

AI for assisting the reporting of chest x-rays: challenges of quantitatively evaluating live implementation

Chris Sherlaw-Johnson (1), Stuti Bagri (1), Malina Bodea (1), Emma Dodsworth (1), Nadia Crellin (1), Holly Elphinstone (2), Naomi Fulop (2), Amanda Halliday (4), Kevin Herbert (3), Rachel Lawrence (2), Joanne Lloyd (4), Raj Mehta (4), Pei Li Ng (2), Steve Morris (3) Holly Walton (2), Angus Ramsay (2)

(1) Nuffield Trust, (2) University College London, (3) University of Cambridge, (4) Public contributor

Background

- We conducted a rapid evaluation of AI in chest diagnostics funded by the National Institute for Health and Care Research (NIHR).
- This was a mixed methods evaluation focusing on multiple aspects of impact and implementation.

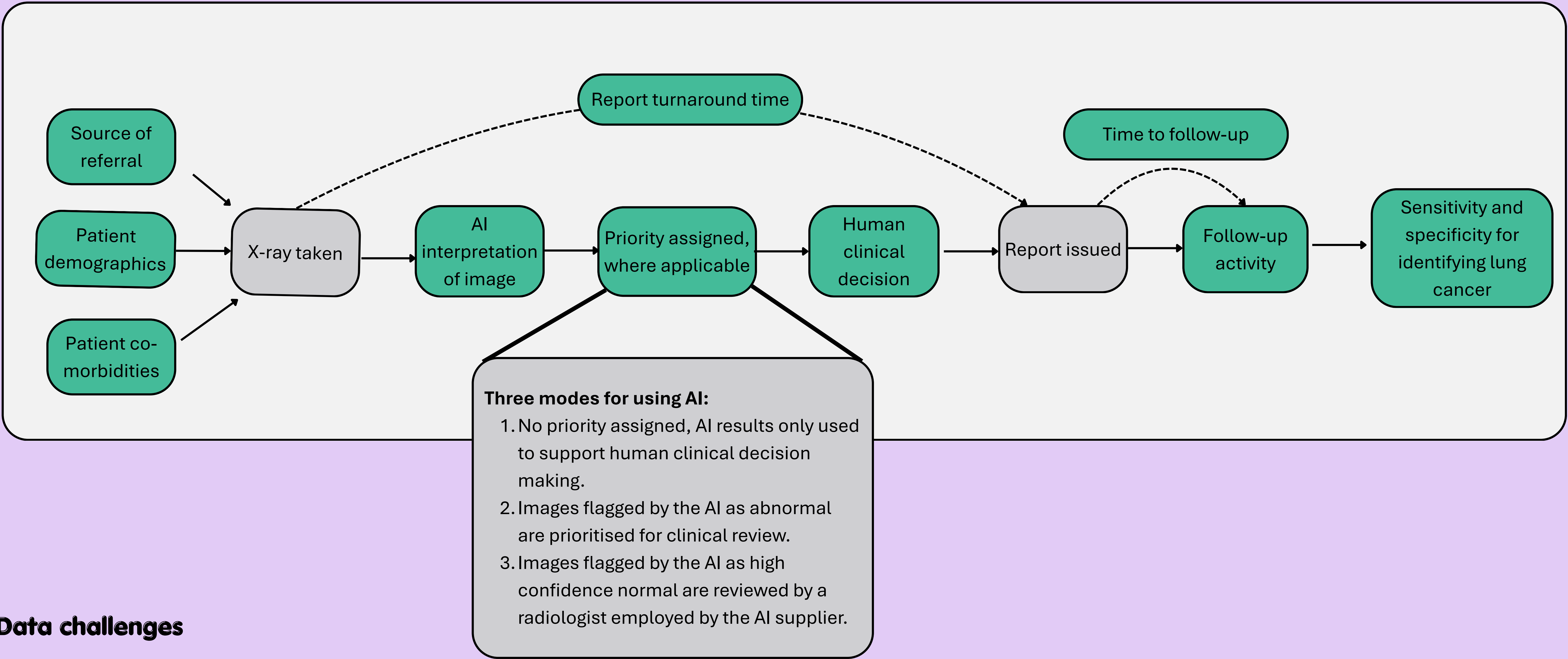
Aim of this sub-study

- The aim of this sub-study was to identify challenges with the quantitative evaluation and their implications for monitoring and evaluating the effectiveness of live clinical AI interventions.

Methods and materials

- Mixed methods rapid evaluation of 49 acute hospital trusts implementing AI for chest diagnostics through initiatives funded by NHS England between 2023 and 2025.
- Knowledge gathered from:
 - 31 interviews with imaging network teams, trust staff and AI suppliers.
 - the processes of acquiring and analysing quantitative data for eight sites.

Flowchart of x-ray processing with AI with an outline of data requirements



Data challenges

- Many sites had difficulty extracting the data required for a robust evaluation. Linkage across data systems couldn't be exploited.
 - This created problems with measuring longer term impact of AI such as on cancer occurrence.
 - A general absence of pathway indicators meant that CT scans could not be linked to prior chest x-rays with certainty.
 - Outcomes were not linked to relevant patient clinical histories.
- AI engagements were not always coded .
- Lack of control groups.

Where challenges were overcome

- One local imaging network was able to coordinate deployment across sites , a unified platform for the AI and a central data repository with relevant linkage.

Other challenges affecting quantitative evaluation

- Implementation varied between sites making it hard to generalise the impact of AI.
- Interventions were a combination of introducing the AI and changes to the process or resourcing to accommodate the AI. This made it difficult to unpick the impact of the AI on its own.

Recommendations

1. IT systems and data linkage need to be developed to enable the capture of relevant data and allow effective local monitoring of safety an at the outset.d performance.
2. IT limitations should inform decisions to deploy at the outset.
3. Site-level monitoring is important to allow for the different ways that AI has been implemented.
4. Methods such as phased implementation within a site might mitigate lack of control groups.

Project findings
slide set:



Project
webpage:



RSET
webpage:



RSET
newsletter
sign-up:

