeep Singh, MD, MMSc

IMPORTANCE The Epic Sepsis Model (ESM), a proprietary sepsis prediction model, is implemented at hundreds of US hospitals. The ESM's ability to identify patients with sepsis has not been adequately evaluated despite widespread use.

OBJECTIVE To externally validate the ESM in the prediction of sepsis and evaluate its potential clinical value compared with usual care.

DESIGN, SETTING, AND PARTICIPANTS This retrospective cohort study was conducted among 27 697 patients aged 18 years or older admitted to Michigan Medicine, the academic health system of the University of Michigan, Ann Arbor, with 38 455 hospitalizations between December 6, 2018, and October 20, 2019.

EXPOSURE The ESM score, calculated every 15 minutes.

MAIN OUTCOMES AND MEASURES Sepsis, as defined by a composite of (1) the Centers for Disease Control and Prevention surveillance criteria and (2) *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision* diagnostic codes accompanied by 2 systemic inflammatory response syndrome criteria and 1 organ dysfunction criterion within 6 hours of one another. Model discrimination was assessed using the area under the receiver operating characteristic curve at the hospitalization level and with prediction horizons of 4, 8, 12, and 24 hours. Model calibration was evaluated with calibration plots. The potential clinical benefit associated with the ESM was assessed by evaluating the added benefit of the ESM score compared with contemporary clinical practice (...) timely administration of antibiotics) ...

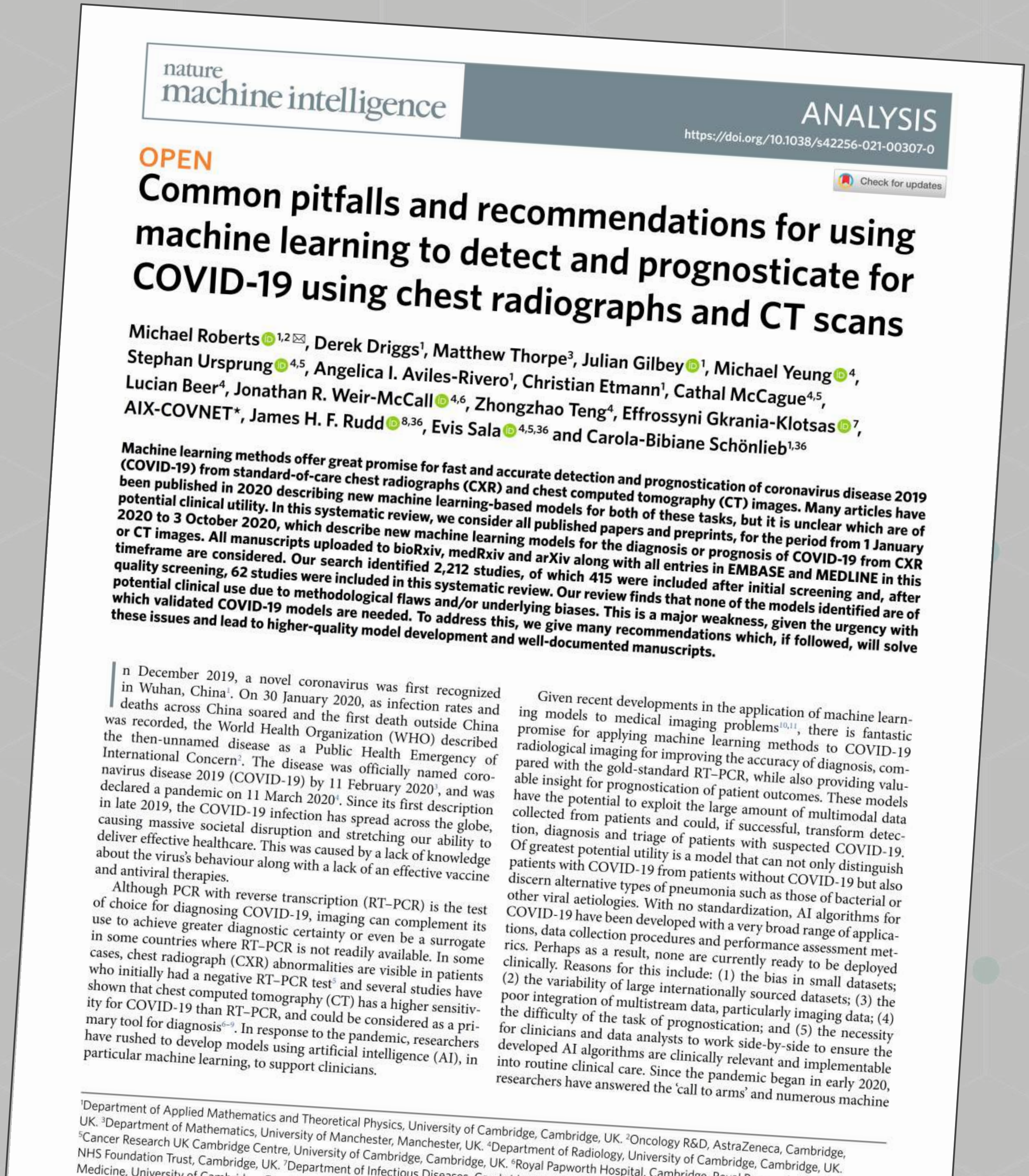
Editorial page 1040
Multimedia
Supplemental content
CME Quiz at jamacmelookup.com and CME Questions page 1148

