Creating an evidence standards framework for artificial intelligence enabled digital health technologies

Trystan Macdonald, MBChB[1,2], Viknesh Sounderajah, MSc[3], Harriet Unsworth, PhD[4], Verena Wolfram, PhD[4], Bernice Dillon, PhD[4], Mark Salmon, MBA[4], Felix Greaves, PhD[4], Xiaoxuan Liu, PhD[1,2], Alastair Denniston, PhD[1,2], Hutan Ashrafian, PhD[3], Ara Darzi, MD[3], Carolyn Ashurst, PhD[5], Chris Holmes, PhD[5], Adrian Weller, PhD[5]
for the NICE Artificial Intelligence Evidence Standards Working Group

[1] Institute of Inflammation and Ageing, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK, [2] University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK, [3] Institute of Global Health Innovation, Imperial College London, London, UK, [4]National Institute for Health and Care Excellence, London, UK, [5]The Alan Turing Institute, London, UK

Background.

The National Institute for Health and Care Excellence (NICE) is an independent public body of the Department of Health and Social Care in England which develops quality standards for those providing and commissioning health services. In 2020 it commissioned academic partners to develop AI specific evidence standards for an update of its' Evidence Standards Framework (ESF) for Digital Health Technologies (DHTs), a tool intended to be used by UK commissioners to evaluate the clinical and economic benefits of adopting DHTs.

Methods.

These standards were developed through a Delphi consensus study consisting of two online rounds and a final expert consensus meeting. There were a total of 83 international participants in the study as a whole.

Results.

1	**The vendor should provide documentation clearly describing the AI model, in line with current best practice (e.g. Sendak et al. (7)). This should include regulatory status, intended use (input data source and type, outputs, target population, value proposition, place in care pathway, decision support requirements), details of training (demographic breakdown, dates, and location of training and validation data), performance measures (test data, subgroup analyses), safety and performance limitations (risks, contra-indications and people for whom it should not be used, when to discontinue use)**
2	**The vendor should provide evidence how the AI-DHT performs in subgroups within the intended use population. This may be stratified by relevant factors such as race, ethnicity, gender, sex and any other clinically relevant groups**
3	**The vendor must provide a description of how incomplete or outlier data was handled in the development of the by the AI-DHT and how these will be handled post-deployment**
4	**The vendor must describe the training datasets to show that relevant, high-quality data was used to train the algorithm, in line with current best-practice (e.g. UK Algorithmic Transparency Data Standard, Datasheets for Datasets)**
5	**The vendor should document the approaches used in the design, development, and testing of the AI-DHT to detect and minimise potential harm to users, which may be either the direct or indirect consequence of the use of the AI-DHT**
6	**The vendor should provide a definition of the level of autonomy with which the AI-DHT is expected to work. This definition must use plain-English and be understandable to the end-user.**
7	**The vendor must document the outputs for the algorithm with confidence levels where relevant.**
8	**The vendor and commissioner should agree a post-deployment oversight process. This may include regular monitoring of an AI-DHTs performance by an agreed party with reports to commissioners at intervals or when certain circumstances are met e.g. a significant deterioration in performance**
9	**Before deployment of the DHT, the vendor and commissioner must agree a post-deployment change management plan. This is an agreement on how the algorithm(s) in the DHT is/are expected to change over time, and how the resulting changes in the effectiveness of the DHT will be measured.**
10	**The vendor should provide a description of any known factors that may impact the performance of a DHT at a new deployment site, and describe what actions will be taken (before and during deployment) to ensure that the DHT performs to the expected level in the new setting**
11	**The vendor must provide a full description of the input data for the AI-DHT.**
12	**Vendors should provide commissioners with sufficient information to counsel patients on the use, limitations, risks and benefits of AI-DHTs so that they may make an informed decision regarding their involvement in their care.**
13	**The vendor must provide a description of the expected training requirements for end-users to safely and effectively use the AI-DHT, and how these needs will be met in practice**

Conclusion.

Incorporated within the updates NICE ESF these standards provide a first in class tool for UK commissioners to assess the clinical and economic impact of DHTs whilst safeguarding patient privacy, safety and fairness.