BMJ Health & Care Informatics

Codesigned standardised referral form: simplifying the complexity

Scott Laing ,[1] Sarah Jarmain,[2] Jacobi Elliott,[3,4,5] Janet Dang,[6] Vala Gylfadottir,[5] Kayla Wierts,[7] Vineet Nair[4]

To cite: Laing S, Jarmain S, Elliott J, *et al*. Codesigned standardised referral form: simplifying the complexity. *BMJ Health Care Inform* 2024;**31**:e100926. doi:10.1136/bmjhci-2023-100926

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjhci-2023-100926).

Received 05 October 2023
Accepted 01 June 2024

[1]Family Medicine, McMaster University, Kitchener, Ontario, Canada
[2]McMaster University, Hamilton, Ontario, Canada
[3]St Joseph's Health Care London, London, Ontario, Canada
[4]Western University, London, Ontario, Canada
[5]Middlesex London Ontario Health Team, London, Ontario, Canada
[6]Ontario Health, Toronto, Ontario, Canada
[7]eHealth Centre of Excellence, Kitchener, Ontario, Canada

Correspondence to
Dr Scott Laing;
scottwlaing@gmail.com

ABSTRACT

Background Referring providers are often critiqued for writing poor-quality referrals. This study characterised clinical referral guidelines and forms to understand which data consultant providers require. These data were then used to codesign an evidence-based, high-quality referral form.

Methods This study used both observational and quality improvement approaches. Canadian referral guidelines were reviewed and summarised. Referral data fields from 150 randomly selected Ontario referral forms were categorised and counted. The referral guideline summary and referral data were then used by referring providers, consultant providers and administrators to codesign a referral form.

Results Referral guidelines recommended 42 types of referral data be included in referrals. Referral data were categorised as patient demographics, provider demographics, reason for referral, clinical information and administrative information. The percentage of referral guidelines recommending inclusion of each type of referral data varied from 8% to 77%. Ontario referral forms requested 264 different types of referral data. Digital referral forms requested more referral data types than paper-based referral forms (55.0±10.6 vs 30.5±8.1; 95% CI p<0.01). A codesigned referral form was created across two sessions with 29 and 21 participants in each.

Discussion Referral guidelines lack consistency and specificity, which makes writing high-quality referrals challenging. Digital referral forms tend to request more referral data than paper-based referrals, which creates administrative burdens for referring and consultant providers. We created the first codesigned referral form with referring providers, consultant providers and administrators. We recommend clinical adoption of this form to improve referral quality and minimise administrative burdens.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Healthcare providers are burning out due to increasing administrative burdens, and referral processes contribute to these administrative burdens.

WHAT THIS STUDY ADDS

⇒ Canadian referral guidelines are ambiguous and not specific enough to facilitate improvements to referral content quality.
⇒ Digital referral forms are significantly longer than paper referral forms, which contributes to administrative burdens.
⇒ The codesigned referral form template clearly defines which referral data elements to include in referral forms while reducing requests for extraneous information.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This study may help inform future referral forms, digital referral system development and data structures, and health policies to minimise the administrative burdens faced by healthcare providers, while improving the quality of referrals.

INTRODUCTION

Referral letters to consultant providers have been criticised for their poor quality due to the omission of relevant and important referral data dating back to the early 1990s.[1] More recent literature has identified that referral letters lack important information, such as patient contact information, reason for referral, presumptive diagnosis, symptoms and physical exam findings.[2 3] Poor quality and incomplete referrals can delay patient care, leading to patient harm and decreased quality of care.[2 4] The cause of care delays is in part due to administrative burdens, as consultant providers must request missing information from referring providers,[5] who likely did not realise that essential referral data were missing in their initial referral. Attempts have been made to mitigate this issue by defining essential referral data through surveys,[1] creation of referral quality scoring systems[6 7] and referral guidelines.[8] However, studies continue to critique the quality of referrals.[4] To our knowledge, no published literature has characterised the referral data that is requested by consultant providers based on clinically used referral forms at a system level. Nor has any study codesigned a referral form with referring providers, consultant providers and administrators.

Creating consensus on which referral data are required by consultants is important for improving referral quality and the transition

to digital referral systems, like eReferral.[9] Development of digital referral systems requires clearly defined data fields[10] over traditional free-text letters. The main benefits of eReferral are that referring providers can send referrals via the internet instead of fax, find consultant providers closer to the patient or who have shorter wait times and patients receive email notifications about their referrals as they are triaged and booked.[9] The timing of this study is important since eReferrals are becoming more common in Canada,[11] meaning there is an opportunity to create standardised referral forms prior to widespread clinical adoption. To do this, we followed the Canadian Medical Associations' 2014 recommendation to codesign referral forms.[8]

This codesign initiative was also in response to the increasing administrative burdens on primary care providers. In 2023, primary care providers in Ontario, Canada were spending 19.1 hours per week on administrative tasks.[12] These administrative burdens arise from: detailed clinical documentation and data entry; inefficient user interfaces; cognitive burdens caused by reminders and irrelevant or redundant patient data and management of clinical messages and inboxes.[13] This is consistent with other findings that healthcare providers are spending at least 2 hours on administrative tasks for each hour of direct patient contact.[14] Importantly, primary care is experiencing the highest level of administrative burdens, leading to provider burnout.[13] In Canada, 53% of primary care providers report burnout, 61% report experiencing significant emotional distress, 64% report their jobs are highly stressful, 76% report a significant increase in workload since 2020 and many plan to stop providing patient care in the next 1 to 3 years.[15] Accordingly, it is essential that initiatives like this are undertaken to reduce administrative burdens and improve provider experiences to avoid future health human resource crises.

This study aimed to establish consensus on which referral data are essential for high-quality referrals. This was accomplished by characterising Canadian referral guidelines and the referral data fields on publicly available and clinically used referral forms from Ontario, Canada. Referring providers, consultant providers and clinic administrators then codesigned a standardised referral form based on these findings. This codesigned referral form was then clinically used on the eReferral platform. Subsequently, both referring and consultant providers were surveyed to report their clinical experience using this codesigned referral form. The primary outcome of this study was the creation of an evidence-based codesigned referral form.

METHODS

Review of referral guidelines and policy statements

A review of the current Canadian referral guidelines was completed. Guidelines and policy statements were collected from national, provincial and territorial medical licensing bodies or medical associations (online supplemental table S1). Each document was reviewed, and all referral data recommended to include in referrals were recorded and categorised by data type in tabular format.

Characterisation of referral data fields on clinically used referral forms

Clinically used and publicly available paper referral forms from Ontario, Canada were collected from OSCAR EMR[16] and South West Primary Care Alliance (SWPCA).[17] These websites are the most comprehensive repositories of paper referral form stored as PDFs (Portable Document Format) and images in Ontario. All referral forms on the websites were extracted using Web Scraper[18] into a Microsoft Excel[19] spreadsheet containing the clinic or consultant provider name, geographic region, specialty, type of referral and URL to each form. All digital referral forms on eReferral[9] and the corresponding clinic or consultant provider name, geographic region, specialty, type of referral and URL were provided by the eHealth Centre of Excellence[20] in a Microsoft Excel[19] spreadsheet. These two files were combined, then all forms were manually reviewed to exclude administrative tools, clinical tools, laboratory requisitions, government programme application forms, diagnostic imaging forms and duplicates since the focus of this study was consultation request forms.

One-hundred and fifty referral forms were randomly selected from the included forms. Each form was assigned a random number, sorted by number and the top 150 using Microsoft Excel.[19] Each form was manually reviewed by author SL, who is a practising family physician. Each referral data field and corresponding format (informational, free text, check boxes, attachments) were recorded using Microsoft Excel.[19] All fields were highlighted after recording to ensure complete data capture. Referral data were later categorised based on the specific data requested. Clinical judgement was required to differentiate between referral data types.

Statistical analysis

All data recording and statistical analysis were performed using Microsoft Excel.[19] The total number of unique referral data fields on each form was determined and the mean number of data fields for paper and digital referral forms was calculated. The average number of informational, checkboxes, free text entry and attachment requests were then calculated for paper and digital referrals. Unpaired t-tests assuming unequal variances were then performed to determine whether there are differences in the average number of unique referral data fields between paper and digital referral forms. A subanalysis was also performed with unpaired t-tests assuming unequal variances for each format of requested referral data.

Codesigned referral form

Two codesign sessions were completed where referring providers, consultant providers and healthcare

administrators collaborated to create a standardised and generic referral form. Participants were recruited by email through local hospital and primary care organisations. Participating referring providers included primary care physicians and nurse practitioners. Participating consultant providers included specialist physicians and nurse practitioners. Participating administrative providers included secretarial staff and clinic managers. Participants were provided with prereadings detailing the review of referral guidelines and review of referral data, outlined above. Participants discussed each category of referral data in a facilitated open forum and identified which information they felt should always be included in all referrals. Participants' decisions and comments were recorded during the session, then used to create a standardised referral form.

In the second session, participants discussed each section of the referral form and commented on their impression of the created form. Participants were asked to identify any referral data that were missing or that required revision. Participants' decisions and comments were recorded during the session and then appropriate revisions were made to the standardised referral form.

Codesigned referral form user experience

The codesigned referral form was then used in Ocean eReferral for a period of 5 months prior to seeking feedback. Referring providers that sent referrals using the codesigned referral form were contacted to provide feedback on their experience. These referring providers were contacted via email and provided a URL to a survey about their experience using the codesigned referral form. Specifically, the question *How was your experience completing Ocean eFax referral forms compared to normal fax-based referrals?* was used to assess providers' experience using the codesigned referral form. Respondents rated their experience using a Likert scale indicating, Excellent, Good, Fair, Poor, Very Poor, or Not applicable.

Ethics approval

This study did not require ethics approval as it used publicly available information and is quality improvement in nature as per the University of Ottawa Research Ethics Board.

RESULTS

Review of referral guidelines and policy statements

Review of the national and provincial referral guidelines identified 7 categories of referral data and 42 specific types of referral data for inclusion in referral letters (table 1). The number of guidelines recommending each type of referral data was variable from 1 (8%) to a maximum of 10 (77%). There were no referral data types that were recommended to include in referrals by all guidelines. Referral guidelines were noted to provided ambiguous referral data inclusion recommendations. These ambiguities arise from general statements to include 'patient information', 'primary care provider information' and 'clinical information' which were not specifically defined.

Characterisation of referral data fields on clinically used referral forms

A total of 622 documents were collected from OSCAR EMR, SWPCA and Ocean eReferral (online supplemental table S2). Four hundred and fifteen documents remained after excluding 9 administrative documents, 8 clinical tools, 6 COVID-19 resources, 163 diagnostic imaging forms, 7 duplicates, 10 laboratory forms, 1 long-term care application and 3 patient information sheets (online supplemental table S2). The included referral form represented 42 different specialties (online supplemental table S2). The 150 randomly selected forms represented 32 different specialties (table 2) and had representation from all geographic areas within Ontario, Canada (online supplemental table S3).

Review of the 150 randomly selected referral forms identified 264 unique types of referral data that were requested by consultant providers (table 3). This means that consultants requested 222 more unique referral data types than were identified in the referral guidelines (table 3). Additionally, 23 types of social history referral data were requested, which were not included in any referral guidelines (table 3). Administrative referral data were limited in the referral guidelines but were frequently requested by consultant providers. Referral data fields were classified into four different formats, information for referring providers, checkboxes, free-text entry or attachment requests. A full list of all referral data identified in the referral forms is available in online supplemental table S4.

The average number of referral data fields per digital referral was significantly higher than paper referrals (55.0±10.6 vs 30.5±8.1; 95% CI p<0.01; figure 1). Subgroup analysis (figure 2) demonstrated that digital referrals have significantly higher average number of informational data (13.7±1.7 vs 8.6±2.9; 95% CI p<0.01), checkboxes (13.7±1.7 vs 4.8±3.4; 95% CI p<0.01) and free-text entry (31.1±5.0 vs 16.1±6.0; 95% CI p<0.01) requests. The average number of attachment requests was not significantly different between digital and paper referrals (1.1±1.4 vs 1.0±1.0; 95% CI p=0.38).

Codesigned referral form

A total of 29 and 21 participants attended the first and second codesign sessions, respectively (table 4). During the first session, participants decided on which referral data were included in the standardised referral form (table 5). Codesign design participants indicated that some specialties may need custom referral forms with additional information requests (online supplemental table S5). The customisations identified during the codesign sessions included options to select specific providers or locations, referral eligibility criteria, required prereferral testing, disease-specific clinical guidance, unique data to triage appropriate clinic location and specific

Table 1 Summary of Canadian referral guidelines separated into categories and specific referral data that were recommended to include in referrals

Category (N=7)	Referral data element (N=42)	Number of guidelines recommending (%)
Patient demographics	Patient contact info	8/13 (62)
	Patient health number	8/13 (62)
	Patient name	7/13 (54)
	Patient date of birth	6/13 (46)
	Patient address	3/13 (23)
	Gender	3/13 (23)
	"Patient information"	2/13 (15)
	Language	2/13 (15)
Referring provider demographics	Referring provider contact info	4/13 (31)
	Referring provider name	3/13 (23)
	Referring provider info	2/13 (15)
Primary care provider demographics	Primary care provider info	4/13 (31)
Consultant provider demographics	Description of consultant's referral process	4/13 (31)
	Description of fees not covered	3/13 (23)
	Consultant provider name	2/13 (15)
	Consultant provider contact info	1/13 (8)
	Consultant provider service type	1/13 (8)
	Description of services offered	1/13 (8)
	Description of services not offered	1/13 (8)
Reason for referral	Reason(s) for consultation	9/13 (69)
	Expected outcome	5/13 (38)
	Indicate if requested by a third party	2/13 (15)
	Clinical question	2/13 (15)
	Information being sought	1/13 (8)
Clinical information	Relevant investigations	10/13 (77)
	Urgency	7/13 (54)
	Past medical history	7/13 (54)
	Medications	7/13 (54)
	Physical exam	6/13 (46)
	Allergies	5/13 (38)
	Relevant consultant notes	5/13 (38)
	Clinical information	3/13 (23)
	Other involved healthcare providers	3/13 (23)
	Current and past management of specific issue	2/13 (15)
	Patient's clinical stability	1/13 (8)
	Duration of issue	1/13 (8)
	Key symptoms/red flags	1/13 (8)
	Comorbidities	1/13 (8)
	Pending investigations	1/13 (8)
Administrative information	Date of referral	7/13 (54)
	Option for referrer inform patient of appointment	2/13 (15)
	Confirmation patient is aware of referral	1/13 (8)

Table 2 Distribution of specialties and number of referral forms randomly selected for review

Specialty (N=32)	eReferral forms	Paper forms	Total
Cardiology		5	5
Chiropody		1	1
Dermatology	9	2	11
Endocrinology		3	3
Gastroenterology	4	1	5
Genetics		1	1
Home care	3	1	4
Mental health and addictions	17	2	19
Multi-specialty		1	1
Nephrology		1	1
Neurology	4	4	8
Oncology		3	3
Optometry		1	1
Orthopaedics	11	2	13
Pain medicine		2	2
Palliative	1		1
Paediatrics	12	4	16
Physiatry		1	1
Physiotherapy		2	2
Podiatry		1	1
Psychiatry	6	7	13
Public health		1	1
Respirology		3	3
Sleep medicine		4	4
Thoracic surgery		1	1
Trans health	1	1	2
Urology	5		5
Vascular		3	3
Women's health/ gynaecology	13	3	16
Geriatrics		1	1
Internal medicine		1	1
Telemedicine		1	1
Total	**86**	**64**	**150**

criteria to triage referral priority. During the codesign discussion, participants expressed that customisations would help referring providers provide better referral, improve patient care and reduce unnecessary or inappropriate referrals.

The codesign participants also provided several recommendations to improve referral form quality beyond the specific referral data. Participants recommended that referral forms be brief and minimise administrative burdens, leverage electronic medical records to

Table 3 Categories of unique referral data on referral forms and in referral guidelines

Categories (N=8)	Unique data elements (N)	
	Referral forms	Guidelines
Clinical information	78	15
Administrative information	53	3
Patient demographics	52	8
Consultant provider demographics	25	7
Social history	23	0
Referring provider demographics	18	3
Primary care provider demographics	8	1
Reason for referral	7	5
Total	**264**	**42**

auto-complete referral forms, only ask for referral data that referring providers will reasonably have, and avoid collection of referral data that do not facilitate referral triage and eligibility decision-making. Codesign participants recommended that if consultant providers require more detailed information, then this could be collected either prior to the initial consultation via a patient completed intake questionnaire or during the initial consultation.