The Epic® sepsis algorithm

"At our selected score threshold of 6, the ESM had a **hospitalization-level sensitivity of 33%, specificity of 83%.**"

Wong et al., 2021

Wearable-derived metrics of sleep and/or physical health were found to worsen physical and mental health in some.


nature machine intelligence

ANALYSIS

https://doi.org/10.1038/s42256-021-00307-0

OPEN

Check for updates

Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans

Michael Roberts[1,2], Derek Driggs[1], Matthew Thorpe[3], Julian Gilbey[1], Michael Yeung[4], Stephan Ursprung[4,5], Angelica I. Aviles-Rivero[1], Christian Etmann[1], Cathal McCague[4,5], Lucian Beer[4], Jonathan R. Weir-McCall[4,6], Zhongzhao Teng[4], Effrossyni Gkrania-Klotsas[7], AIX-COVNET*, James H. F. Rudd[8,36], Evis Sala[4,5,36] and Carola-Bibiane Schönlieb[1,36]

Machine learning methods offer great promise for fast and accurate detection and prognostication of coronavirus disease 2019 (COVID-19) from standard-of-care chest radiographs (CXR) and chest computed tomography (CT) images. Many articles have been published in 2020 describing new machine learning-based models for both of these tasks, but it is unclear which are of potential clinical utility. In this systematic review, we consider all published papers and preprints, for the period from 1 January 2020 to 3 October 2020, which describe new machine learning models for the diagnosis or prognosis of COVID-19 from CXR or CT images. All manuscripts uploaded to bioRxiv, medRxiv and arXiv along with all entries in EMBASE and MEDLINE in this timeframe are considered. Our search identified 2,212 studies, of which 415 were included after initial screening and, after quality screening, 62 studies were included in this systematic review. Our review finds that none of the models identified are of potential clinical use due to methodological flaws and/or underlying biases. This is a major weakness, given the urgency with which validated COVID-19 models are needed. To address this, we give many recommendations which, if followed, will solve these issues and lead to higher-quality model development and well-documented manuscripts.

In December 2019, a novel coronavirus was first recognized in Wuhan, China[1]. On 30 January 2020, as infection rates and deaths across China soared and the first death outside China was recorded, the World Health Organization (WHO) described the then-unnamed disease as a Public Health Emergency of International Concern[2]. The disease was officially named coronavirus disease 2019 (COVID-19) by 11 February 2020[3], and was declared a pandemic on 11 March 2020[4]. Since its first description in late 2019, the COVID-19 infection has spread across the globe, causing massive societal disruption and stretching our ability to deliver effective healthcare. This was caused by a lack of knowledge about the virus's behaviour along with a lack of an effective vaccine and antiviral therapies.

Although PCR with reverse transcription (RT–PCR) is the test of choice for diagnosing COVID-19, imaging can complement its use to achieve greater diagnostic certainty or even be a surrogate in some countries where RT–PCR is not readily available. In some cases, chest radiograph (CXR) abnormalities are visible in patients who initially had a negative RT–PCR test[5] and several studies have shown that chest computed tomography (CT) has a higher sensitivity for COVID-19 than RT–PCR, and could be considered as a primary tool for diagnosis[6–9]. In response to the pandemic, researchers have rushed to develop models using artificial intelligence (AI), in particular machine learning, to support clinicians.

