Lessons from the NHS AI Awards:
experience of evaluating AI tools in real-world environments

EE564

Benjamin Caswell-Midwinter[1], Babak Jamshidi[1], Catriona Inverarity[1,2], Emily Kwong[1,2], Chloe Black[1,2], Fowzia Ibrahim[4], Lauren Gatting[1], Minjie Gao[3], Anthony Tsang[1], Akashdeep Singh Chauhan[1], Alison Griffiths[1], Jo Waller[5], Juan I Baeza[3], Jon Hindmarsh[3], Salma Ayis[4], Angie A Kehagia[1], Anna Barnes[1,2].

1. King's Technology Evaluation Centre (KiTEC), London Institute for Healthcare Engineering, King's College London. 2. Medical Physics and Clinical Engineering, Guy's and St Thomas's NHS Foundation Trust, 3. Department of Public Services Management & Organisation, King's Business School, King's College London, London, UK, 4. School of Life Course & Population Sciences, Faculty of Life Sciences & Medicine, King's College London. 5. Wolfson Institute of Population Health, Queen Mary University of London

Introduction

The AI in Health and Care Award ran in partnership with the Accelerated Access Collaborative (AAC) and NHS AI Lab 2020-2022; to accelerate the testing and evaluation of the most promising AI technologies that meet the strategic aims set out in the NHS Long Term Plan.

Technology Specialist Evaluation Teams (TSET) were commissioned to perform the evaluation of the successful AI technologies with market authorisation but insufficient evidence to merit large-scale commissioning or deployment.

KiTEC were partnered with 4 award winners (1,2)

Image-segmentation (workflow efficiencies): possible cost savings through staff changes, increased precision and fewer side effects experienced by patients, time to treatment reduced.

Detect Vertebral Compression Fractures (service): possible cost savings by preventative treatment, increased standardisation between radiology departments to report incidental findings.

Detect ECG abnormalities (service): Address staff shortages, and waiting times, cost savings through preventative treatments

Detect breast lesions (screening): Address staff shortages.

Methods

Framework: Systems Engineering for better healthcare "improving improvement" (3)

Roadmap: Evidence Standards Framework for Digital Health Technologies: NICE (4)

Real-world Data (5): Observational where possible.

Stats: Generalised Linear Models, with additional post-hoc analysis

Early Health Economics: Short time horizons
- Decision Tree (pathways)
- Complex Process Evaluation (6)
- Qualitative Research Analysis methods.

Evaluation Domains: safety, accuracy, effectiveness, value, fit with sites, implementation, feasibility of scaling up, sustainability (AI in Health cand Care Award) (7)

Analysis and results

1.**Understand the context** – this step includes stakeholder mapping, clinical pathway mapping, organisational and geographical context.

2.**Define the problem** – this step helps describe the clinical need/gap that the technology is fulfilling and consequently defines a robust value-add or value-proposition

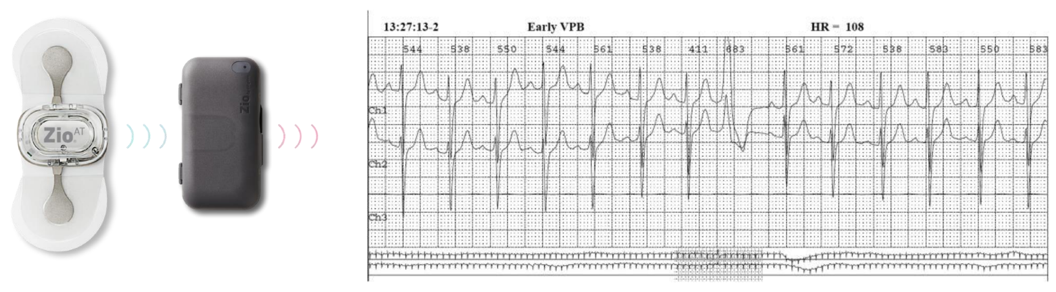
3.**Develop the solution** – The solution is not only the technology itself but how the technology is implemented and used in order to provide the solution to the clinical need/gap. The stakeholder engagement alongside the clinical pathway mapping will be used to identify the customers that will contribute to this solution.

4.**Collect the evidence** – Data from each of the 3 previous steps is used to develop a data collection protocol that is both comprehensive and convenient if the data doesn't already exist.

5.**Make the case** – Once the data has been collated and analysed the results will then be interpreted and evidence extracted in such a way as to address the requirements of, for example, a NICE health technology evaluation process.