Codesigned referral form user experience

A total of 147 referring providers that sent referrals using the codesigned referral form were contacted to provide feedback on their experience. Eighteen responses were

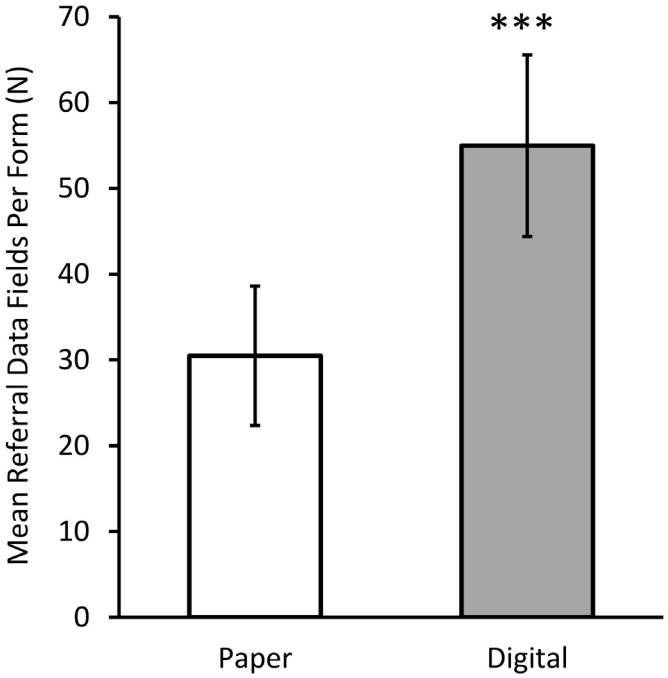

Figure 1 Mean number of referral data types requested per paper and digital referral form (***p≤0.01).

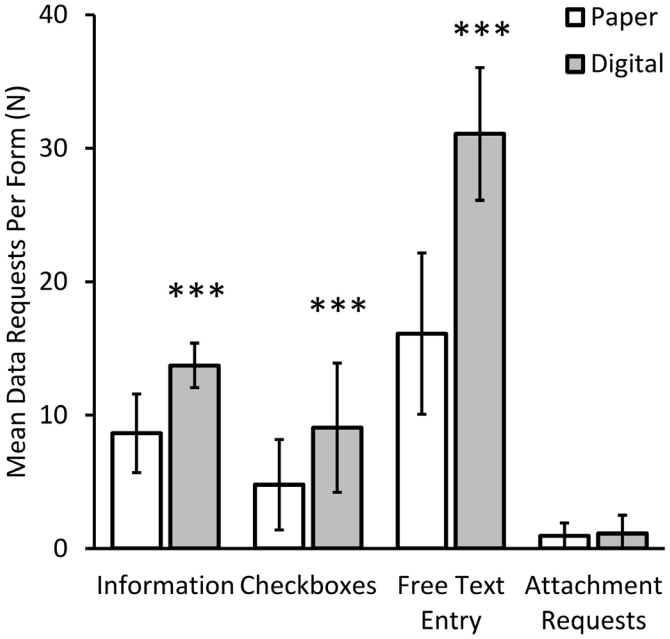

Figure 2 Mean number of referral data types requested within each category for paper and digital referral forms (***p≤0.001).

received from providers in three different regions of Ontario (Central, Toronto and West). All 18 responses were from providers working in primary care clinics. Of those who responded, 11 (61%) were primary care physicians, 4 (22%) were nurse practitioners, 1 (6%) was an allied health practitioner, 1 (6%) was an office administrator and 1 (6%) was a referral clerk. Respondents rated their experience using the codesigned referral form positively, with 14 of 17 respondents rating their experience as excellent or good (2 excellent and 12 good).

DISCUSSION

Referring providers have been criticised for writing poor-quality referrals for many years.[1–3] This study identified inconsistencies and ambiguities within Canadian referral guidelines. No referral data type was consistently identified for inclusion by all referral guidelines. Surprisingly, only 7 of 13 (54%) referral guidelines specifically recommended the patient's name and 8 of 13 (62%) recommended the patient's contact information, be included in the referral. The remaining guidelines either made no recommendation[21] or generally specified 'patient information'.[22 23] We propose that the lack of consensus and specificity in referral guidelines contributes to why referring providers unintentionally omit essential referral data, leading to low-quality referrals.

Table 4 Categorisation of codesign participants by codesign session

Role	Session 1 (N)	Session 2 (N)
Administrators	10	6
Primary care physicians	11	7
Primary care nurse practitioners	3	1
Specialist physicians	1	4
Specialist nurse practitioners	1	0
Facilitators	3	3
Total	**29**	**21**

To our knowledge, this study presents the first characterisation of referral data fields from clinically used referral forms across multiple specialties. The requested referral data should contain all clinically relevant information that consultants need.[24] We, therefore, used this data to codesign a standardised referral form with referring providers, consultants and administrators as recommended by the Canadian Medical Association.[8] Codesign participants wanted referral forms to be short and simple; however, this study demonstrated that newer digital referral forms requested more information from referring providers. One potential reason why digital referral forms are longer and more complex is because digital referrals are not restricted to a single physical sheet of paper. Limiting digital referral form length is important because additional referral data does not correlate with consultants' confidence in triaging appointments.[25] Referral form length will also increase administrative burdens for referring and consultant providers, which correlates with provider burnout[13 14] and intention to stop practicing.[15] Therefore, we recommend that consultants' providers adopt shorter, standardised, evidence-based referral forms, such as the one codesigned here.

The codesign of this referral form is a step toward providing clearer referral guidelines to improve referral quality. Some studies have recommended that referring providers require more training on how to write referrals,[4 26] however, a Cochrane review from 2008 identified that education alone is insufficient.[27] Instead, we recommend following England's National Health Services' Sustainability Model, creating interventions that target processes, staff education and organisational improvements.[28] The Cochrane review concluded that clear referral guidelines (staff education) released in conjunction with a referral form (process change) can significantly improve referral quality.[27]

A randomised trial in Norway also demonstrated that a combination of provider education and referral form improves referral quality.[29] However, the Norwegian referral form was based on disease-specific clinical guidelines and consultant opinion. The codesign approach used in this study facilitated dialogue between referring providers, consultant providers and administrators, leading to more nuanced learning. Specifically, the codesign participants highlighted that specifying referral data fields is only one part of the problem. Participants expected high-quality referral forms to be brief, reduce administrative burdens, leverage technology to facilitate form completion and only request referral data

Table 5 Codesigned, standardised, generic referral form data template. When possible, 'yes' and 'no' questions were formatted so that a checked box indicates a 'yes' response

Consultant clinic's information	
Clinic name	Clinic's telephone number
Address line 1	Clinic's fax number
Address line 2	Clinic's Email
City, province/territory	Service languages
Postal code	
Important practice announcements: **OPTIONAL** ► Brief description, only used when important updates are required (eg, temporary clinic closures) Service specialty: ► List (eg, cardiology, respirology) Accepted referral indications: ► List (eg, atrial fibrillation, asthma)	Does not see: ► List of indications (eg, valve disorders, COPD) Uninsured services offered: ► List+costs Consultant names: ► List Parking available? Yes/No Parking cost: Free/Cost Site accessible? Yes/No
Patient information	
Surname:	First name:
Date of birth:	Gender: ▫ Male ▫ Female ▫ Other
Health card number:	Health card number version code:
Street address 1:	Mobile number:
Street address 2:	Home number:
City:	Business number:
Province:	Email:
Postal code:	Best method of contact: ▫ Mobile▫ Home ▫ Business ▫ Email
	▫ Voicemails acceptable
Where appropriate, please provide the following	
Preferred language: ▫ English ▫ French ▫ Other	
If 'other' language, preferred language:	▫ Translator required
▫ Appointment booking contact (if not patient)	
Contact name:	Phone number (if different than patient):
Special considerations (third party insurance, accessibility, barriers, tips for care delivery):	
Preferred consultant or location **OPTIONAL FOR GROUP/CENTRAL INTAKES** Only one preferred consultant or location can be selected.	
Preferred consultant (dropdown list with 'shortest wait time' or provider names)	
Preferred Location (dropdown list with 'shortest wait time' or provider names)	
Reason for referral (* indicates required field)	
Urgency*: ▫ Routine ▫ Urgent	
Rationale for urgent referral*:	
Goal of referral: ▫ Advice/Question ▫Co-Management ▫ Diagnostic Clarification ▫ 2nd Opinion ▫ Re-Referral ▫ 3rd Party Request ▫ Other	
Goal of referral*:	
Name of suspected diagnosis/problem triggering referral*:	
Describe advice needed or clinical question:	
Brief description of history, management and investigations*:	
Cumulative patient profile	
▫ CPP attached separately	
▫ CPP excluded (removes next four lines)	

Continued

Table 5 Continued

Consultant clinic's information	
Current medications:	
Current problems:	
Past medical history:	
Allergies:	
Supporting documentation	
Please attach all relevant laboratory and diagnostic investigations from last 6 months.	
☐ Personal health information that is medically relevant has not been disclosed at the request of the patient.	
Referrer's information	
Site name:	Phone number:
Address line 1:	Fax number:
Address line 2:	Billing number:
City:	Professional ID number:
Province:	Signature:
Postal code:	Clinician type:
Copy of referral and status updates to:	

Grey cells have conditional logic to only appear if the grey option is selected in the preceding row to minimise the screen space required to view the form. Some sections are optional and are clearly demarcated as such. This template is most useful for digital referrals where details can be automatically completed, and conditional logic applied as demonstrated. CPP = Cumulative Patient Profile, which is a patient summary. COPD = chronic obstructive pulmonary disease.

that facilitate consultant triage. These finding came organically from the codesign open forum and there is increasing awareness of the value that codesign brings to digital health technology development.[30] Given these findings, it is possible that previous efforts to standardise referral forms have failed because without codesign, referral forms tend towards being longer, more complex and request information that referring providers do not have or is better collected directly from patients.

The next step from this study is to further implement the codesigned referral form in clinical practice. Additionally, we recommend revision to existing referral guidelines to provide clearer direction for referring and consulting providers. Once this is completed, then further quality improvement cycles may be completed to further refine and define the components of high-quality referrals and referral forms.

Strengths and limitations

The main limitation of this study was that all reviewed referral forms and codesign participants were from Ontario. We attempted to mitigate any local practice patterns by collecting forms from all geographic regions within Ontario. However, there may be difference in the referral data that consultants require in different regions. Codesign participants also expressed this concern and suggested that the codesigned referral form could be customised for different regions as needed to assist with referral triage or decision-making. Additionally, this study only assessed referral forms for consultation requests. Accordingly, these findings and the codesigned referral form will not be adequate for all referral types, such as diagnostic, home care, allied health or laboratory services. Finally, the referral form review was completed by a single author due to study constraints. However, each codesign participant had the opportunity to review and discuss these findings, which should mitigate potential biases in the primary outcome—the codesigned referral form.

CONCLUSION

This study has demonstrated that referral guidelines lack consistency and specificity, which may contribute to poor-quality referrals. To our knowledge, this is the first study that has characterised the referral data requested by consultant providers. These data were used to codesign a referral form with referring providers, consultant providers and administrators, which should be adequate for consultation referrals across multiple specialties. Implementation of this codesigned referral form is expected to improve referring providers' experience by reducing administrative burdens and improve referral quality by more clearly defining essential referral data fields. Further studies will be needed to assess and improve the codesigned referral form's impact on referral quality, referral appropriateness, patient safety, and provider experiences.

Contributors SL, SJ, JE, JD, VG and VN contributed to the design of this study. SL contributed to the referral guideline and referral form review. All members were involved in the codesign session, data analysis and manuscript production. SL is the guarantor of this study.

Funding This study has been funded by contributions from Ontario Health via the Central Waitlist Management fund provided to the Middlesex-London Ontario Health Team and from the eHealth Centre of Excellence.

Competing interests SL does consultancy for the eHealth Centre of Excellence, VN does consultancy for the Middlesex London OHT; SJ was the clinician lead for the Middlesex London OHT; SL co-chairs the eHealth Centre of Excellence Referral Best Practices Working Group, JD was a volunteer Board Director for 360 Degree Nurse Practitioner Clinic, Peterborough, ON.

Patient consent for publication Not applicable.

Ethics approval University of Ottawa Research Ethics Board.

Provenance and peer review Not commissioned; externally peer-reviewed.

Data availability statement Data are available upon reasonable request.

ORCID iD

Scott Laing http://orcid.org/0000-0002-9704-459X

REFERENCES

1 Newton J, Eccles M, Hutchinson A. Communication between general practitioners and consultants: what should their letters contain? *BMJ* 1992;304:821–4.
2 Graydon SL, Thompson AE. Triage of referrals to an outpatient rheumatology clinic: analysis of referral information and triage. *J Rheumatol* 2008;35:1378–83.
3 Chetcuti K, Farrugia R, Cassar K. GP referral letters: time for a template? *Malta Med J* 2009;21:26–9.
4 Holly M, Courtney H, Julia M, *et al*. A retrospective evaluation of the quality of referrals to IBD specialist care and its influence on patient outcomes. *Am J Gastroenterol* 2018;113:S15.
5 Lenihan L, Cullen IM, Durkan G, *et al*. GP referrals to the rapid access prostate cancer (RAPC) clinic in Galway; are they adequate? *Ir J Med Sci* 2013;182:S337.
6 François J. Tool to assess the quality of consultation and referral request letters in family medicine. *Can Fam Physician* 2011;57:574–5.
7 Eskeland SL, Brunborg C, Seip B, *et al*. First quality score for referral letters in gastroenterology—a validation study. *BMJ Open* 2016;6:e012835.
8 Canadian Medical Association. *Streamlining patient flow from primary to specialty care: a critical requirement for improved access to specialty care*. 2014:8.
9 CognisantMD. Ocean eReferral network. 2021. Available: https://www.oceanereferralnetwork.ca/about/
10 Ontario Health. Ontario eReferral HI7 FHIR implementation Guide-draft V0.10.1. 2021. Available: https://simplifier.net/guide/ca-on-eReferral-r4-iguide-v0.10.1/Introduction
11 eHealth Centre of Excellence. eReferral adoption strategies throughout the COVID-19 pandemic. 2020. Available: https://ehealthce.ca/userContent/documents/Research%20and%20Evaluation/eReferral%20case%20studies/One%20Pager%20-%20Change%20Mgmt%20During%20COVID%202020_Dec%2011%202020_Final.pdf
12 Ontario College of Family Physicians. Family medicine: a profession in crisis. 2023. Available: https://ontariofamilyphysicians.ca/wp-content/uploads/2023/09/family-medicine-a-profession-in-crisis_ocfp-report.pdf
13 Budd J. Burnout related to electronic health record use in primary care. *J Prim Care Community Health* 2023;14:21503319231166921.
14 Sinsky C, Colligan L, Li L, *et al*. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. *Ann Intern Med* 2016;165:753–60.
15 Stressed out and burned out: the global primary care crisis. 2022. Available: https://www.commonwealthfund.org/publications/issue-briefs/2022/nov/stressed-out-burned-out-2022-international-survey-primary-care-physicians
16 OSCAR Canada Users Society. OSCAR eForms: Ontario specific. 2021. Available: http://oscarcanada.org/oscar-users/emr-resource/eform/eform-examples/ontario-specific
17 South West Primary Care Alliance. EMR resources [Internet]. n.d. Available: https://www.swpca.ca/EMRResource/
18 Web scraper. n.d. Available: https://www.webscraper.io/
19 Microsoft Corporation. Microsoft excel spreadsheet software. n.d. Available: https://www.microsoft.com/en-us/microsoft-365/excel
20 eReferral - eHealth centre of excellence. 2021. Available: https://ehealthce.ca/eReferral.htm
21 Newfoundland and Labrador Medical Association. NLMA guideline: advice for physicians regarding referrals, consultations, and missed hospital appointments. 2022.
22 College of Physicians & Surgeons of Ontario. Transitions in care policy. 2021. Available: https://www.cpso.on.ca/Physicians/Policies-Guidance/Policies/Transitions-in-Care
23 Yukon Medical Council. The referral consultation process. 2022.
24 Jenkins S, Arroll B, Hawken S, *et al*. Referral letters: are form letters better? *Br J Gen Pract* 1997;47:107–8.
25 Jiwa M, Meng X, O'Shea C, *et al*. Impact of referral letters on scheduling of hospital appointments: a randomised control trial. *Br J Gen Pract* 2014;64:e419–25.
26 Speed CA, Crisp AJ. Referrals to hospital-based rheumatology and orthopaedic services: seeking direction. *Rheumatology* 2005;44:469–71.
27 Akbari A, Mayhew A, Al-Alawi MA, *et al*. Interventions to improve outpatient referrals from primary care to secondary care. *Cochrane Database Syst Rev* 2008;2008:CD005471.
28 Maher L, Gustafson D, Evans A. *Sustainability model*. NHS Institute for Innovation and Improvement, 2010. Available: https://www.england.nhs.uk/improvement-hub/wp-content/uploads/sites/44/2017/11/NHS-Sustainability-Model-2010.pdf
29 Wåhlberg H, Valle PC, Malm S, *et al*. Impact of referral templates on the quality of referrals from primary to secondary care: a cluster randomised trial. *BMC Health Serv Res* 2015;15:353.
30 Busse TS, Jux C, Laser J, *et al*. Involving health care professionals in the development of electronic health records: scoping review. *JMIR Hum Factors* 2023;10:e45598.