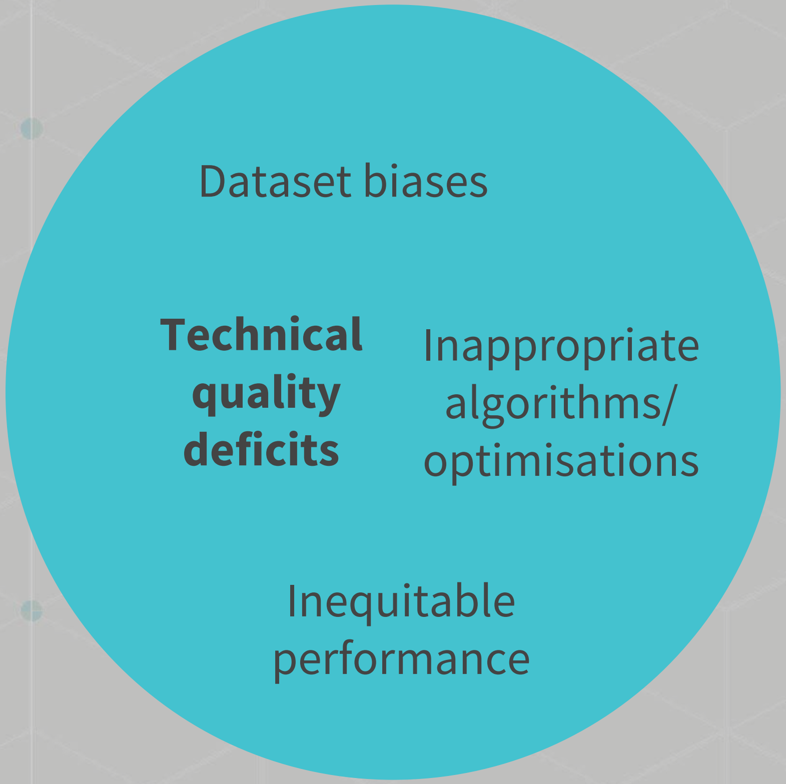
Given recent developments in the application of machine learning models to medical imaging problems[10,11], there is fantastic promise for applying machine learning methods to COVID-19 radiological imaging for improving the accuracy of diagnosis, compared with the gold-standard RT–PCR, while also providing valuable insight for prognostication of patient outcomes. These models have the potential to exploit the large amount of multimodal data collected from patients and could, if successful, transform detection, diagnosis and triage of patients with suspected COVID-19. Of greatest potential utility is a model that can not only distinguish patients with COVID-19 from patients without COVID-19 but also discern alternative types of pneumonia such as those of bacterial or other viral aetiologies. With no standardization, AI algorithms for COVID-19 have been developed with a very broad range of applications, data collection procedures and performance assessment metrics. Perhaps as a result, none are currently ready to be deployed clinically. Reasons for this include: (1) the bias in small datasets; (2) the variability of large internationally sourced datasets; (3) the poor integration of multistream data, particularly imaging data; (4) the difficulty of the task of prognostication; and (5) the necessity for clinicians and data analysts to work side-by-side to ensure the developed AI algorithms are clinically relevant and implementable into routine clinical care. Since the pandemic began in early 2020, researchers have answered the 'call to arms' and numerous machine

[1]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK. [2]Oncology R&D, AstraZeneca, Cambridge, UK. [3]Department of Mathematics, University of Manchester, Manchester, UK. [4]Department of Radiology, University of Cambridge, Cambridge, UK. [5]Cancer Research UK Cambridge Centre, University of Cambridge, Cambridge, UK. [6]Royal Papworth Hospital, Cambridge, Royal ... NHS Foundation Trust, Cambridge, UK. [7]Department of Infectious Diseases ... Medicine, University of Cam...

AI/ML for COVID-19 screening and detection

"[Of 62 studies, o]ur review finds that **none of the models** are of potential clinical use."

Roberts et al., 2021


Dataset biases
Technical quality deficits
Inappropriate algorithms/ optimisations
Inequitable performance