6.**Manage the plan** – The data collection trial can be managed in such a way as to simultaneously gather information around the deployment, implementation and sustainability of the technology at scale.

Ambulatory ECG service



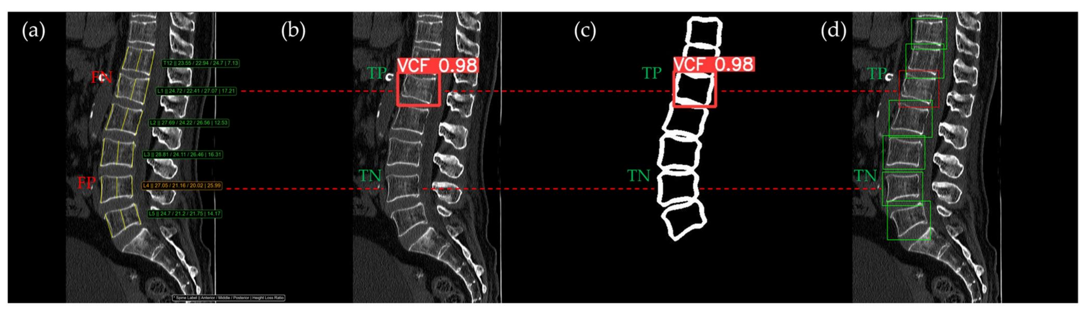
Population: adults referred for ambulatory ECG post stroke or cardiac arrest
Intervention: ZioXT service, ECG patch posted to company.
Comparator: ambulatory ECG holter service standard NHS pathway
Outcomes: Time to treatment, and resource costs

Statistical Analysis highlighted a decrease in time to treatment, reduced number of repeat tests, and rapid availability of report.

Health Economics found the service to be cost saving for bulk orders of consumables, and due to reduced clinic visits.

Stakeholder Engagement – Stroke services benefited from avoiding cardiology waiting lists and patients preferred it to the conventional 6 lead holter, cardiology departments were difficult to reach possibly due to lack of healthcare scientist cardio-physiology workforce involvement.

Opportunistic vertebral fracture detection tool



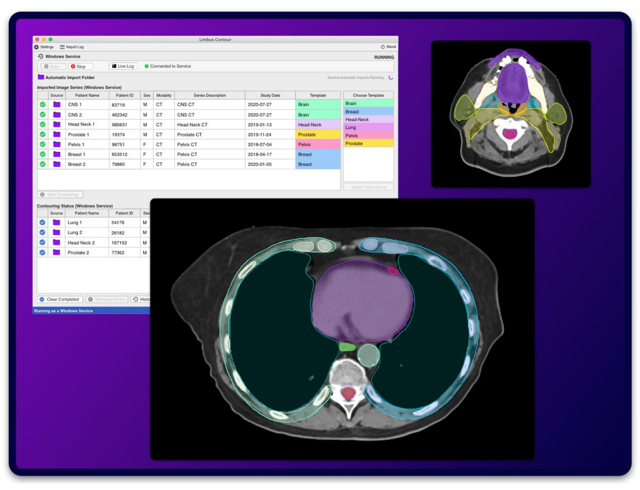
Population: adults referred for CT scan of thorax, abdo, pelvis
Intervention: AI enabled vertebral compression fracture (VCF) detection and auto referral to fracture clinic
Comparator: Standard care pathway, manual detection of VCF and referral.
Outcomes: time to referral and resource costs

Statistical Analysis revealed non-inferiority of detection of fractures both retrospective and prospective data collection compared to manual read.

Health Economics: increased cost due to increase in number of referrals but no improvement in patient outcomes (no follow-up data). Unclear who would bear the cost of implementation radiology or fracture liaison service.

Stakeholder Engagement revealed heterogeneous practices in reporting incidental findings, an onerous installation process of the software, and lack of IT infrastructure enabling exchange of information between departments