BMJ Health & Care Informatics

Prediction of coronary artery disease based on facial temperature information captured by non-contact infrared thermography

Minghui Kung,[1,2] Juntong Zeng ,[3,4,5] Shen Lin,[3,4,5,6,7] Xuexin Yu,[1] Chang Liu,[1] Mengnan Shi,[1] Runchen Sun,[3,4,5] Shangyuan Yuan,[1,2] Xiaocong Lian,[8] Xiaoting Su,[3,4,5] Yan Zhao,[3,4,6,7] Zhe Zheng ,[3,4,5,6,7] Xiangyang Ji[1,8]

To cite: Kung M, Zeng J, Lin S, *et al*. Prediction of coronary artery disease based on facial temperature information captured by non-contact infrared thermography. *BMJ Health Care Inform* 2024;**31**:e100942. doi:10.1136/bmjhci-2023-100942

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjhci-2023-100942).

MK, JZ, SL and XY contributed equally.

Received 19 October 2023
Accepted 06 March 2024

For numbered affiliations see end of article.

Correspondence to
Dr Zhe Zheng;
zhengzhe@fuwai.com

Dr Xiangyang Ji;
xyji@tsinghua.edu.cn

ABSTRACT

Background Current approaches for initial coronary artery disease (CAD) assessment rely on pretest probability (PTP) based on risk factors and presentations, with limited performance. Infrared thermography (IRT), a non-contact technology that detects surface temperature, has shown potential in assessing atherosclerosis-related conditions, particularly when measured from body regions such as faces. We aim to assess the feasibility of using facial IRT temperature information with machine learning for the prediction of CAD.

Methods Individuals referred for invasive coronary angiography or coronary CT angiography (CCTA) were enrolled. Facial IRT images captured before confirmatory CAD examinations were used to develop and validate a deep-learning IRT image model for detecting CAD. We compared the performance of the IRT image model with the guideline-recommended PTP model on the area under the curve (AUC). In addition, interpretable IRT tabular features were extracted from IRT images to further validate the predictive value of IRT information.

Results A total of 460 eligible participants (mean (SD) age, 58.4 (10.4) years; 126 (27.4%) female) were included. The IRT image model demonstrated outstanding performance (AUC 0.804, 95% CI 0.785 to 0.823) compared with the PTP models (AUC 0.713, 95% CI 0.691 to 0.734). A consistent level of superior performance (AUC 0.796, 95% CI 0.782 to 0.811), achieved with comprehensive interpretable IRT features, further validated the predictive value of IRT information. Notably, even with only traditional temperature features, a satisfactory performance (AUC 0.786, 95% CI 0.769 to 0.803) was still upheld.

Conclusion In this prospective study, we demonstrated the feasibility of using non-contact facial IRT information for CAD prediction.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The current conventional approaches for initial coronary artery disease (CAD) assessment in clinical practice mainly rely on pretest probability tools based on traditional risk factors and symptoms, which often exhibit limited prediction performance.

⇒ Infrared thermography (IRT), a non-contact technology that captures surface temperature, has shown promising potential in assessing various atherosclerosis-related conditions but has not yet been evaluated for its clinical feasibility in predicting CAD.

WHAT THIS STUDY ADDS

⇒ For suspected individuals referred for confirmatory CAD evaluation, we demonstrated that human facial temperature information captured by the non-contact IRT can be effectively used by advanced machine learning algorithms for predicting CAD.

⇒ Both an end-to-end, deep-learning-based facial IRT image analysis approach and an interpretable facial temperature variable extraction approach exhibited superior performance for CAD prediction, compared with conventional clinical methods.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Novel biophysiological information from facial temperature offers the possibility of real-time, non-contact CAD detection, which could potentially be adopted in clinical practice to improve the accuracy of CAD assessment and optimise the current clinical workflow.

INTRODUCTION

Coronary artery disease (CAD) is the leading cause of mortality and imposes a significant disease burden worldwide.[1] Accurate CAD assessment is crucial to inform appropriate downstream care. Current guidelines rely on pretest probability (PTP) tools to estimate CAD probability in suspected patients.[2 3] However, these tools suffer from issues of subjectivity, modest precision and limited generalisability.[3–5] Although supplementary cardiovascular examinations such as electrocardiography and coronary artery calcium (CAC) score, or complex clinical models incorporating additional risk factors

of comorbidities and laboratory markers, could improve CAD probability estimation, they often present challenges regarding procedural complexity, time efficiency and limited availability.[6–10] Therefore, there is a need for more accurate CAD prediction tools that efficiently integrate these different aspects of additional CAD-related information.

Infrared thermography (IRT) is a non-contact, real-time imaging technology that captures temperature distribution and variations on the object's surface by detecting self-emitted infrared radiation.[11] This non-invasive approach has emerged as a promising tool for disease assessment, as it can identify areas of abnormal blood circulation and inflammation activity through the measurement of skin temperature patterns. Studies in recent years have revealed strong associations between human body IRT temperature information and various conditions related to atherosclerotic cardiovascular disease (ASCVD), including carotid and peripheral artery diseases (PAD),[12 13] diabetes,[14] hyperlipidaemia,[15] metabolic syndrome[16] and inflammatory conditions.[17 18] Among these studies, the human face has received particular attention due to its convenience and the previously reported link between human facial features and CAD risk.[19 20] However, previous IRT studies have used simplistic, low-dimensional IRT information extracted and analysed with conventional statistical methods, which limited their ability to objectively and comprehensively quantify and use the wealth of information contained in IRT images. The advent of machine learning (ML) technology to extract, process and integrate complex information has shown impressive capability in harnessing the myriad of imaging information for various disease predictions.[21–23] Therefore, we hypothesised that the IRT information measured from human faces, with the aid of ML technology, could be fully used for CAD prediction in a non-contact manner.

This study aims to investigate the feasibility of using non-contact captured facial IRT temperature information for CAD prediction.

METHODS

Study design and participants

This is a prospective, single-centre, cross-sectional study (ClinicalTrials.gov Identifier: NCT04941560). Eligible adult participants undergoing invasive coronary angiography (ICA) or coronary CT angiography (CCTA) at the National Center for Cardiovascular Disease, Fuwai Hospital were enrolled (detailed inclusion and exclusion criteria in online supplemental method S1). Informed consents were obtained from all eligible patients, with permission to use their facial IRT images, as well as required medical record data, for research-only deidentified analysis. Our study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis reporting guideline (online supplemental table S1).[24]

Data collection

Trained clinical researchers collected baseline information and conducted IRT filming. The participants' presenting complaints, lifestyles, socioeconomic status, medical and family history, and medication usage were documented. The IRT filming was conducted in a confined room with air conditioning-controlled environmental temperature prior to the ICA or CCTA examination. Participants were seated in a stationary position, looking horizontally and naturally at an IRT camera (FLIR A315, FLIR Systems, USA) fixed at a distance of 1.5 m. The IRT filming commenced after proper positioning and alignment of the participant's face and a 3 min resting period. The entire filming process lasted for at least 5 s with the participant maintaining a still and centred position in the IRT capturing frame. Further demographic information, clinical history and risk factors, baseline blood biochemistry results and confirmatory CAD workup findings were obtained by reviewing participants' electronic medical records.

Data preparation and labelling

For each participant, one facial IRT image was selected and underwent preparation procedures before analyses, including greyscale conversion, background cropping and uniform resizing (online supplemental method S2). The prediction of interest in this study is the presence of CAD or not, as evidenced by ICA or CCTA findings, defined as a coronary lesion stenosis ≥50%. Two interventional cardiologists or radiologists, blinded to the study design and patient information, independently reviewed ICA or CCTA findings to evaluate the presence and/or degree of CAD lesions. Discrepancies were resolved through a third reviewer invited for final consensus.

Clinical and IRT image models for CAD prediction

To develop and evaluate CAD prediction models, we performed five repetitions of fivefold cross-validations with random shuffling.

i. IRT image model: We employed an advanced deep-learning algorithmic framework optimised for relatively small-sample training while effectively leveraging relevant information to achieve satisfactory prediction performance. This framework comprises two essential components: the contrastive language-image pretraining image encoder, known for its exceptional zero-shot capabilities in extracting high-fidelity image features without task-specific training[25] and a vision transformer layer incorporating self-attention mechanisms to capture global context and relationships within the image for better integration of local and global features.[26] Additionally, a single fully connected layer served as the final classifier (detailed algorithm description and training process in online supplemental method S3; algorithm framework in online supplemental figure S1). This streamlined framework operated in an end-to-end manner for CAD prediction based on one single IRT image.

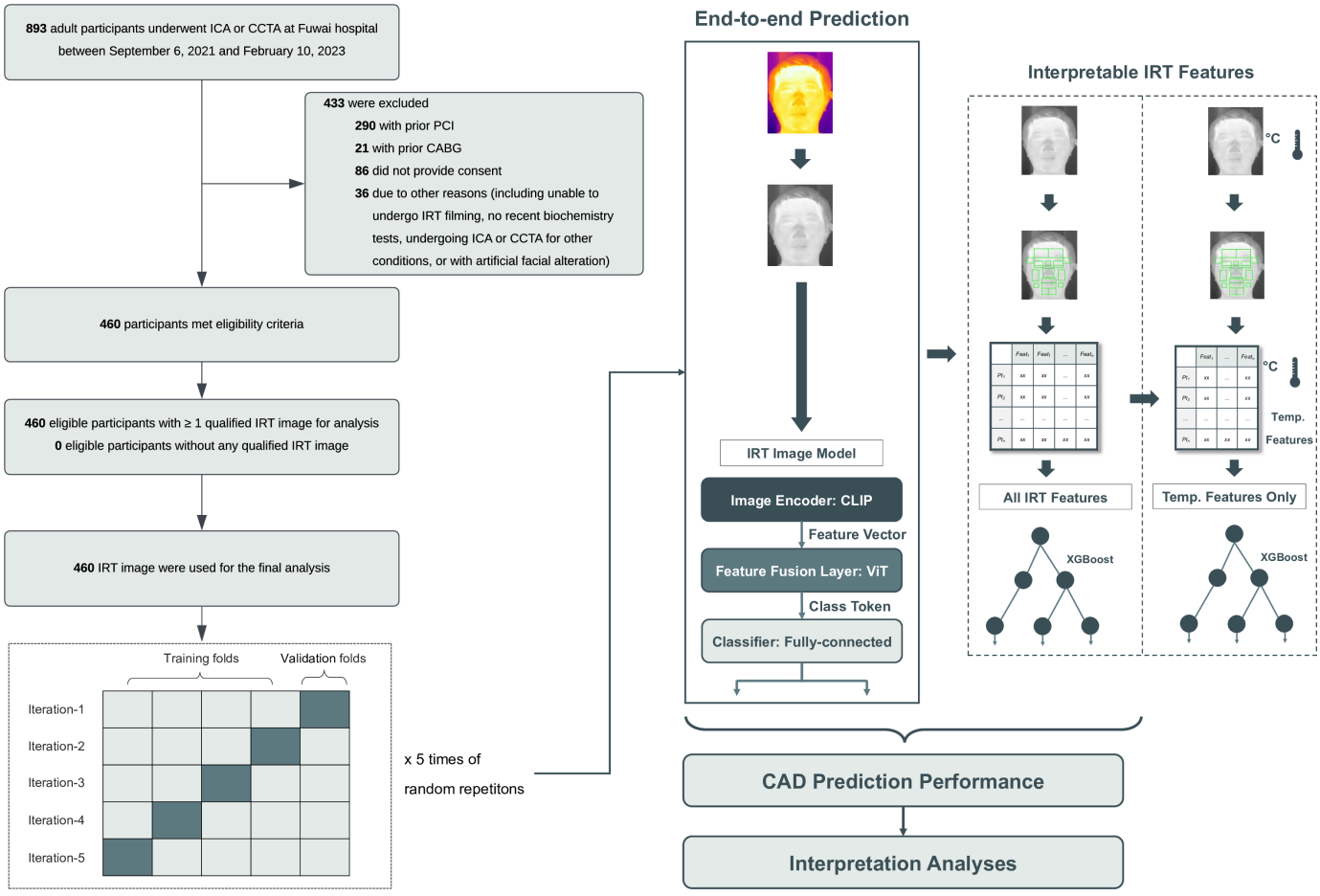

Figure 1 Flow chart of the study dataset and design. CABG, coronary artery bypass grafting; CAD, coronary artery disease; CCTA, coronary CT angiography; CLIP, contrastive language-image pretraining; ICA, invasive coronary angiography; IRT, infrared thermography; PCI, percutaneous coronary intervention; Temp., temperature; ViT, vision transformer.

ii. Models with clinical variables: Two CAD prediction models with clinical information were constructed for comparison with the IRT image model. (1) The guideline-recommended PTP model for CAD prediction, which requires the patient's age, sex and presenting symptom characteristics,[3 7] served as the clinical baseline for predicting CAD. (2) A hybrid model that incorporated both clinical and IRT information. Specifically, this model fused the clinical variables from the PTP model with the IRT information from the IRT image model, in order to assess whether there was any additional performance improvement from this joint data input.

IRT image model interpretation

To enhance our understanding of how IRT information contributes to CAD prediction, we conducted a series of interpretation analyses to gain insights into the IRT image model:

i. Occlusion experiments: To quantify the contribution of different IRT facial regions to model's predictions, we sequentially occluded the corresponding region of interest (ROI) for each of the 10 facial regions. We then measured the individual impact of each occlusion on the model's performance.

ii. Saliency map visualisation: The gradient-weighted class activation map (Grad-CAM) method was employed to visually identify key areas in each facial IRT image that the algorithm focuses on for CAD prediction (online supplemental method S4).[27]

iii. Dose–response analyses: To explore the potential causal relationship between facial IRT information and CAD status, we investigated the association between individuals' CAD risk predicted by the IRT model and the CAD lesion severity.

iv. CAD surrogate label prediction: To further explore potential mechanisms by which IRT information may contribute to CAD prediction, we hypothesised that the IRT model's predictive potential may derive from identifying various CAD-contributing or related aspects, represented by surrogate labels of ASCVD risk factors and other cardiovascular or inflammation markers. We tested this hypothesis by evaluating the performance of IRT models in predicting these surrogate labels.

Interpretable IRT features for CAD prediction

To further validate our hypothesis regarding the predictive value of IRT information for CAD and to obtain more human-interpretable insights, we extracted a diverse

Table 1 Baseline characteristics

	Overall (n=460)	CAD (n=322)	No CAD (n=138)	P value
Age, mean (SD)	58.4 (10.4)	60.4 (9.7)	53.8 (10.6)	<0.001
Female sex, n (%)	126 (27.4)	74 (23.0)	52 (37.7)	0.002
Smoking, n (%)	219 (47.6)	177 (55.0)	42 (30.4)	<0.001
BMI, mean (SD)	25.5 (3.0)	25.6 (3.0)	25.2 (3.0)	0.155
Menopause, n (%)	107 (84.9)	71 (95.9)	36 (69.2)	<0.001
Early ASCVD family history, n (%)	18 (3.9)	15 (4.7)	3 (2.2)	0.128
Hypertension, n (%)	267 (58.0)	215 (66.8)	52 (37.7)	<0.001
Hyperlipidaemia, n (%)	348 (75.7)	295 (91.6)	53 (38.4)	<0.001
Diabetes mellitus, n (%)	112 (24.3)	96 (29.8)	16 (11.6)	<0.001
Cerebrovascular event, n (%)	67 (14.6)	59 (18.3)	8 (5.8)	0.001
Peripheral artery disease, n (%)	48 (10.4)	44 (13.7)	4 (2.9)	0.001
Congestive heart failure, n (%)	63 (13.7)	32 (9.9)	31 (22.5)	0.001
Chronic kidney disease, n (%)	5 (1.1)	4 (1.2)	1 (0.7)	1.00
COPD, n (%)	7 (1.5)	5 (1.6)	2 (1.4)	1.00
Atrial Fibrillation, n (%)	35 (7.6)	21 (6.5)	14 (10.1)	0.250
Chronic inflammatory disease, n (%)	18 (3.9)	14 (4.3)	4 (2.9)	0.637
CAD symptoms, n (%)				
No symptoms	77 (16.7)	42 (13.0)	35 (25.4)	0.002
Non-anginal	102 (22.2)	70 (21.7)	32 (23.2)	
Atypical	146 (31.7)	102 (31.7)	44 (31.9)	
Typical	135 (29.3)	108 (33.5)	27 (19.6)	
Regular medications				
Aspirin, n (%)	191 (41.5)	173 (53.7)	18 (13.0)	<0.001
Beta blocker, n (%)	116 (25.2)	92 (28.6)	24 (17.4)	0.016
Statin, n (%)	210 (45.7)	173 (53.7)	37 (26.8)	<0.001
Nonstatin lipid-lowering drugs, n (%)	11 (2.4)	7 (2.2)	4 (2.9)	0.740
ACEI/ARB, n (%)	125 (27.2)	103 (32.0)	22 (15.9)	0.001
CCB, n (%)	121 (26.3)	94 (29.2)	27 (19.6)	0.042
Fast glucose, mean (SD)	6.3 (2.0)	6.5 (2.2)	5.7 (1.3)	<0.001
Total cholesterol, mean (SD)	4.3 (1.2)	4.2 (1.2)	4.7 (1.1)	<0.001
Triglyceride, mean (SD)	1.7 (1.7)	1.7 (1.9)	1.5 (0.9)	0.058
HDL, mean (SD)	1.2 (0.3)	1.2 (0.3)	1.3 (0.3)	<0.001
LDL, mean (SD)	2.5 (1.0)	2.4 (0.9)	2.9 (1.0)	<0.001
Haemoglobin A1c%, mean (SD)	6.3 (1.2)	6.4 (1.2)	5.9 (0.7)	<0.001
ESR, mean (SD)	8.0 (9.6)	8.3 (10.2)	6.7 (5.8)	0.069
CRP, mean (SD)	3.6 (5.2)	3.7 (5.5)	3.0 (3.4)	0.231
LVEF, mean (SD)	63.2 (6.2)	62.5 (6.6)	65.1 (4.5)	<0.001
Coronary confirmatory exam, n (%)				<0.001
ICA	379 (82.4)	310 (96.3)	69 (50.0)	
CCTA	81 (17.6)	12 (3.7)	69 (50.0)	
Coronary Lesion severity, n (%)				<0.001
No coronary stenosis >50%	138 (30.0)	/	138 (100.0)	
One vessel	89 (19.3)	89 (27.6)	/	
Two vessels	74 (16.1)	74 (23.0)	/	
Left main or three or more vessels	159 (34.6)	159 (49.4)	/	

ACEI/ARB, ACE inhibitor or angiotensin receptor blocker; ASCVD, atherosclerotic cardiovascular diseases; BMI, body mass index; CAD, coronary artery disease; CCB, calcium channel blocker; CCTA, coronary CT angiography; COPD, chronic obstructive pulmonary disease; CRP, C reactive protein; ESR, erythrocyte sedimentation rate; HDL, high-density lipoprotein; ICA, invasive coronary angiography; LDL, low-density lipoprotein; Lp(a), lipoprotein(a); LVEF, left ventricular ejection fraction.