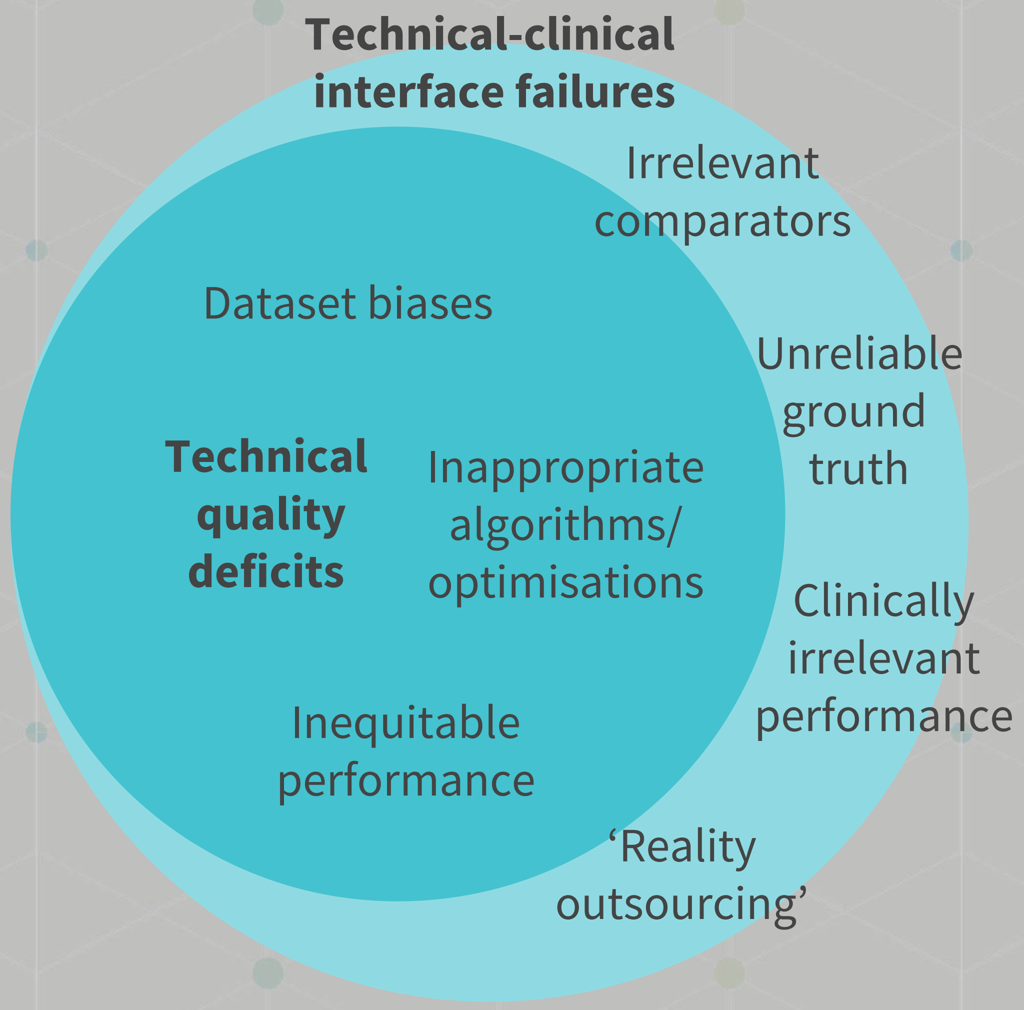
Stuff people are talking about

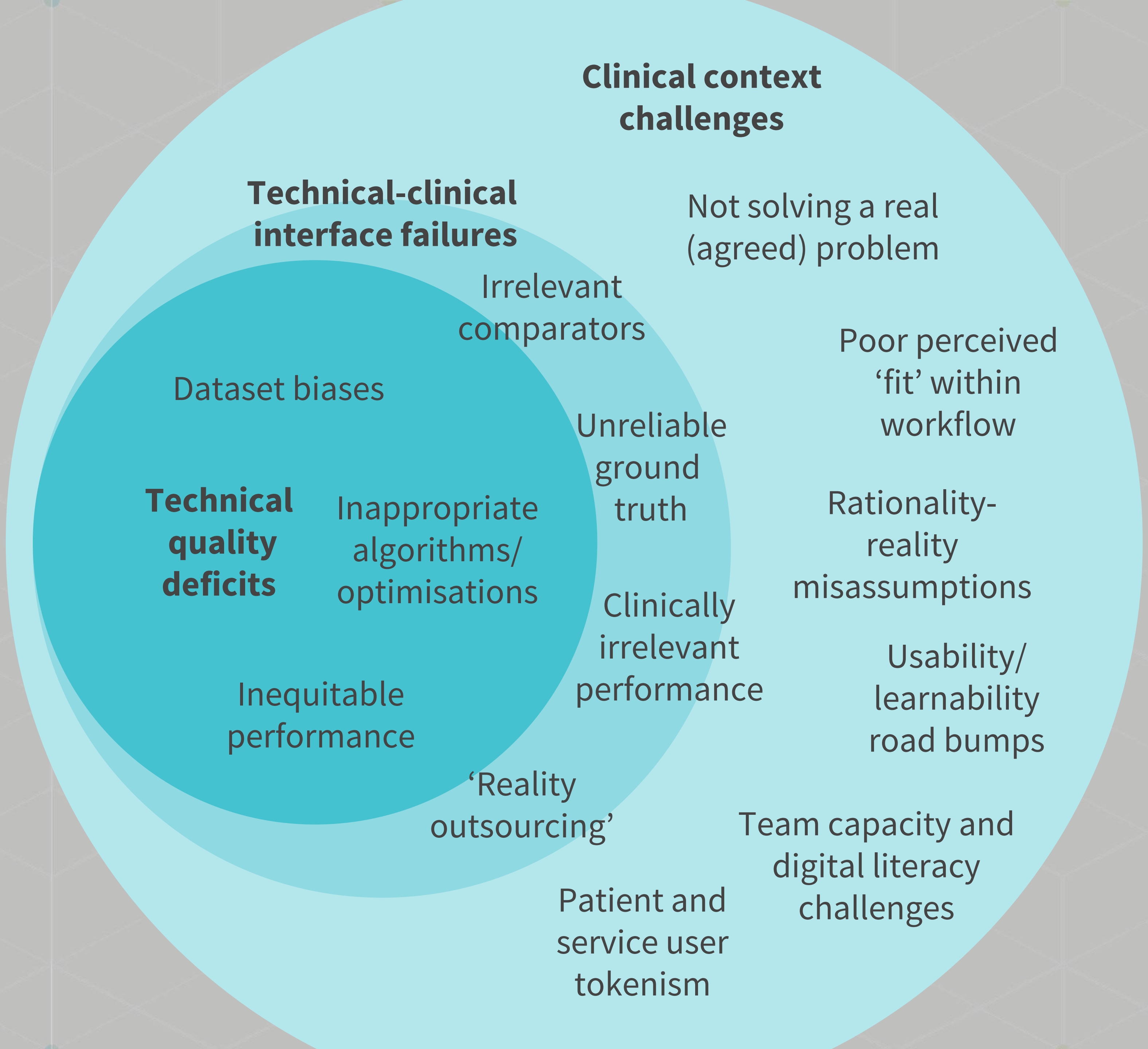
Dataset biases

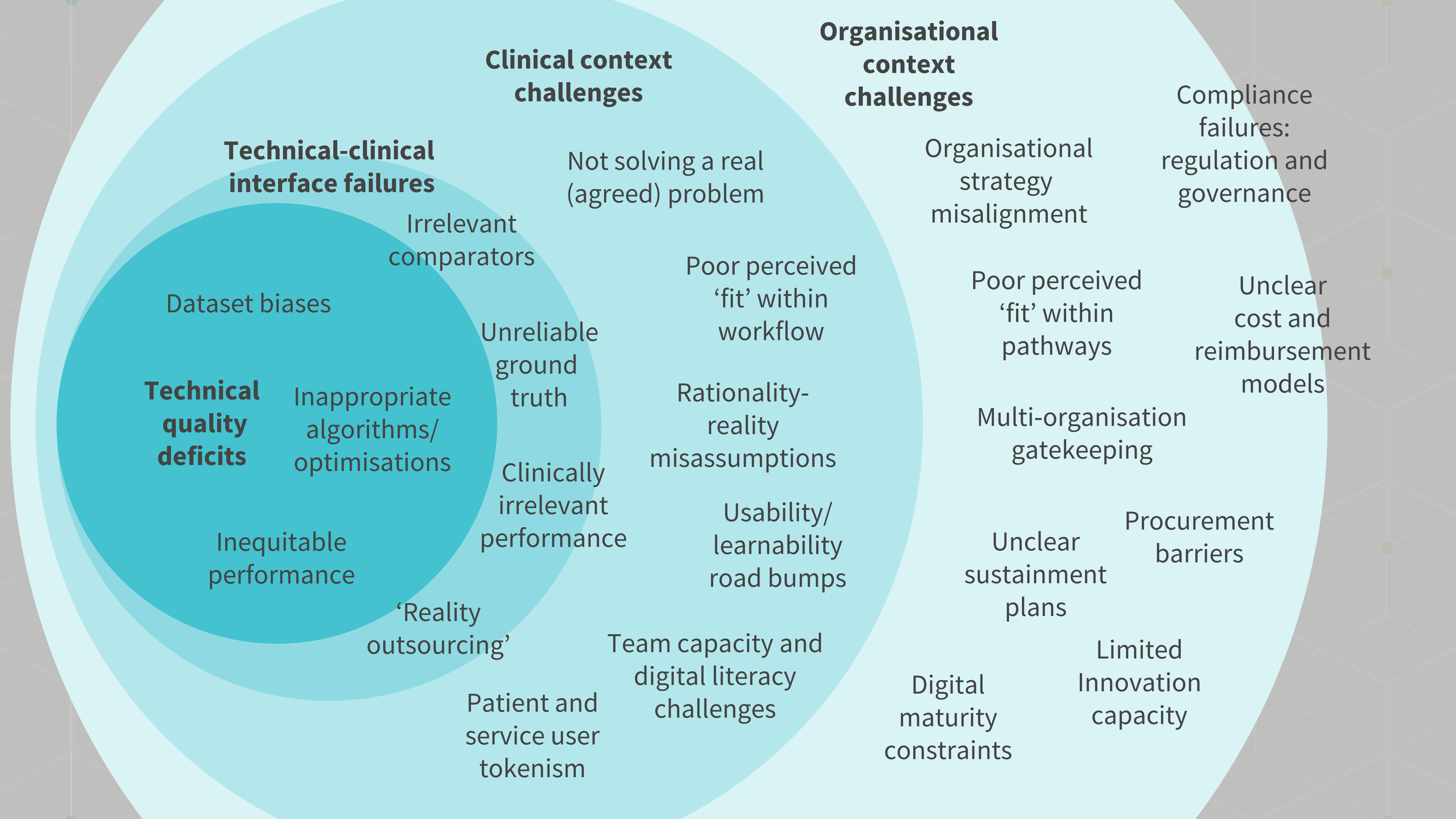
Technical quality deficits

Inappropriate algorithms/ optimisations

Inequitable performance


Technical-clinical interface failures
Irrelevant comparators
Dataset biases
Unreliable ground truth
Technical quality deficits
Inappropriate algorithms/ optimisations
Clinically irrelevant performance
Inequitable performance
'Reality outsourcing'


Clinical context challenges
Technical-clinical interface failures
Not solving a real (agreed) problem
Irrelevant comparators
Poor perceived 'fit' within workflow
Dataset biases