Image segmentation



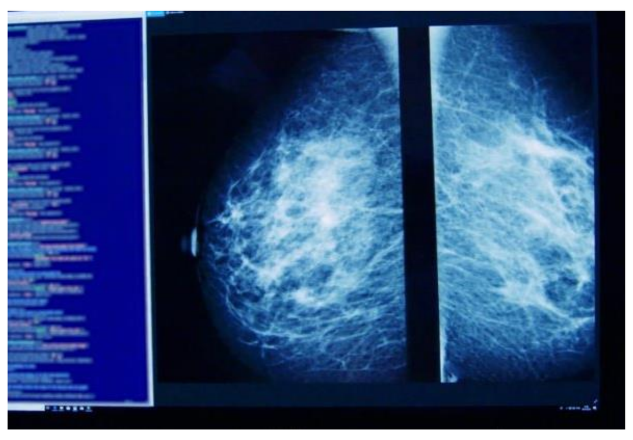
Population: adults referred for external beam radiotherapy treatment of primary tumour in: Head and Neck, Lung, pelvis and breast.
Intervention: AI enabled contouring of normal anatomy.
Comparator: manual delineation of anatomy
Outcomes: time to complete, acceptability of AI result, resource costs

Statistical Analysis highlighted a decrease in time to complete the task but was attenuated across the whole workflow and that there was a high acceptability of AI results.

Health Economics found the service to be cost/time saving for all cancers with the highest cost/time saving being for the breast cancer workflow due to the low number of corrections needed and therefore avoiding the costs associated with highly experienced, skilled staff who could be redeployed elsewhere.

Stakeholder Engagement – highlighted an overwhelmingly positive attitude to the use of this kind of tool, seen as labour saving and trustworthy (associated with duration of use)

Breast screening – lesion detection



Population: Females 50-70 referred for routine mammography screening
Intervention: AI enabled abnormality detection
Comparator: Standard care pathway, dual manual reporting of abnormality
Outcomes: sensitivity, specificity and resource costs

Statistical Analysis revealed non-inferiority of detection of lesions retrospective and prospective data collection compared to manual read.

Health Economics: increased cost due to increase in false positives requiring an additional read with no increase in number of cancers detected.

Stakeholder Engagement highlighted difficulties with integration into existing IT infrastructure and national reporting database and an onerous installation and commissioning process of the software.

Conclusion

The degree to which a technology can be cost saving is heavily influenced by

- its **position in the workflow** hence the importance of a detailed process evaluation and stakeholder mapping and engagement before implementation, is often dominated by staff costs
- Its **position in the system**: referral clinics vs screening, results direct to referrer vs via electronic records, at home or in the clinic.
- **Who Pays**: NHS vs local provider. There maybe cost savings overall but opportunity costs in one department vs another. Local Dept budgets are unconnected.
- **Increase in treatment costs** due to higher productivity and/or higher detection rates.

The degree to which a technology is successfully implemented is influenced by

- **User trust:** immediate demonstration of accuracy and precision, time savings, ease of installation and commissioning, a certain degree of explainability.
- **Value Add**: AI does not obviate the need for human healthcare professionals but can allow highly skilled staff to do more complex tasks. Indirect consequences may include reduced unpaid overtime, increased patient facing time, higher job satisfaction, lower sickness, decreased attrition/turnover.
- Robustness and flexibility **of existing IT infrastructure** and presence of R&D innovation teams. Access to data is an issue, due to poor understanding of GDPR, cybersecurity by non-healthcare support staff.

Recommendations: Use Digital Twin technology to allow "*in-silico*" simulations of clinical pathways for operations research informed early health economic modelling.

Acknowledgements

The projects described here were funded by the Artificial Intelligence in Health and Care Award, one of the NHS AI Lab programmes. The competitive award scheme was run by the Department of Health & Social Care (DHSC), supported by the National Institute for Health & Care Research (NIHR) and the Accelerated Access Collaborative (AAC).

References

1. https://www.nihr.ac.uk/ai-health-and-care-awards-funded-projects-2020
2. https://www.nihr.ac.uk/ai-health-and-care-award-funded-projects-2021 .
3. University of Cambridge Engineering Design Centre. Improving improvement: a toolkit for Engineering Better Care. iitoolkit.com accessed 28th October 2025.
4. https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies
5. https://www.nice.org.uk/corporate/ecd9/chapter/overview
6. Process evaluation of complex interventions: Medical Research
7. Council guidance. G Moore et al. BMJ 2015;350:h1258
8. https://www.england.nhs.uk/aac/what-we-do/how-can-the-aac-help-me/ai-award/evaluation-for-artificial-intelligence/

Contact information

London Institute for Healthcare Engineering
100 Lambeth Palace Road
London SE1 7AR
anna.barnes@kcl.ac.uk

Medical Physics & Clinical Engineering
Tabard House
Guy's & St Thomas's NHS Foundation Trust
London SE1 1YP
catriona.inverarity@kcl.ac.uk