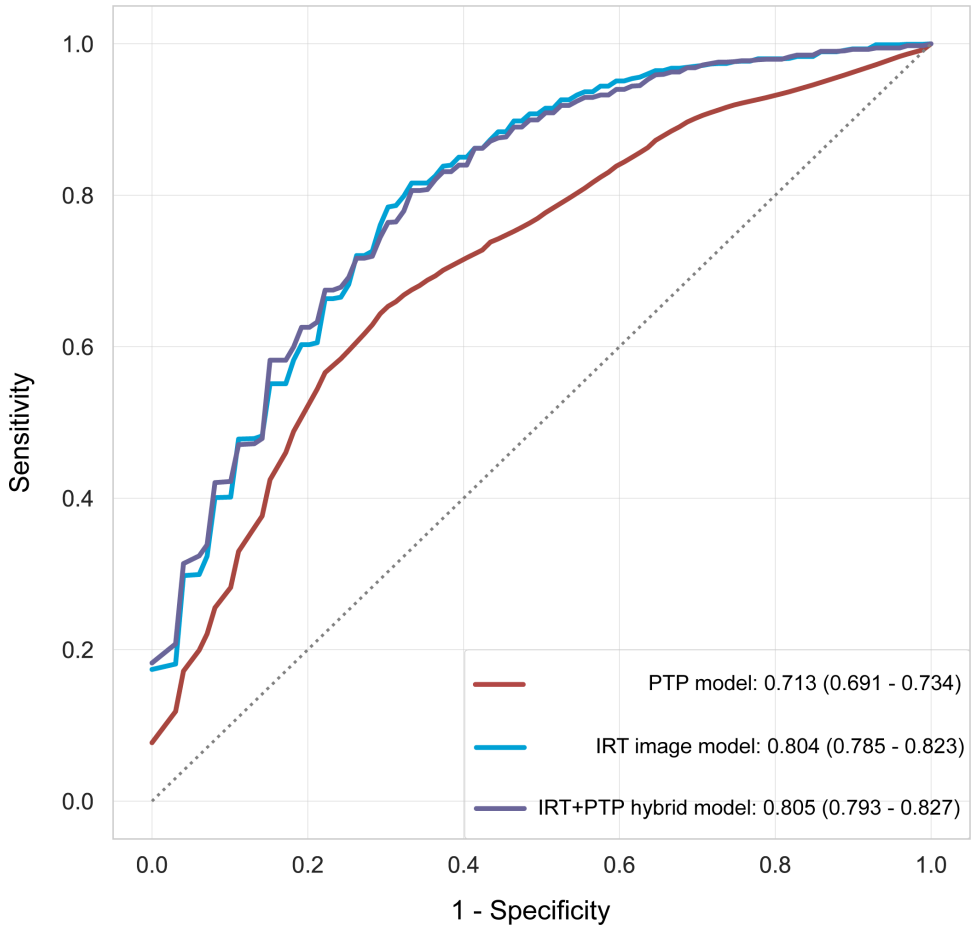

Figure 2 Receiver operating characteristic curves of models performance for CAD prediction. The legend in the right lower corner indicates different CAD prediction models and their corresponding AUC estimates, as well as the 95% CIs. AUC, area under the curve; CAD, coronary artery disease; IRT, Infrared thermography; PTP, pretest probability.

set of IRT tabular features from the IRT image. These features served as purer and more intuitive representations of underlying IRT information, reflecting facial temperature distribution. These extracted IRT features were categorised into two main levels: whole-face level and ROI-specific level. At the ROI-specific level, we partitioned the image into 18 facial ROIs (online supplemental method S5) and extracted features, respectively, resulting in a total of 619 ROI-specific IRT features. In addition, nine features were extracted at the whole-face level. A total of 628 IRT features encompassed four categories, namely: traditional temperature features, first-order texture features, second-order texture features and the fractal analysis feature (detailed description and a complete list of IRT features in online supplemental method S5 and table S2).

We employed the XGBoost algorithm, a gradient-boosted decision tree approach,[28] to integrate these extracted interpretable IRT features and assess their predictive values for CAD. We evaluated the performance of two approaches: one using all the interpretable IRT features and the other using only the traditional temperature features. The former comprehensive IRT feature approach aimed to approximate as much volume of IRT information as that used in the end-to-end IRT image model. Whereas, the traditional temperature feature-only approach was to explore the predictive values of traditional temperature variables, which can be more readily available in clinical practice even if an IRT camera is not readily accessible. We further leveraged the feature importance functionality inherent in tree-based ML models to obtain rankings of individual facial IRT features, which assigned importance scores to each feature based on their contributions to the overall model performance.

Statistical analysis

Data are presented as mean with SD or median with IQR for continuous variables, and percentages for categorical variables. Student's t-test or Wilcoxon rank-sum test was used to compare continuous variables, while the χ^2 test or Fisher's exact test was used for categorical variables. The model's discrimination performance was evaluated by area under the curve (AUC) with 95% CIs. All comparisons were two sided, with statistical significance defined as $p<0.05$, without adjustment for multiple comparisons. MATLAB V.R2021b (MathWorks, Massachusetts, USA) and Python V.3.10.5 were used for data preprocessing and model development, and R V.4.0.3 (R Foundation for Statistical Computing, Vienna, Austria) was used for plotting and statistical analysis.

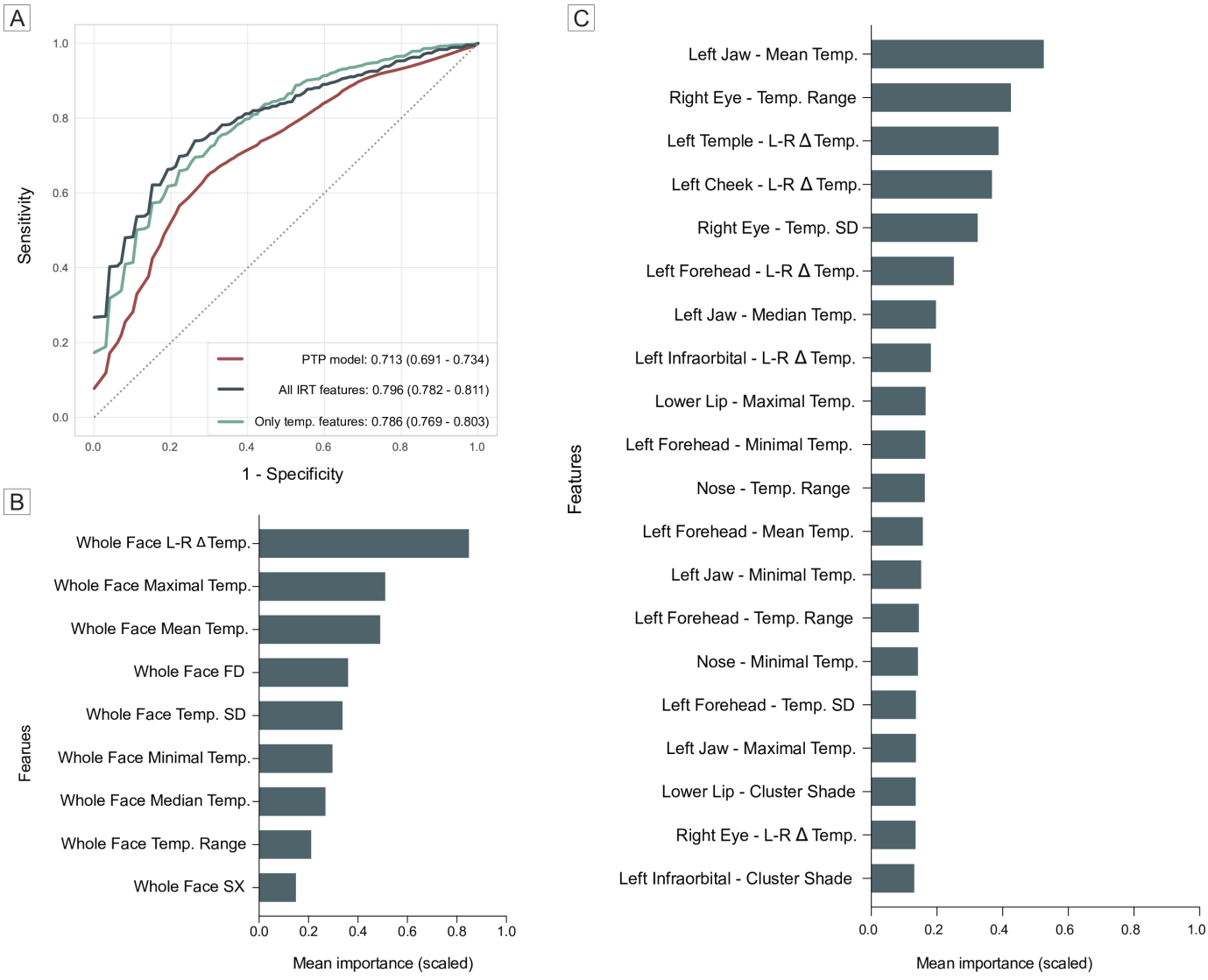

Figure 3 Analyses of the interpretable IRT features for coronary artery disease (CAD) prediction. (A) Predictive performance for using all or traditional temperature-only IRT features for CAD prediction, as compared with the PTP model; (B) the ranking of the scaled importance value of the whole-face level features; (C) the ranking of the scaled importance value of the top 20 region of interest-level features. FD, fractal dimension; IRT, infrared thermography; L-R Δ, left-right difference; PTP, pretest probability; SX, sum of extrema; Temp., temperature; Δ, value difference.

RESULTS

Study participant overview

Between 6 September 2021 and 10 February 2023, a total of 893 adult participants undergoing ICA or CCTA evaluation were screened. After excluding 433 individuals according to study criteria, 460 eligible participants were included. All participants underwent standard IRT filming, and their image quality was assessed, with all participants having at least one qualified IRT image, constituting the final analysis dataset (figure 1). Among this final dataset (460 participants with corresponding 460 IRT images), the mean age was 58.4 (SD 10.4) and 126 individuals (27.4%) were female. A total of 322 participants (70.0%) were confirmed to have CAD. Table 1 presents the baseline characteristics between CAD and non-CAD participants. Compared with non-CAD participants, those with CAD were older, more likely to be male, had a greater prevalence of lifestyle, clinical and laboratory risk factors for CAD, as well as more frequent use of primary prevention medications.

CAD prediction model performance

The performance of the individual CAD prediction models in the validation sets under the current five-repeated fivefold cross-validation design is summarised in online supplemental table S3. In comparison to the guideline-recommended PTP model (AUC 0.713, 95% CI 0.691 to 0.734), the IRT image model exhibited a considerably higher performance (AUC 0.804, 95% CI 0.785 to 0.823). Furthermore, when integrating clinical variables from the PTP models with the IRT image as joint input, the resulting IRT-PTP hybrid model (AUC 0.805, 95% CI 0.793 to 0.827) did not yield a significant difference in

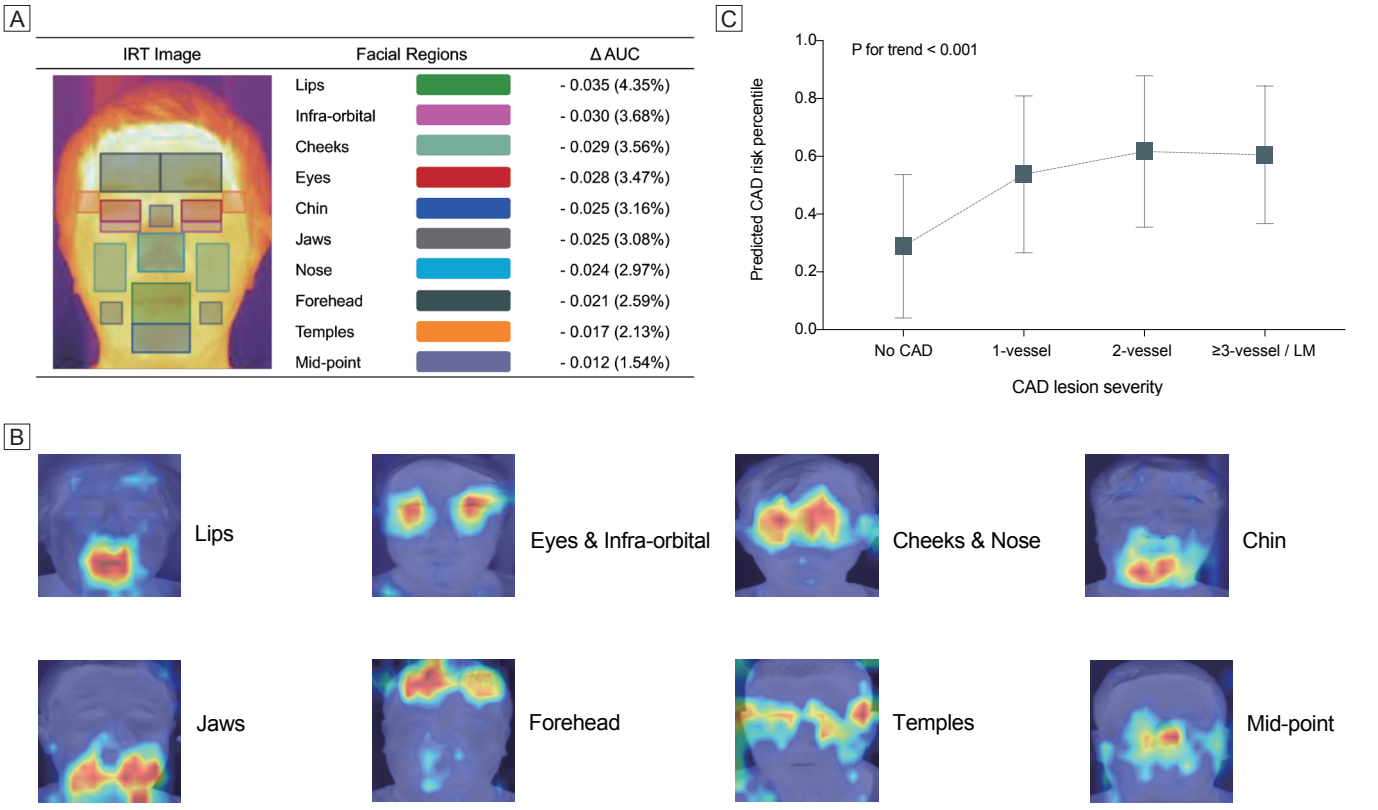

Figure 4 Interpretation and visualisation of the IRT image model. (A) Results of the occlusion tests in assessing the effect of individual facial regions after occlusion on the IRT image model's predictive performance, measured by the degree of AUC reduction (ΔAUC); (B) visualisation of examples with specific facial regions deemed important for IRT image model prediction highlighted by the Gradient-weighted Class Activation Map methods; (C) dose–response relationship between the CAD lesion severity and the IRT image model predicted CAD risk percentiles. AUC, area under the curve; CAD, coronary artery disease; IRT, infrared thermography; LM, left main.

performance improvement compared with the IRT image model alone (figure 2).