Unreliable ground truth
Technical quality deficits
Inappropriate algorithms/ optimisations
Rationality-reality misassumptions
Clinically irrelevant performance
Usability/ learnability road bumps
Inequitable performance
'Reality outsourcing'
Team capacity and digital literacy challenges
Patient and service user tokenism


Organisational context challenges
Clinical context challenges
Technical-clinical interface failures
Technical quality deficits
Dataset biases
Inappropriate algorithms/ optimisations
Inequitable performance
Irrelevant comparators
Unreliable ground truth
Clinically irrelevant performance
'Reality outsourcing'
Not solving a real (agreed) problem
Poor perceived 'fit' within workflow
Rationality-reality misassumptions
Usability/ learnability road bumps
Team capacity and digital literacy challenges
Patient and service user tokenism
Organisational strategy misalignment
Compliance failures: regulation and governance
Poor perceived 'fit' within pathways
Unclear cost and reimbursement models
Multi-organisation gatekeeping
Unclear sustainment plans
Procurement barriers
Digital maturity constraints
Limited Innovation capacity

So, what can you do to prevent these disasters?

So, what can you do to prevent these disasters?

Evidence-informed tools and methods for de-risking digital health projects…

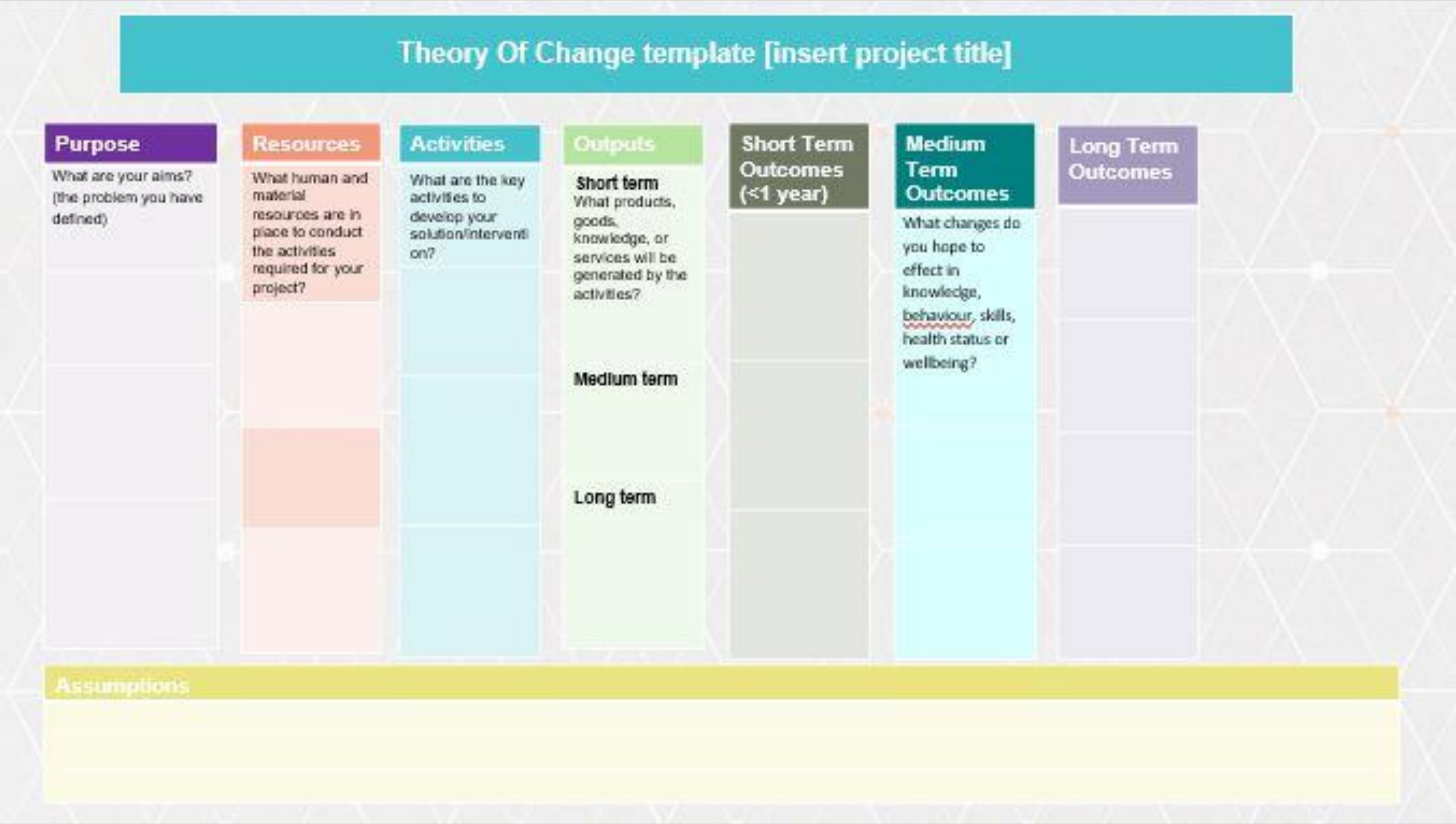
Step 1: Identify where you are in the LHS cycle

Step 2

- Build a multi-disciplinary team with expertise in:
  - Your clinical problem
  - Digital health research expertise
  - Software development expertise or folks who know how to talk to them
  - Statisticians and/or big data experts to manage the data
  - Consumers

Step 3

- Build a logic model of your digital tool, with attention to capturing assumptions and unintended consequences



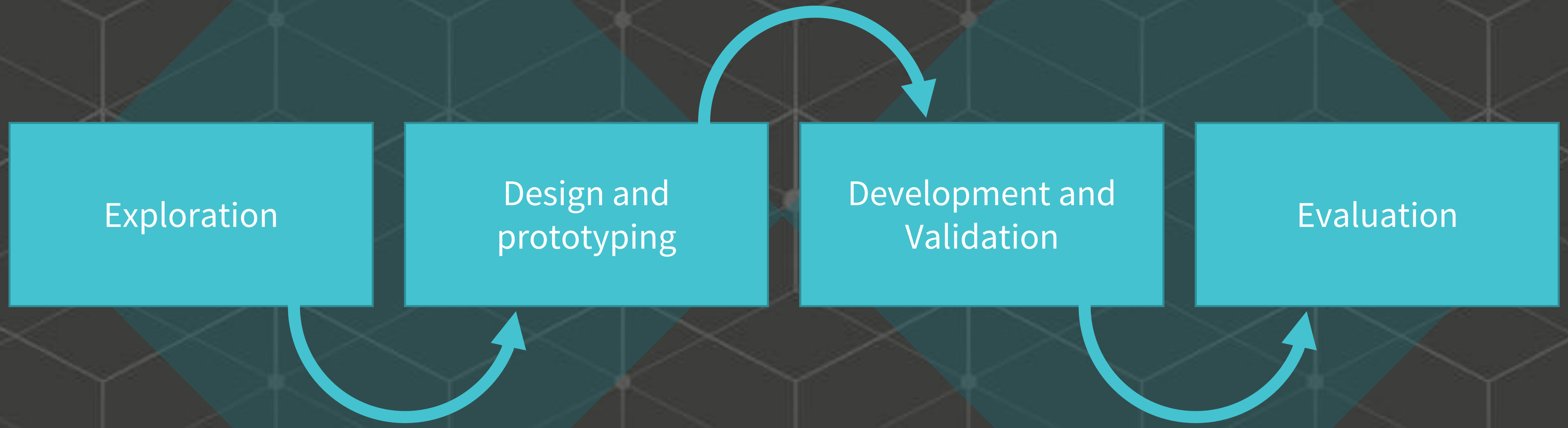
Step 4

Prototype:

1. Scan the evidence and environment for pre-existing tools and evidence
2. Co-design your digital tool with stakeholders
3. Seek extensive feedback on early prototypes before spending money on development
4. Test and model necessary changes to workflows
5. Don’t engage developers too early

Step 5

- Work with digital health researchers to design your evaluation, economics analyses, and implementation plans.
- Engage early with regulatory, legal, and IP experts around commercialisation plans.