Interpretable IRT features for CAD prediction

Based on the manually extracted interpretable IRT features for further validation, both the all IRT feature approach (AUC 0.796, 95% CI 0.782 to 0.811) and the traditional temperature feature-only approach (AUC 0.786, 95% CI 0.769 to 0.803) demonstrated superior performance (figure 3A), which closely aligned with the performance of the end-to-end IRT image model in utilisation of IRT information for CAD prediction.

The relative importance rankings of the interpretable IRT features for CAD prediction are depicted in figure 3B,C. At the whole-face level (figure 3B), of the three most significant features, the most influential one was the overall left-right temperature difference, followed by the maximal facial temperature, mean facial temperature and fractal dimension of facial temperature. Among the three most influential ROI-specific features (figure 3C), the mean temperature of the left jaw region exhibited the highest impact, followed by the temperature range of the right eye region and the left-right temperature difference of the left temple regions.

Interpretation of the IRT image model

The occlusion experiments (figure 4A) demonstrated varying degrees of reduction in the IRT image model performance when occluding different ROIs for any of the 10 facial regions. The largest decrease was observed when occluding the upper and lower lips (ie, the oral and perioral) region (ΔAUC=−0.035, 4.35%), followed by the left and right infraorbital (ΔAUC=−0.030, 3.68%) and cheeks (ΔAUC=−0.029, 3.56%), etc. In addition, examples of facial regions in the IRT image deemed important for the IRT image model prediction were visualised using the Grad-CAM method (figure 4B). Moreover, a trend of higher predicted CAD risk percentile was observed as CAD severity increased (figure 4C).

Table 2 presents the potential of the modified IRT image model to predict various surrogate labels associated with CAD. For ASCVD traditional risk factors, the image model demonstrated good performance in identifying hyperlipidaemia (0.831, 95% CI 0.811 to 0.850), male sex (0.988, 95% CI 0.985 to 0.991), smoking (0.749, 95% CI 0.694 to 0.804), body mass index (mean absolute error (MAE) 2.593, 95% CI 2.147 to 3.038), HbA1C% (MAE 0.772, 95% CI 0.686 to 0.859), etc. Furthermore, the model also exhibited potential in identifying other cardiovascular (eg, NT-proBNP>300 pg/mL, 0.636

Table 2 IRT model prediction for surrogate labels contributing or related to CAD

Surrogate labels	AUC (95% CI)	MAE (95% CI)
ASCVD traditional risk factors		
Hyperlipidaemia	0.831 (0.811 to 0.850)	/
Hypertension	0.640 (0.607 to 0.673)	/
Diabetes mellitus	0.659 (0.573 to 0.745)	/
Male	0.988 (0.985 to 0.991)	/
Age	/	8.23 (7.543 to 8.914)
Body mass index	/	2.593 (2.147 to 3.038)
Smoking	0.749 (0.694 to 0.804)	/
Early ASCVD family history	0.691 (0.587 to 0.795)	/
HbA1C%	/	0.772 (0.686 to 0.859)
Inflammation and other cardiovascular markers		
Chronic inflammatory diseases	0.631 (0.536 to 0.726)	/
Elevated ESR level*	0.645 (0.524 to 0.766)	/
Elevated Inflammatory Markers†	0.601 (0.539 to 0.663)	/
NT-proBNP>300 pg/mL	0.636 (0.593 to 0.678)	/

*The elevated level refers to the laboratory value higher than the upper bound of reporting normal range.
†Inflammatory markers include ESR, C reactive protein and Interleukin-6.
ASCVD, atherosclerotic cardiovascular diseases; AUC, area under the curve; CAD, coronary artery disease; CRP, C reactive protein; ESR, erythrocyte sedimentation rate; HbA1C%, Hemoglobin A1C%; IRT, infrared thermography; MAE, mean absolute error; NT-proBNP, N-terminal pro-B-type natriuretic peptide.

(95% CI 0.593 to 0.678)) and inflammation-related labels (eg, chronic inflammatory diseases, 0.631 (95% CI 0.536 to 0.726), elevated erythrocyte sedimentation rate, 0.645 (95% CI 0.524 to 0.766), etc).

DISCUSSION

In this study, we have demonstrated the feasibility of using IRT temperature information from human faces to predict CAD in a non-contact manner. Our developed deep-learning IRT image model for CAD prediction achieved superior performance compared with the current guideline-recommended PTP model that relied on traditional risk factors and clinical presentation for CAD assessment. The current findings highlighted the promising potential of facial temperature information in CAD assessment, which could be harnessed through either the end-to-end IRT image-based deep-learning approach or through a more interpretable temperature variable approach in clinical practice (figure 5).

The feasibility of IRT information for CAD prediction was built on previous evidence between IRT and ASCVD-related conditions. For ASCVD risk factors, previous studies demonstrated that combining temperature and textural features from facial IRT images with clinical risk factors achieved high prediction accuracy for type II diabetes.[14] Associations were also found between body surface temperature measured by IRT in specific regions and blood lipid levels.[15] Distinct IRT distribution patterns, especially temperature asymmetry, have also been observed in individuals at high risk or with established CAD.[29] Inflammation, an increasingly recognised non-traditional risk factor contributing to ASCVD,[30–32] has also been reflected in IRT images in various chronic inflammatory conditions.[17 18] Therefore, it is possible that IRT information reflective of inflammation activity could be used in ASCVD prediction and evaluation. The potential of IRT in assessing established ASCVD diseases has also been explored in previous studies, including PAD from IRT measurements in peripheral extremities[13] and carotid atherosclerosis detected by IRT obtained from neck and facial regions.[12 33] In addition, studies have also investigated the dynamic temperature changes captured through IRT to reflect vascular function, which was further shown to be well correlated with ASCVD risk, CAC score and myocardial perfusion defects.[34–36] However, previous studies generally employed simplistic approaches for IRT information extraction and analysis, which could limit their ability to comprehensively and objectively integrate the full breadth of IRT information for disease assessment. In our study, we conducted surrogate label prediction experiments to replicate and validate these previous findings. The observed overall strong performance of our IRT models in predicting these CAD-related surrogate labels further strengthens the pathophysiological plausibility and validity of facial IRT information for CAD prediction.

Internal validity and interpretability were prioritised in establishing the feasibility of IRT models in predicting CAD in the current study. The IRT image model employed a state-of-the-art deep-learning framework, allowing for

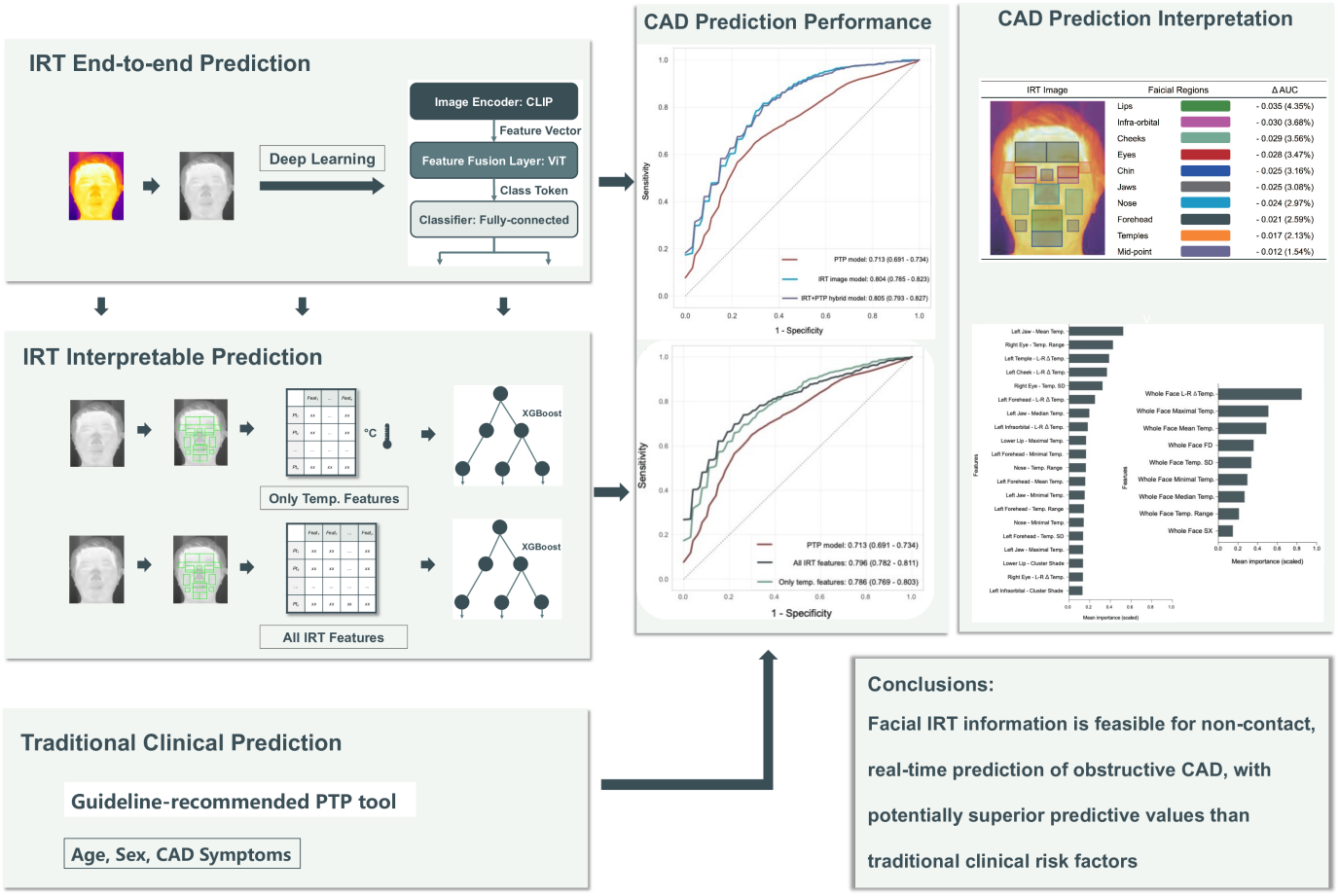

Figure 5 Central illustration. CAD, coronary artery disease; CLIP, contrastive language-image pretraining; FD fractal dimension; IRT, infrared thermography; L-R Δ, left-right difference; PTP, pretest probability; SX, sum of extrema; Temp., temperature; ViT, vision transformer; Δ, value difference.

robust extraction of high-fidelity image features and reliable prediction for our specific downstream task, even with a relatively small training sample size. Notably, the addition of clinical variables to the IRT image model did not yield further improvements compared with the standalone end-to-end IRT image-based approach, suggesting that the facial IRT information extracted by the algorithm may already encompass relevant clinical information associated with CAD. Model interpretation also confirmed that the deep-learning algorithm focused on potentially relevant facial IRT areas and helped identify important facial regions contributing to predictions. Furthermore, the observed dose–response relationship between predicted CAD risk and CAD severity further bolstered the model's credibility. The predictive value of IRT information for CAD was further validated by the interpretable IRT tabular features, which could also avoid potential inclusion of irrelevant image details that might give away the prediction label and thus inflate performance.[37] Importantly, this interpretable IRT tabular feature-based approach demonstrated relatively consistent performance as the deep-learning IRT image model. With these human-interpretable IRT features, we also gained insights into specific aspects of facial IRT temperature information deemed important for the CAD predictions, with prominent aspects such as facial temperature asymmetry and distribution non-uniformity.

The feasibility of IRT temperature-based CAD prediction suggests potential future applications and research opportunities. As a biophysiological-based health assessment modality, IRT provides disease-relevant information beyond traditional clinical measures that could enhance ASCVD and related chronic condition assessment. The non-contact, real-time nature of the end-to-end IRT image model allows for instant disease assessment at the point of care, which could streamline clinical workflows and save time for important physician–patient decision-making. In addition, it has the potential to enable mass prescreening for more cost-effective adoption of downstream screening modalities (eg, CAC score). Deploying IRT-based assessment in a non-contact and passive monitoring manner could also enable continuous evaluation of disease progression in the daily living spaces outside of regular clinic visits.[38] Depending on resource availability, the temperature-based CAD assessment could be adopted accordingly with satisfactory performance, from the more widely available traditional temperature features that could be measured with regular thermometer, to the end-to-end IRT-based imaging approach that uses validated IR cameras with good reproducibility and minimal operator

training. Importantly, IR temperature-based prediction tools have several inherent advantages that enhance their trustworthiness for healthcare providers, including its physiologically sound mechanism, high reproducibility and user-friendly operation.

Several limitations should be acknowledged in the current study. First, the relatively small sample size may have limited the performance of current IRT algorithms. To address this limitation, we employed ML algorithms with simplistic structure optimised for small-sample prediction tasks, which minimised the training requirements while still achieving valid and satisfactory performance. Second, the study was conducted in a single-centre cohort, necessitating external validation from diverse patient populations in multicentre studies. Lastly, the study participants were patients referred for confirmatory CAD examinations, and therefore, represented a higher PTP spectrum, which could limit the generalisability of current findings. Future research should include a broader spectrum of patients for CAD evaluation.

CONCLUSION

In this diagnostic study, we have examined and established the feasibility of using non-contact captured human facial temperature information by IRT in predicting CAD. Our developed IRT prediction models, based on advanced ML technology, have exhibited promising potential compared with the current conventional clinical tools. Further investigations incorporating larger sample sizes and diverse patient populations are needed to validate the external validity and generalisability of current findings.

Author affiliations
[1]Department of Automation, Tsinghua University, Beijing, China
[2]Shenzhen International Graduate School, Tsinghua University, Shenzhen, Guangdong, China
[3]National Clinical Research Center of Cardiovascular Diseases, Fuwai Hospital, National Center for Cardiovascular Diseases, Beijing, China
[4]State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center for Cardiovascular Diseases, Beijing, China
[5]Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
[6]Department of Cardiovascular Surgery, Fuwai Hospital, National Center for Cardiovascular Diseases, Beijing, China
[7]Key Laboratory of Coronary Heart Disease Risk Prediction and Precision Therapy, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
[8]Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University, Beijing, China

Acknowledgements We thank Yue Zhang for her assistance with patient enrolment and logistics, and all participating staff members in the catheterisation laboratory and the radiology department for their assistance in the collection and interpretation of ICA and CCTA.

Contributors ZZ, XJ and SL conceived the overall study. JZ and SL designed the experiments. JZ, RS, SL, SY, XL and YZ performed the data acquisition, extracting and cleaning. MK, JZ and XY performed the data processing and conducted the experiments. MK, XY, CL and MS designed and implemented the algorithm. JZ, MK and XS analysed the data. ZZ and XJ directed the project. All authors contributed to the interpretation of the results and JZ drafted the final manuscript, which was reviewed, revised and approved by all authors. ZZ is the guarantor of this study.

Funding This work was supported by grants from the National High Level Hospital Clinical Research Funding of Fuwai Hospital, Chinese Academy of Medical Sciences (2022-GSP-GG-28 to ZZ, 2022-GSP-QN-10 and 2023-GSP-RC-10 to SL), the National Natural Science Foundation of China (Key Programme; 81830072 to ZZ), and the National Natural Science Foundation of China (61827804 and 62131011 to XJ).

Disclaimer The funders had no role in study design, data collection, data analysis, data interpretation, writing of the report or the decision to submit the paper for publication.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by name of the ethics committee: Ethics Committee of Fuwai Hospital, CAMS and PUMC, reference number of ethics approval: 2021-1471. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; internally peer reviewed.

Data availability statement Data are available on reasonable request. The data collected and analysed in the current study cannot be shared publicly due to patient privacy. All data were approved for research purposes in Fuwai hospital only. Any external use requires additional consent and ethical approval from Fuwai hospital institutional review board and may be available from the corresponding author on reasonable request.

ORCID iDs
Juntong Zeng http://orcid.org/0000-0003-1235-3180
Zhe Zheng http://orcid.org/0000-0002-9162-6492

REFERENCES

1 Vos T, Lim SS, Abbafati C, *et al*. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the global burden of disease study 2019. *Lancet* 2020;396:1204–22.
2 Fihn SD, Blankenship JC, Alexander KP, *et al*. 2014 ACC/AHA/AATS/PCNA/SCAI/STS focused update of the guideline for the diagnosis and management of patients with stable ischemic heart disease: a report of the American college of cardiology/American heart association task force on practice guidelines, and the American association for thoracic surgery, preventive cardiovascular nurses association, society for cardiovascular angiography and interventions, and society of thoracic surgeons. *J Am Coll Cardiol* 2014;64:1929–49.
3 Knuuti J, Wijns W, Saraste A, *et al*. 2019 ESC guidelines for the diagnosis and management of chronic coronary syndromes. *Eur Heart J* 2020;41:407–77.
4 Juarez-Orozco LE, Saraste A, Capodanno D, *et al*. Impact of a decreasing pre-test probability on the performance of diagnostic tests for coronary artery disease. *Eur Heart J Cardiovasc Imaging* 2019;20:1198–207.
5 Cheng VY, Berman DS, Rozanski A, *et al*. Performance of the traditional age, sex, and angina typicality-based approach for estimating pretest probability of angiographically significant coronary artery disease in patients undergoing coronary computed tomographic angiography: results from the multinational coronary CT angiography evaluation for clinical outcomes: an international multicenter registry (CONFIRM). *Circulation* 2011;124:2423–32, .