Exploration
Design and prototyping
Development and Validation
Evaluation

Validitron

Who are the users?
Who are the stakeholders?

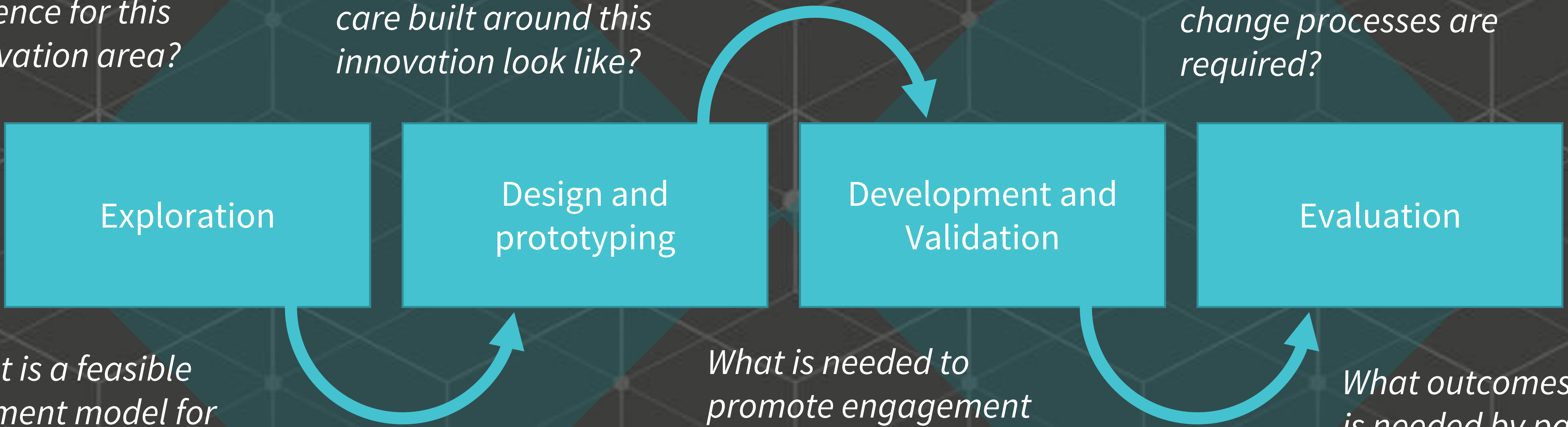
How and where can clinical information generated by this innovation best be used?

How can the usability, acceptability and safety of this innovation be optimised (for realistic clinical scenarios)?

What is the state of evidence for this innovation area?

What would a model of care built around this innovation look like?

What training, implementation and/or change processes are required?



What is a feasible payment model for this innovation?

What is needed to promote engagement and sustained, meaningful use?

What outcomes evidence is needed by payers and for compliance?

What regulatory frameworks will it need to operate within?

Validitron

How can automated digital/behavioural data collection enable outcomes/experience measurement?

Team expertise

Dr Debbie Passey	Digital health implementation science and digital readiness
Dr Hasan Ferdous	Human factors for digital health
Ms Isabella Preston	Digital health UX/UI design
Dr Kara Burns	Consumer involvement and digital health equity specialist
Dr Mahima Kalla	Digital health co-design
Dr Meredith Layton	Non-traditional evaluation design
Dr Olivia Metcalf	Digital health intervention specialist
Dr Omar Dabash	AIML/LLM evaluation in digital health
Dr Quy Nguyen Dinh	Situational awareness
Dr Portia Cornell	Digital health economics
Dr Teresa Wulandari	Simulation-based research in digital health
Dr Joël Foussuo Tagne	Digital health in health services


Simulation-based research

Sandbox

Private access to primary/secondary EMRs for development and testing.

Rapid prototyping with FHIR, SMART and CDS.

Private LLM cloud for dev/test with clinical and sensitive data.

Dr Jane Goller
MSPGHS

1. Making it easier to access the skills needed to explore the dimensions that affect implementation success.

2. Validating digital health ideas with representative users as early as possible.

Ways to collaborate with us

- Free one hour consultation with senior researchers
- Up to five hours free grant writing support and advice for grants that intend to use the Validitron
- Bespoke help with wherever you are on your digital health journey:
  - Logic modelling workshops
  - Co-design facilities and expertise
  - Prototyping, UX and HCI support
  - Access to Sandbox and Simulation facilities
  - Evaluation design
  - Implementation science and health economics support

Connect with the Validitron

- Digital health simulation facilities.
- Codesign, implementation science and evaluation advice.
- End-to-end digital health human factors and validation study design and execution.
- Grant collaborations.

validitron-team@unimelb.edu.au

www.validitron.com.au