6 Genders TSS, Steyerberg EW, Hunink MGM, *et al*. Prediction model to estimate presence of coronary artery disease: retrospective pooled analysis of existing cohorts. *BMJ* 2012;344:e3485.
7 Reeh J, Therming CB, Heitmann M, *et al*. Prediction of obstructive coronary artery disease and prognosis in patients with suspected stable angina. *Eur Heart J* 2019;40:1426–35.
8 Fordyce CB, Douglas PS, Roberts RS, *et al*. Identification of patients with stable chest pain deriving minimal value from noninvasive testing: the PROMISE minimal-risk tool, a secondary analysis of a randomized clinical trial. *JAMA Cardiol* 2017;2:400–8.
9 Budoff MJ, Mayrhofer T, Ferencik M, *et al*. Prognostic value of coronary artery calcium in the PROMISE study (prospective multicenter imaging study for evaluation of chest pain). *Circulation* 2017;136:1993–2005.
10 Villines TC, Hulten EA, Shaw LJ, *et al*. Prevalence and severity of coronary artery disease and adverse events among symptomatic patients with coronary artery calcification scores of zero undergoing coronary computed tomography angiography: results from the CONFIRM (coronary CT angiography evaluation for clinical outcomes: an international multicenter) registry. *J Am Coll Cardiol* 2011;58:2533–40.
11 Tattersall GJ. Infrared thermography: a non-invasive window into thermal physiology. *Comp Biochem Physiol A Mol Integr Physiol* 2016;202:78–98.
12 Saxena A, Ng EYK, Lim ST. Infrared (IR) thermography as a potential screening modality for carotid artery stenosis. *Comput Biol Med* 2019;113:103419.
13 Piva G, Crepaldi A, Zenunaj G, *et al*. The value of infrared thermography to assess foot and limb perfusion in relation to medical, surgical, exercise or pharmacological interventions in peripheral artery disease: a systematic review. *Diagnostics (Basel)* 2022;12:12.
14 Thirunavukkarasu U, Umapathy S, Janardhanan K, *et al*. A computer aided diagnostic method for the evaluation of type II diabetes mellitus in facial thermograms. *Phys Eng Sci Med* 2020;43:871–88.
15 Thiruvengadam J, Anburajan M, Menaka M, *et al*. Potential of thermal imaging as a tool for prediction of cardiovascular disease. *J Med Phys* 2014;39:98–105.
16 Gao M-J, Xue H-Z, Cai R, *et al*. A preliminary study on infrared thermograph of metabolic syndrome. *Front Endocrinol (Lausanne)* 2022;13:851369.
17 Pauk J, Ihnatouski M, Wasilewska A. Detection of inflammation from finger temperature profile in rheumatoid arthritis. *Med Biol Eng Comput* 2019;57:2629–39.
18 Schiavon G, Capone G, Frize M, *et al*. Infrared thermography for the evaluation of inflammatory and degenerative joint diseases: a systematic review. *Cartilage* 2021;13:1790S–1801S.
19 Christoffersen M, Frikke-Schmidt R, Schnohr P, *et al*. Visible age-related signs and risk of ischemic heart disease in the general population: a prospective cohort study. *Circulation* 2014;129:990–8.
20 Lin S, Li Z, Fu B, *et al*. Feasibility of using deep learning to detect coronary artery disease based on facial photo. *Eur Heart J* 2020;41:4400–11.
21 Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25:44–56.
22 Zeng J, Shao J, Lin S, *et al*. Optimizing the dynamic treatment regime of in-hospital warfarin anticoagulation in patients after surgical valve replacement using reinforcement learning. *J Am Med Inform Assoc* 2022;29:1722–32.
23 Johnson KW, Torres Soto J, Glicksberg BS, *et al*. Artificial intelligence in cardiology. *J Am Coll Cardiol* 2018;71:2668–79.
24 Collins GS, Reitsma JB, Altman DG, *et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;350:g7594.
25 Radford A, Kim JW, Hallacy C, *et al*. Learning transferable visual models from natural language supervision. 2021. 8748–63. Available: https://proceedings.mlr.press/v139/radford21a.html [accessed 14 Jun 2023]
26 Dosovitskiy A, Beyer L, Kolesnikov A, *et al*. An image is worth 16X16 words: transformers for image recognition at scale. *Arxiv* 2020.
27 Selvaraju RR, Cogswell M, Das A, *et al*. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int J Comput Vis* 2020;128:336–59.
28 Chen T, Guestrin C. Xgboost: a scalable tree boosting system. Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining; 2016:785–94.
29 Choda G, Rao G. Thermal imaging for the diagnosis of early vascular dysfunctions: a case report. 2020.
30 Kaptoge S, Di Angelantonio E, Lowe G, *et al*. C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: an individual participant meta-analysis. *Lancet* 2010;375:132–40.
31 Pai JK, Pischon T, Ma J, *et al*. Inflammatory markers and the risk of coronary heart disease in men and women. *N Engl J Med* 2004;351:2599–610.
32 Ridker PM, Hennekens CH, Buring JE, *et al*. C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *N Engl J Med* 2000;342:836–43.
33 Yang Y, Liu J. Detection of atherosclerosis through mapping skin temperature variation caused by carotid atherosclerosis plaques. *J Therm Sci Eng Appl* 2011;3.
34 Gul KM, Ahmadi N, Wang Z, *et al*. Digital thermal monitoring of vascular function: a novel tool to improve cardiovascular risk assessment. *Vasc Med* 2009;14:143–8.
35 Ahmadi N, Hajsadeghi F, Gul K, *et al*. Relations between digital thermal monitoring of vascular function, the framingham risk score, and coronary artery calcium score. *J Cardiovasc Comput Tomogr* 2008;2:382–8.
36 Ahmadi N, Usman N, Shim J, *et al*. Vascular dysfunction measured by fingertip thermal monitoring is associated with the extent of myocardial perfusion defect. *J Nucl Cardiol* 2009;16:431–9.
37 Zech JR, Badgeley MA, Liu M, *et al*. Confounding variables can degrade generalization performance of radiological deep learning models. *PLoS Med* 2018;15:11.
38 Haque A, Milstein A, Fei-Fei L. Illuminating the dark spaces of Healthcare with ambient intelligence. *Nature* 2020;585:193–202.

BMJ Health & Care Informatics

Barriers and facilitators to learning health systems in primary care: a framework analysis

Georgia Fisher ,[1] Maree Saba,[1] Genevieve Dammery,[1] Louise A Ellis,[1] Kate Churruca,[1] Janani Mahadeva,[2] Darran Foo,[1,2] Simon Wilcock,[2] Jeffrey Braithwaite[1]

To cite: Fisher G, Saba M, Dammery G, *et al*. Barriers and facilitators to learning health systems in primary care: a framework analysis. *BMJ Health Care Inform* 2024;**31**:e100946. doi:10.1136/bmjhci-2023-100946

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjhci-2023-100946).

Received 23 October 2023
Accepted 01 June 2024

[1]Centre for Healthcare Resilience and Implementation Science, Australian Institute of Health Innovation, Sydney, New South Wales, Australia
[2]MQ Health General Practice, Macquarie University, Sydney, New South Wales, Australia

Correspondence to
Dr Jeffrey Braithwaite; jeffrey.braithwaite@mq.edu.au

ABSTRACT

Background The learning health system (LHS) concept is a potential solution to the challenges currently faced by primary care. There are few descriptions of the barriers and facilitators to achieving an LHS in general practice, and even fewer that are underpinned by implementation science. This study aimed to describe the barriers and facilitators to achieving an LHS in primary care and provide practical recommendations for general practices on their journey towards an LHS.

Methods This study is a secondary data analysis from a qualitative investigation of an LHS in a university-based general practice in Sydney, Australia. A framework analysis was conducted using transcripts from semistructured interviews with clinic staff. Data were coded according to the theoretical domains framework, and then to an LHS framework.

Results 91% (n=32) of practice staff were interviewed, comprising general practitioners (n=15), practice nurses (n=3), administrative staff (n=13) and a psychologist. Participants reported that the practice alignment with LHS principles was influenced by many behavioural determinants, some of which were applicable to healthcare in general, for example, some staff lacked *knowledge* about practice policies and *skills* in using software. However, many were specific to the general practice environment, for example, the *environmental context* of general practice meant that administrative staff were an integral part of the LHS, particularly in facilitating partnerships with patients.

Conclusions The LHS journey in general practice is influenced by several factors. Mapping the LHS domains in relation to the theoretical domains framework can be used to generate a roadmap to hasten the journey towards LHS in primary care settings.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The learning health system (LHS) concept is gaining traction in multiple healthcare settings yet remains relatively underexamined in primary care, particularly through the lens of implementation science.

WHAT THIS STUDY ADDS

⇒ This study uses an established implementation science framework to describe key facilitators and barriers to the cultivation of an LHS in a primary care setting.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ We compare these factors to the small existing body of literature in this area and propose practical solutions to implement the principles of the LHS into primary care practice.

BACKGROUND

Primary care is the 'frontline' of healthcare; it is the first point of contact with the health system for most people[1 2] and thus an essential component of care delivery. Primary care can reduce overall health costs and relieve pressure on other areas of the health system; for example, by reducing the number of preventable or unnecessary presentations to emergency departments.[3] In many countries, including the UK and Australia, primary care is chiefly provided through general practitioners (GPs).[4 5] However, general practice is under pressure. Ageing populations and an increase in chronic disease have heightened the demand for primary care services.[6 7] Growth in the workforce has flatlined[8 9] with fewer GPs providing care for more people,[10] and many GPs unsure of the viability of their practice.[11] More recently, the unprecedented challenge of a global pandemic has necessitated system-wide reorganisation[12] and placed many additional stresses on GPs and the system in which they work.[13] The solutions to these entrenched implementation issues are by no means easy or short term, but in the interim, general practice needs a viable framework to guide the steps towards a sustainable and high-performing primary care system. In response, the concept of a learning health system (LHS) has been proposed.[14] According to the National Academy of Medicine (NAM; then the Institute of Medicine), an LHS is a system that '*consistently delivers reliable performance and constantly improves, systematically and*

seamlessly, with each care experience—in short, a system with an ability to learn'.[15]

LHSs have been embraced by multiple providers who have reported a variety of benefits, including increases in evidence-based care delivery, improved clinical outcomes, higher levels of patient-centred care and reductions in adverse events.[16] The core characteristics of an LHS identified by the NAM include: (1) science and informatics that provide real-time access to knowledge and digitally capture care delivery; (2) patient–clinician partnerships, where patients are engaged, empowered participants in care; (3) incentives that reward high-value care and transparency; and (4) a continuous learning culture that is supported by the system and its leaders.[17] More recently, a fifth characteristic has been identified, structure and governance, that aligns policy and regulation to facilitate research, collaboration and learning.[18] An LHS can manifest at the micro level of the practice right through to the macro level of the healthcare system. This makes the LHS model well suited for primary care, and able to help support the performance of individual general practices and their interactions with the larger healthcare system. However, despite their promise for primary care, most reports describe LHS in tertiary hospital settings, with few that focus on the unique context of general practice or its providers on the frontlines of care.[19]

Even less frequent in the literature are reports of primary care LHS that are underpinned by principles of implementation science, a field that aims to establish what works and why in the translation of research evidence into practice.[20] The simplicity of the five-part LHS framework is somewhat deceptive; not only are its components multifaceted and their role unpredictable,[21] but they must also be applied in the broader complex adaptive system of healthcare.[22] Subsequently, there are many factors that affect the success of the LHS in the real world. Implementation science frameworks provide an evidence-based explanation of such factors, enabling us to leverage facilitators, and overcome barriers. An established method of doing so is via the theoretical domains framework (TDF), which brings together multiple theories of behaviour change into a single 14-item framework.[23] In the present study, we used the TDF to conduct secondary analysis of data obtained in our previous investigation of an LHS in the general practice setting. We aimed to identify and describe the barriers and facilitators to adopting LHS principles specific to each of the five components of the LHS framework, and to provide evidence-based implementation recommendations for general practices who are making the journey towards an LHS.

METHODS

This study is a secondary analysis of data generated in a qualitative investigation of an LHS in primary care.[24] Our original investigation brought together researchers from the Australian Institute of Health Innovation (AIHI) and staff from MQ Health General Practice (MQGP) in a qualitative study that used an embedded research approach and that was codesigned by the research team from AIHI, and clinicians and senior clinic administrators from MQGP.

Study setting and context

MQGP is a not-for-profit, university-based general practice that operates in the northern suburbs of Sydney, Australia across two sites: one adjacent to a hospital on the university campus, and one in a suburban location.[24] The practice is part of the broader entity of MQ Health, which also comprises specialist clinics, an inpatient hospital, and allied health, medical imaging, radiotherapy and on-site pathology services. Most MQGP staff are employees of MQ Health and have access to educational resources available to employees of Macquarie University. Due to its university affiliation, MQGP is actively involved in research and teaching activities and has a strong record of quality improvement initiatives. MQGP also participates in its local Primary Health Network (PHN), which is a government-initiated, independent organisation that aims to streamline and coordinate primary care services. At the time of the study, MQGP employed 17 GPs, 4 clinic nurses, 13 administrative staff and a clinical psychologist across both sites.

Embedded research approach

In our embedded research approach,[25] a research assistant from AIHI (GD) was introduced to all MQGP staff at a clinic practice meeting in July 2021, and then worked alongside practice staff until December 2021. The embedded researcher was included on all staff emails, liaised closely with the practice's business manager and GPs and attended the practice's 'strategy day'. The embedded researcher was also involved in the coordination and data collection of the present study.

Data collection and recruitment

We conducted semistructured interviews with MQGP staff. The research team used the modified five-characteristic NAM LHS framework[18] to design the interview questions, which were then reviewed by multiple clinical and administrative staff at MQGP to ensure their clarity and relevance to the practice. All practice staff were invited to take part in the study via email, where they were provided with participant information and consent forms that outlined the purpose of the research study. There was no sample size calculation for the study. Instead, we aimed to interview a sample that was representative of all clinic staff. Interviews were conducted in October 2021 by a senior research fellow (LE) or the embedded research assistant (GD), either in person at the general practice or via teleconference. The interviewers had prior training in qualitative research methods and interviewing.

Analysis

Interviews were audio recorded and transcribed verbatim. To deidentify the data, staff were given a unique code that consisted of their role (ADMIN, administrative staff; GP,

Table 1 Elements of the LHS and TDF frameworks

LHS components	
Science and informatics	Real-time access to knowledge and digital capture of the care experience.
Patient–clinician partnerships	Engaged, empowered patients and families that are full partners in a patient-centred system.
Incentives	Incentives aligned for value that actively encourage ongoing improvement of care and full transparency.
Continuous learning culture	Leadership-instilled culture of learning and supportive system competencies that encourage staff skill development.
Structure and governance	Policies, governance and regulations aligned to facilitate research, collaboration and learning.
TDF determinants	
Knowledge	An awareness of the existence of something.
Skills	An ability or proficiency acquired through practice.
Social and professional roles and identity	A coherent set of behaviours and displayed personal qualities of an individual in a social or work setting.
Beliefs about capabilities	Acceptance of the truth, reality or validity about an ability, talent or facility that a person can put to constructive use.
Optimism	The confidence that things will happen for the best or that desired goals will be attained.
Beliefs about consequences	Acceptance of the truth, reality or validity about outcomes of a behaviour in a given situation.
Reinforcement	Increasing the probability of a response by arranging a dependent relationship, or contingency, between the response and a stimulus.
Intentions	A conscious decision to perform a behaviour or a resolve to act in a certain way.
Goals	Mental representations of desired outcomes or end states.
Memory, attention and decision processes	The ability to retain information, focus selectively on aspects of the environment and choose between two or more alternatives.
Environmental context and resources	Any circumstance of a person's situation or environment that discourages or encourages the development of skills and abilities, independence, social competence and adaptive behaviour.
Social influences	Those interpersonal processes that can cause individuals to change their thoughts, feelings or behaviours.
Emotion	A complex reaction pattern, involving experiential, behavioural and physiological elements, by which the individual attempts to deal with a personally significant matter or event.
Behavioural regulation	Anything aimed at managing or changing objectively measured actions.

LHS framework adapted from[18]. TDF framework reproduced from [23] under the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/)
LHS, learning health system; TDF, theoretical domains framework.

general practitioner; NUR, nursing staff) and a random number. Deidentified interview transcripts were imported into NVivo V.20. The secondary analysis of study data was conducted by two members of the research team who were independent of the original study data collection (GF, MS), who conducted a deductive framework analysis[26] with the TDF and LHS framework (the Independent Analysis Team, IAT). The components of each of these frameworks are detailed in table 1.

Both members of the IAT first coded an interview transcript together in real time, categorising the data into the domains of the TDF. Then, the IAT independently coded five transcripts, iteratively checking agreement and discussing conflicts after each. After the fifth transcript, the IAT's mean±SD agreement across all TDF domains was 87.6%±10.1%. The IAT researchers then each independently coded half of the remaining transcripts. Next, they used the modified five-component NAM LHS framework[18] to organise the data in each TDF determinant; this second deductive process ensured data coded to each TDF determinant were also described in relation to the key tenets of an LHS. Counts of the number of participants who made statements coded to each TDF determinant and each LHS component were recorded. Finally, the IAT researchers met and inductively generated belief statements that were relevant to each domain of the TDF and each component of the LHS. Final results were reviewed for validity by four members of the original study team: two senior academics from AIHI (LE, JB), one GP from MQGP (DF) and the embedded research assistant (GD).

RESULTS

A total of 32 out of 35 (91%) practice staff were interviewed, which included GPs (n=15), practice nurses

Table 2 Key TDF determinants and associated belief statements for each domain of the LHS framework

	Key TDF domains		Belief statements	Exemplar quote
Science and informatics	Environmental context and resources	+	MQGP's affiliation with both Macquarie University and the PHN facilitated real-time access to data.	*'Through the university we can access to a lot of information that would normally be subscription.'* (GP5)
		+	Data associated with COVID-19 tended to be made more rapidly available.	*'Take COVID as an example…protocols and guidelines are changing week to week.'* (GP1)
	Social and professional roles and identity	~	Different professions had different access to science and informatics.	*'I have access to things on the ETG and AMA…those are things that are just available for registrars.'* (GP8)
	Knowledge	–	Low knowledge about how to access information prevented engagement with LHS.	*'I have no idea what [software] is, which sounds absolutely dreadful. No one's brought it up with me.'* (ADMIN13)
	Memory attention and decision processes	–	Use of science and informatics was based around their perceived difficulty or cognitive load.	*'[I'm] probably not sure [about using the software]…if it is quite too difficult or if it's, something that would actually take up time.'* (ADMIN2)
	Beliefs about consequences	+	Belief that technology would decrease workload and have positive impacts on care facilitated its use.	*'I mean [the technology] is definitely useful and how that could be used, would be to recall patients who are falling through the cracks.'* (GP2)
Patient–clinician partnerships	Reinforcement	+	Engagement in partnerships was driven by the rewards they provided.	*'The way I know whether we're providing a good service is, patient feedback.'* (ADMIN1)
	Environmental context and resources	+	Strong and clear leadership and management promoted partnerships.	*'[At MQGP]…the doctor, patient, the nurse, and admin staff work as a team, and that is again to empower the patient.'* (ADMIN12)
		+	Unique structure of a general practice meant that partnerships were also driven by non-clinical staff.	*'Normally—if the patient's happy to—they just tell [the reception staff] then they will communicate it straight to [senior administration staff].'* (ADMIN10)
	Professional role and identity	~	Professional role mediated nature of partnerships, for example, administrative staff were unable to give medical advice.	*'And that's hard because you, you want to try and help ease their anxiety. But at the same time, you can't give clinical advice.'* (ADMIN6)
	Beliefs about consequences	+	Positive beliefs about consequences of partnerships facilitated engagement in them.	*'There are some patients who…I believe they could definitely add some beautiful insight. Then there are some which would create more chaos.'* (ADMIN7)
Incentives	Reinforcement	+	Engagement with incentives was driven by the rewards they provided, both intrinsic and extrinsic.	*'The incentive personally is always to be better so that you can be better for your patients.'* (GP2)
	Professional role and identity	~	Professional role mediated access to incentives.	*'Within our clinic [for] GPs there are some… incentives in terms of particular indicators based on various things they clinically work on.'* (GP3)
	Emotion	–	Perceived inequity in incentives generated negative emotions and was a barrier to engagement.	*'Part of the issue is getting the philosophy of what's a proper incentive system…You don't want it to be competing with your colleagues, you actually want it to be collaborative and to be fair.'* (GP10)
Continuous learning culture	Environmental context and resources	+	Affiliation with both Macquarie University and the PHN facilitated learning opportunities.	*'We enlisted the PHN to run some improvement workshops, and the idea was for it to be team building.'* (ADMIN1)
		+	Leadership and management team perpetuated a strong culture of learning.	*'We are continuously, like, encouraged to learn information that is relevant to what we do every day.'* (ADMIN6)
		+	Distribution of a weekly newsletter facilitated learning.	*'We get a newsletter every week [that] updates protocols on a weekly basis. So, we're aware of those changes.'* (ADMIN12)
	Professional role and identity	~	Professional role influenced access to, and engagement with, learning opportunities.	*'[I have] a lot of different [learning activities] because GPs are always learning.'* (GP9)
	Social influences	+	Learning arose from social interactions with colleagues.	*'A lot of that comes from a peer saying 'I've discovered'.'* (GP10)
	Memory attention and decision processes	+	Clinical staff paid more attention to learning about conditions with which patients presented.	*'When you see the patient and you don't know something, then that raises a flag that…I need to read a bit more.'* (GP2)
	Beliefs about capabilities	+	Belief of inability to keep up with the pace of information prevented learning.	*'There's so much information in there…So, try and find the time to actually update myself is a bit strenuous.'* (ADMIN3)

Continued

Table 2 Continued

	Key TDF domains		Belief statements	Exemplar quote
Structure and governance	Environmental context and resources	–	Complexities and poor communication from Medicare prevented engagement with LHS.	*'Trying to find out what those changes were…there was literally nothing until the first of July, when MBS published the fact sheet.'* (ADMIN1)
		–	General practice is a low-risk environment, which was a barrier to knowledge of incident reporting processes.	*'I think there is an [incident] form…But obviously that's never happened. I don't know where that form is.'* (ADMIN10)
	Memory, attention and decision processes	+	Clear practice policies facilitated decision-making.	*'Having that sort of delegation of roles makes it easier for us.'* (GP5)
	Social influences	+	Colocation facilitated social relationships with other MQ Health clinicians, which mediated organisational engagement.	*'With all collaborations or referrals, it's good to know the person, [to whom] you are referring.'* (GP2)

~, Mediator. +, Facilitator. –, Barrier.
ADMIN, administrative staff; GP, general practitioner; LHS, learning health system; MQGP, MQ Health General Practice; PHN, Primary Health Network; TDF, theoretical domains framework.

(n=3), administrative staff (n=13) and a psychologist (n=1). Three clinicians were unable to attend their scheduled interview, and as data saturation was reached, these interviews were not rescheduled. Interviews lasted between 17 and 50 min (mean 35.5). Participating staff had been working at MQGP for between 3 weeks and 15 years.

Barriers and facilitators to an LHS

The *environmental context and resources available to participants* and their *social and professional role and identity* were key determinants in engagement with most domains of the LHS. *Reinforcement* was particularly important for the development of patient–clinician partnerships and engagement with incentives in the LHS, while a several domains of the TDF had a reciprocal relationship with the practice structure and governance; for example, clear policy facilitated the development of a strong professional identity, which then in turn facilitated access to and understanding of policy. Key barriers and facilitators that are relevant to each domain of the TDF are reported in table 2, according to each of the five LHS components. Figure 1 provides a visual summary of the framework analysis and the relative proportions of each TDF domain described in each LHS component.

Data

The codebook and exemplar quotes on which these results were based are available as online supplemental material associated with this manuscript. The full study dataset is available from the authors on reasonable request, subject to ethical approval.

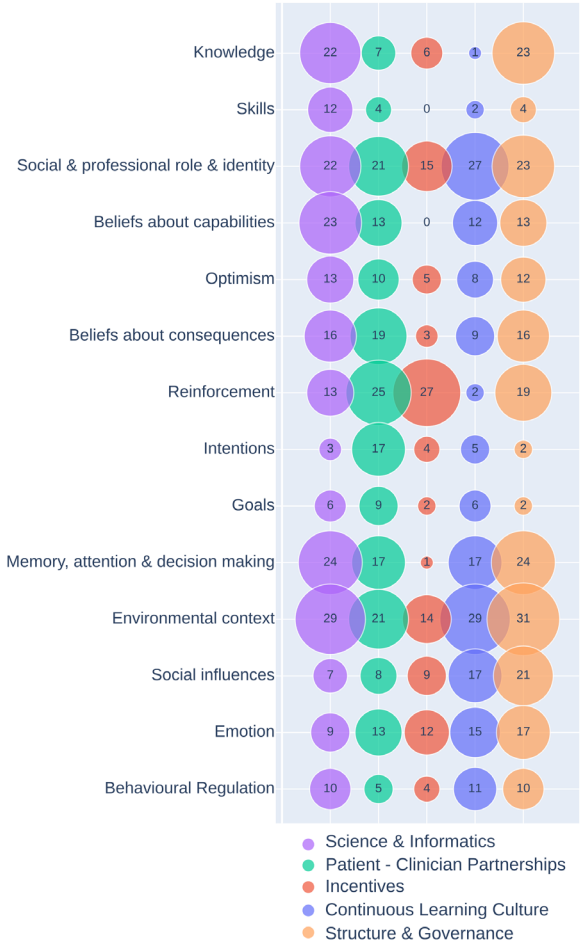

Figure 1 Results from the framework analysis using the learning health system (LHS) and theoretical domains framework (TDF) domains. Each domain of the LHS is represented by a different colour. Each coloured circle represents a TDF determinant. The sizes of the circles represent the number of participants who reported a quote in the respective TDF determinant, which are also written on each circle.

DISCUSSION

In our original study, we presented a case study of an LHS within an Australian primary care setting and showed that it was operating within several dimensions of the LHS framework, and that its staff were willing to embrace additional elements of the LHS.[24] In this secondary analysis, we used the TDF to describe barriers and facilitators to

converting this willingness into reality. In all LHS domains there was a consistently reported influence of *environmental context and resources*; for example, the MQGP affiliation with a university was described as a strong facilitator of learning, and the unique general practice environment was reported to shape patient–clinician partnerships. The *professional role* of participants was a second consistently reported determinant, influencing access and attitudes to learning and incentives. The reported impact of other determinants varied across LHS dimensions; for example, continuous learning culture was mediated by *social influences*, where strong social relationships were reported to facilitate informal learning, while a lack of *knowledge* of clinic structure and governance was described as a barrier to its effectiveness. Overall, our results show that implementing the principles of an LHS in this primary care setting was influenced by many behavioural determinants, some applicable to healthcare in general, but most specific to the general practice structure and environment.

A key strength of the study was its codesign, which allowed it to reflect the goals of both the research team and the staff of MQGP. Further strengths included the high participation rate and broad recruitment strategy, which enabled a comprehensive description of behavioural determinants from the perspective of clinical and non-clinical staff. Additionally, this secondary analysis was conducted by an IAT that did not participate in the original study and were thus less subject to biases from their relationships with practice staff or from the original interviews. The primary limitation of this study was the inclusion of only one organisation, limiting the generalisability of our results to other primary care settings. Generalisability is also limited by the affiliation of the practice with a university which, while a facilitator of the uptake of LHS principles, is relatively uncommon in the Australian context. A final limitation was the timing of the present study, which was conducted during and after significant public health restrictions associated with the COVID-19 pandemic. These restrictions, and their removal, would likely have influenced the responses of participants.

Despite these limitations, our results have encouraging similarities with the few other empirical investigations of primary care LHS that are grounded in implementation science.[19] Pestka and colleagues qualitatively evaluated the lessons learnt from their implementation of a primary care LHS in the USA.[27] They, too, reported that clearly defined roles and the incentivisation of value-based care were facilitators to the development of an LHS, as was the use of a weekly newsletter to communicate essential information. However, their investigation took place in a system of 40 primary care practices, much larger than the two practices described in the present study. The facilitatory effect of a weekly newsletter was diluted by a larger LHS size, where at times people had '*no idea what was going on at other stations*',[27] a finding that was echoed by another investigation of a province-wide primary care LHS in Canada.[28] The same study also reported that the perceived difficulty or cognitive load of a technology was a primary barrier to its use, and that a perceived increase in the quality and efficiency of patient care was a motivation for participants to engage in the LHS,[28] findings similar to our results. However, a key difference between their investigation and our own was the *type* of incentives that motivated participants; in the Canadian province-wide primary care LHS competition or peer pressure were motivators for engagement,[28] while our participants reflected that they were primarily motivated by the rewards of providing better patient care and developing a sense of comradery with their colleagues. These differences may reflect the different social contexts in which the studies were conducted, particularly the influence of the COVID-19 pandemic, in which healthcare workers likely banded together to deal with high levels of uncertainty and stress.

The results of our own and other empirical investigations suggest that while some barriers and facilitators are unique to certain contexts, others are common to many journeys towards a primary care LHS. These are summarised in box 1, which also describes possible strategies for primary care practices to facilitate their journey towards an LHS. A notable facilitator that likely applies to all contexts is *external support*, as many primary care providers work in small independent community practices which limits their access to resources.[29] Affiliations with academic and professional institutions, including the use of codesign and embedded researchers, or collaborations of multiple primary care practices are viable

Box 1 Summary of five key barriers and facilitators to a learning health system (LHS) in primary care and five proposed solutions.

Key barriers
- ⇒ Unclear policy and roles.
- ⇒ Poor data quality.
- ⇒ Complex learning requirements.
- ⇒ Physical distance between teams.
- ⇒ Poor communication with patients.

Key facilitators
- ⇒ Strong leadership.
- ⇒ Desire to help patients.
- ⇒ Shared organisational goals.
- ⇒ Culture of patient-centred care.
- ⇒ Communication of progress and goals.

Key solutions
- ⇒ Formal lines of patient communication and feedback (eg, online reviews).
- ⇒ Diverse modes of care and communication (eg, telehealth).
- ⇒ Weekly practice newsletter to share updates and progress.
- ⇒ Multidisciplinary leadership teams that model a learning culture.
- ⇒ Mentorship and 'buddy systems' between senior and junior staff.

Each point describes a barrier, facilitator or solution described in at least two of the three following papers: (a) Nash *et al*,[28] (b) Pestka *et al*[27] or (c) current study.

strategies that cultivate a primary care LHS. Additionally, our results suggest that it is not only patient–*clinician* partnerships that are important in the primary care LHS, but rather that administrative staff also play an important role in the patient experience. As such, primary care practices that aim to become LHS should invest in training, involvement and retention of all staff, not just those in clinical roles.

CONCLUSION

There are numerous benefits, success factors and barriers in primary care settings making the transition to LHS. These factors should feed into a roadmap to assist primary care settings that are at different stages of the journey towards an LHS.

Acknowledgements The authors gratefully acknowledge all interview participants who kindly gave their time to partake in this research.

Contributors JB, SW, JM, GD, KC and LE conceptualised the study. GD and LE collected the data. GF and MS analysed the data with input from LE and GD. GF created the study figures with input from MS, LE and JB. GF drafted the manuscript. JB is guarantor.

Funding JB is funded by multiple grants including the National Health and Medical Research Council (NHMRC) Partnership Grant for Health Systems Sustainability (ID: 9100002) and NHMRC Investigator Grant (ID: 1176620) on the learning health system and its applications.

Disclaimer The funders had no role in the design, analysis, and interpretation of the research, or drafting of the manuscript.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by Macquarie University Human Research Ethics Committee (reference number: 52021905624229). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

ORCID iD

Georgia Fisher http://orcid.org/0000-0002-7252-7800

REFERENCES

1 Australian Institute of Health and Welfare. Primary health care in Australia. Australian Institute of Health and Welfare; Canberra (AU), 2016. Available: https://www.aihw.gov.au/reports/primary-health-care/primary-health-care-in-australia/contents/about-primary-health-care [Accessed 4 Aug 2023].
2 NHS england. London (ENG): NHS England. Primary care services, Available: https://www.england.nhs.uk/get-involved/get-involved/how/primarycare/ [Accessed 4 Aug 2023].
3 Dolton P, Pathania V. Can increased primary care access reduce demand for emergency care? evidence from England's 7-day GP opening. *J Health Econ* 2016;49:193–208.
4 Australian Institute of Health and Welfare. General practice, Allied health and other primary care services. Australian Institute of Health and Welfare; Canberra (AU), 2023. Available: https://www.aihw.gov.au/reports/primary-health-care/general-practice-allied-health-and-other-primary-c [Accessed 4 Aug 2023].
5 NHS England. NHS England; London (ENG), 4 August 2023. Available: https://www.england.nhs.uk/five-year-forward-view/next-steps-on-the-nhs-five-year-forward-view/primary-care/
6 Australian Institute of Health and Welfare. Ageing and the health system: challenges, opportunities and adaptations. Australian Institute of Health and Welfare; Canberra (AU), 2014. Available: https://www.aihw.gov.au/getmedia/19dbc591-b1ef-4485-80ce-029ff66d6930/6_9-health-ageing.pdf.aspx [Accessed 4 Aug 2023].
7 Raymond A, Bazeer N, Barclay C, *et al*. Our Ageing Population: How Ageing Affects Health and Care Need in England. London: The Health Foundation, 2021.
8 Royal Australian College of General Practitioners. Australia; General Practice: Health of the Nation 2021, 2021. Available: https://www.racgp.org.au/health-of-the-nation/chapter-2-general-practice-access/2-2-gp-workforce
9 Osborne S. One in seven GP posts empty as vacancy rates remain high: pulse. 2021. Available: https://www.pulsetoday.co.uk/news/workforce/one-in-seven-gp-posts-empty-as-vacancy-rates-remain-high
10 Pressures in general practice data analysis: British Medical Association, 2023. Available: https://www.bma.org.uk/advice-and-support/nhs-delivery-and-workforce/pressures/pressures-in-general-practice-data-analysis
11 Royal Australian College of General Practitioners. Australia; General Practice: Health of the Nation 2022, 2022. Available: https://www.racgp.org.au/general-practice-health-of-the-nation-2022/
12 Joy M, McGagh D, Jones N, *et al*. Reorganisation of primary care for older adults during COVID-19: a cross-sectional database study in the UK. *Br J Gen Pract* 2020;70:e540–7.
13 Jefferson L, Golder S, Heathcote C, *et al*. GP wellbeing during the COVID-19 pandemic: a systematic review. *Br J Gen Pract* 2022;72:e325–33.
14 Foo D, Mahadeva J, Lopez F, *et al*. High-performing primary care: Reinvigorating general practice as a learning health system. *Br J Gen Pract* 2023;73:8–9.
15 Smith M, Halvorson G, Kaplan G. What's needed is a health care system that learns: recommendations from an IOM report. *J Am Med Assoc* 2012;308:1637–8.
16 Enticott J, Johnson A, Teede H. Learning health systems using data to drive Healthcare improvement and impact: a systematic review. *BMC Health Serv Res* 2021;21:200.
17 Olsen L, Aisner D, McGinnis JM, eds. Institute of medicine (US) Roundtable on evidence-based medicine. In: *The Learning Healthcare System: Workshop Summary*. Washington (DC): National Academies Press (US), 2007.
18 Zurynski Y, Smith CL, Vedovi A, *et al*. Mapping the learning health system: a Scoping review of current evidence. Sydney: Australian Institute of Health Innovation and the NHMRC Partnership Centre for Health System Sustainability, 2020.
19 Nash DM, Bhimani Z, Rayner J, *et al*. Learning health systems in primary care: a systematic Scoping review. *BMC Fam Pract* 2021;22:126.
20 Bauer MS, Kirchner J. Implementation science: what is it and why should I care. *Psychiatry Res* 2020;283:112376.
21 De Bruin J, Bos C, Struijs JN, *et al*. Conceptualizing learning health systems: A mapping review. *Learn Health Syst* 2023;7:e10311.
22 Braithwaite J. Changing how we think about Healthcare improvement. *BMJ* 2018;k2014.
23 Atkins L, Francis J, Islam R, *et al*. A guide to using the theoretical domains framework of behaviour change to investigate implementation problems. *Implement Sci* 2017;12:77.
24 Dammery G, Ellis LA, Churruca K, *et al*. The journey to a learning health system in primary care: a qualitative case study Utilising an embedded research approach. *BMC Prim Care* 2023;24:22.
25 Churruca K, Ludlow K, Taylor N, *et al*. The time has come: embedded implementation research for health care improvement. *J Eval Clin Pract* 2019;25:373–80.
26 Gale NK, Heath G, Cameron E, *et al*. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Med Res Methodol* 2013;13:117:1–8:.

27 Pestka DL, White KM, DeRoche KK, *et al*. Trying to fly the plane while we were building it'. applying a learning health systems approach to evaluate early-stage barriers and Facilitators to implementing primary care transformation: a qualitative study. *BMJ Open* 2022;12:e053209.
28 Nash DM, Brown JB, Thorpe C, *et al*. The Alliance for healthier communities as a learning health system for primary care: A qualitative analysis in Ontario, Canada. *J Eval Clin Pract* 2022;28:1106–12.
29 Ministry of health and long-term care and eHealth Ontario. electronic health records. In: *Implementation Status Ontario (CA*). Ontario (CA): Ministry of Health and Long-Term Care and eHealth Ontario, 2016. Available: https://www.auditor.on.ca/en/content/annualreports/arreports/en16/v1_303en16.pdf [accessed 4 Aug 2023].

BMJ Health & Care Informatics

Promising algorithms to perilous applications: a systematic review of risk stratification tools for predicting healthcare utilisation

Christopher Oddy ,[1] Joe Zhang,[2,3] Jessica Morley,[4] Hutan Ashrafian[2]

To cite: Oddy C, Zhang J, Morley J, *et al*. Promising algorithms to perilous applications: a systematic review of risk stratification tools for predicting healthcare utilisation. *BMJ Health Care Inform* 2024;**31**:e101065. doi:10.1136/bmjhci-2024-101065

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjhci-2024-101065).

Received 23 February 2024
Accepted 14 May 2024

[1]Department of Anaesthesia, Critical Care and Pain, Kingston Hospital NHS Foundation Trust, London, UK
[2]Imperial College London Institute of Global Health Innovation, London, UK
[3]London AI Centre, Guy's and St. Thomas' Hospital, London, UK
[4]Digital Ethics Center, Yale University, New Haven, Connecticut, USA

Correspondence to
Dr Christopher Oddy;
christopher.oddy1@nhs.net

ABSTRACT

Objectives Risk stratification tools that predict healthcare utilisation are extensively integrated into primary care systems worldwide, forming a key component of anticipatory care pathways, where high-risk individuals are targeted by preventative interventions. Existing work broadly focuses on comparing model performance in retrospective cohorts with little attention paid to efficacy in reducing morbidity when deployed in different global contexts. We review the evidence supporting the use of such tools in real-world settings, from retrospective dataset performance to pathway evaluation.

Methods A systematic search was undertaken to identify studies reporting the development, validation and deployment of models that predict healthcare utilisation in unselected primary care cohorts, comparable to their current real-world application.

Results Among 3897 articles screened, 51 studies were identified evaluating 28 risk prediction models. Half underwent external validation yet only two were validated internationally. No association between validation context and model discrimination was observed. The majority of real-world evaluation studies reported no change, or indeed significant increases, in healthcare utilisation within targeted groups, with only one-third of reports demonstrating some benefit.

Discussion While model discrimination appears satisfactorily robust to application context there is little evidence to suggest that accurate identification of high-risk individuals can be reliably translated to improvements in service delivery or morbidity.

Conclusions The evidence does not support further integration of care pathways with costly population-level interventions based on risk prediction in unselected primary care cohorts. There is an urgent need to independently appraise the safety, efficacy and cost-effectiveness of risk prediction systems that are already widely deployed within primary care.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Risk prediction models that stratify primary care populations according to their likelihood of accessing healthcare resources are generally considered to perform well within similar contexts to those in which they were derived. It is unclear how they perform when deployed in wider global contexts and indeed if their application can be harnessed to reduce resource demands.

WHAT THIS STUDY ADDS

⇒ We find that most models have not been studied in a sufficient diversity of contexts to appraise the robustness of prediction, however, those that have appear to retain their discriminatory ability. The real-world application of these models to reduce healthcare resource use in unselected cohorts has produced disappointing results, with an equal weight of evidence suggesting a harmful effect as a beneficial one in this context.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Our results call into question the common, and costly, practice of commissioning population health management strategies based on risk stratification of whole primary care populations without a concrete understanding of the associated risks.

INTRODUCTION

Risk stratification tools that predict healthcare resource use are widely used in primary care settings.[1–6] These tools are integral to population health management (PHM) strategies around the world, enabled by the availability of routinely collected data from sources such as electronic health records.[7] Risk stratification tools typically use predictive models that are developed through statistical or machine learning (ML) techniques, to generate an individual risk score for some measure of resource use. These scores form a key component of anticipatory care pathways, where those at the highest risk may be targeted for specific interventions aimed at reducing future morbidity.[8–11] The process by which these tools are ideally developed and deployed within healthcare systems is summarised in figure 1.

A growing body of literature describes the development and validation of risk

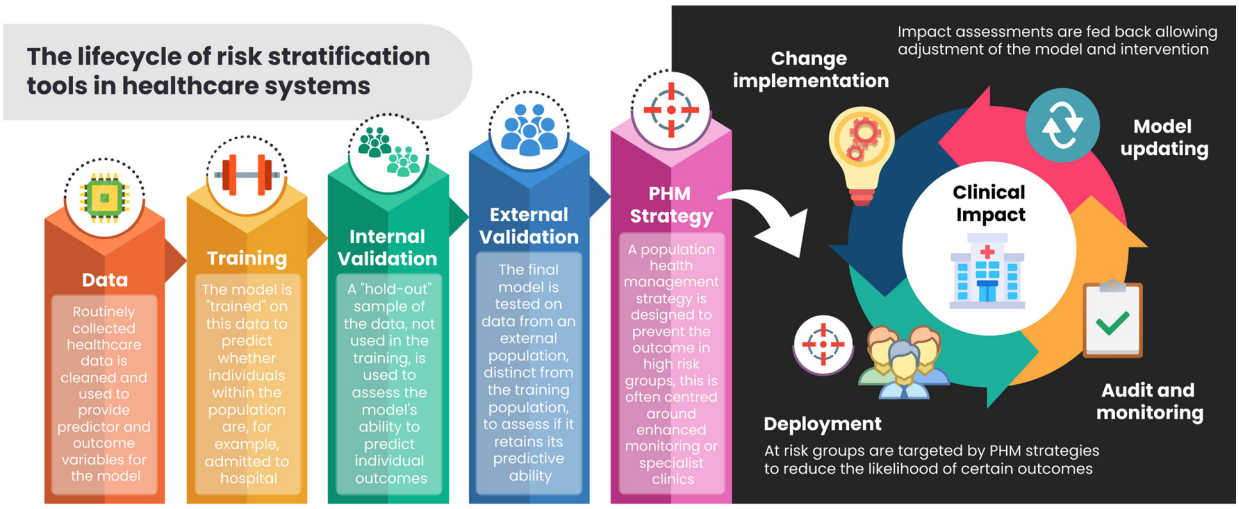

Figure 1 An infographic describing an idealised process for developing and deploying a risk prediction tool within a healthcare system. In black is the deployment cycle, linking risk prediction tools and their associated population health management measures to a lifecycle of evidence generation, impact evaluation and monitoring for negative consequences that are fed back into the model and intervention.

stratification tools in the primary care setting reporting an acceptable discriminatory power for the majority of models.[1 2 12 13] However, existing work broadly focuses on the assessment of model performance within retrospective datasets, with little attention paid to their efficacy in real-world settings, where the clinical impacts of deploying these algorithms within a population are assessed. Commercial literature asserts the efficacy of interventions based on algorithmic case selection in improving key outcomes, such as hospital admission rates, but suffers from a lack of transparency in data and methodology.[14 15]

Predictive models that appear accurate in development are increasingly found to be ineffective or unsafe when deployed in clinical pathways. Predictive performance may be diminished when translated to demographically and culturally distinct populations, or when deployed using electronic health data with differing characteristics. Differences in how healthcare resources are used in local settings, alongside inherent biases inlaid within such technologies, may result in varying clinical effectiveness from inconsistent intervention thresholds, variation in the physical clinical interventions that are deployed, to sociotechnical variation across end-users and processes.[16–20] Resultantly, where an algorithm is deployed into an untested context without real-world evidence for a comparable integrated pathway, there are risks to both patient safety and exacerbation of healthcare inequalities through a lack of fairness in prediction or intervention allocation.

With extensive integration of risk stratification into pathways within primary care systems worldwide it is of paramount importance to establish the current evidence base on which these care-defining interventions can be appraised. We therefore systematically review the available literature concerning risk stratification tools for predicting future healthcare utilisation in primary care populations. We present three aims: (1) to update existing evidence for algorithmic solutions with attention paid to predictive performance and risk of bias in dataset evaluation, as well as real-world clinical outcomes; (2) to describe the transfer of algorithms from initial development to testing and deployment in different global contexts and (3) to evaluate risks in cross-context transfer and application. Based on our findings, we provide recommendations for the responsible evaluation and deployment of predictive risk stratification tools.

METHODS

Search strategy

A systematic search of the MEDLINE, Embase and Global Health databases was carried out on 18 July 2023 via the Ovid platform. PRISMA guidelines were followed throughout the conduct and reporting of this review.[21] A combination of keywords and MeSH terms was used to curate relevant literature, details of which are available in online supplemental material.

Inclusion and exclusion criteria

We defined our inclusion criteria using the Population, Intervention, Control and Outcome method. The population of our analysis was selected to be comparable to the populations in which these models are currently in use. We therefore included only papers that applied algorithms to unselected primary care populations, where deployment was to the entire patient population for a given organisation without selection of particular groups. Prestratified populations, such as specific disease groups, or groups previously identified as high risk for healthcare utilisation, were excluded. Age-stratified populations were permitted as this is a pragmatic selection criterion adopted by the majority of predictive modelling work. Publications applying algorithms to historic research

study datasets or specifically designed questionnaires (ie, not routinely collected or 'real-world data'[22]) were also excluded.

Our intervention was defined as the application of a risk stratification model to an appropriate population in the process of derivation or validation, or to perform case selection as part of a PHM strategy. Models reliant on non-routinely collected data, such as questionnaire results, were excluded.

Outcomes included measures of predictive performance across five main categories: access to primary care services; emergency department attendance; healthcare costs; hospital admissions and readmission. Studies examining risk of readmission were included provided that the study population was not limited to patients with a recent admission. A group formed of those who had recently been admitted would, by definition, no longer be considered unselected and would thus violate our population criteria. Composite (eg, admissions and mortality as a single endpoint) and component (eg, respiratory admissions instead of total admissions) outcomes were excluded. We also considered clinical impact assessments related to a real-world evaluation.

Study selection and quality appraisal

Titles and abstracts were screened by two reviewers (CO/JZ) according to the criteria set out above, with all conflicts decided by a third (JM). Eligible publications were read in full and assessed for exclusions not apparent in the title or abstract, and for methodological quality.[23] Risk of bias was assessed using the Prediction model Risk Of Bias ASsessment Tool.[24]

Data extraction

We extracted information regarding model characteristics, study design and context, predictive performance, and measures of clinical impact from any associated intervention where evaluation took place in a real-world setting. Due to significant heterogeneity in study design and reporting a meta-analysis was not conducted. C-statistics were used as the primary outcome for model performance. A subset of papers did not report discrimination, but instead reported goodness of fit using coefficient of determination (R^2) which were extracted where available. Impact evaluations were described using the terminology and significance testing employed in the original paper, commonly expressed as the absolute difference (AD) between groups or odds ratios (OR).

Model appraisal

Models that appeared in multiple studies were qualitatively appraised by comparing their derivation methodology to subsequent external validation or clinical evaluation studies. For each model we report: the context of its original development; contexts in which the model's predictive performance has been tested; and contexts in which the model's real-world impacts have been assessed. Results were synthesised separately as the outcome of either internal or external validation. Internal validation was defined as any measure of predictive performance within the same population in which the model was derived, and external as any validation using data from a separate population.

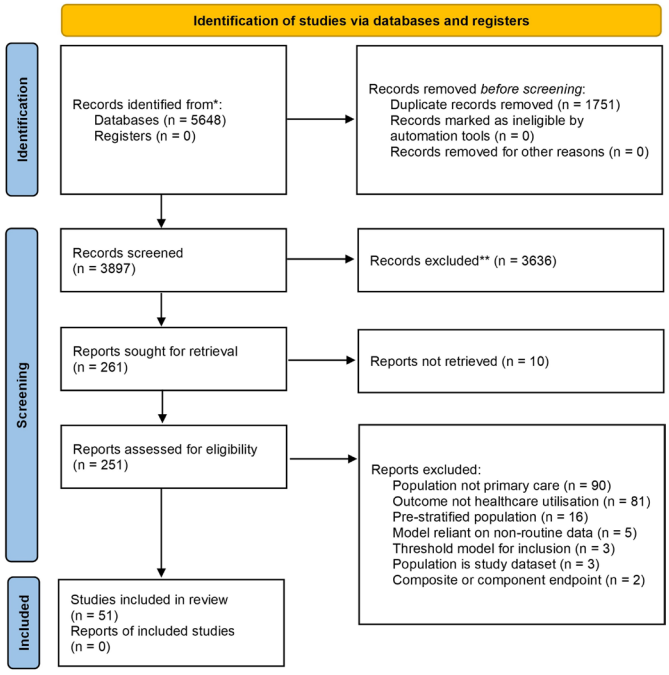

Figure 2 A PRISMA flow diagram showing the process of study selection for our analysis. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

RESULTS

Systematic review

Our review identified 3897 publications eligible for screening after duplicates were removed (figure 2). Of these, 3636 were excluded on the basis of their title or abstract alone leaving 261 that were sought for retrieval. Full texts could not be retrieved for 10 publications, thus 251 were reviewed in full. A total of 51 publications met our criteria and were included in our final analysis (online supplemental table 1).[25–75] Further detail about the identified models, along with our risk of bias analysis, can be found in online supplemental materials.

The majority of studies were based in the USA (23), with the remainder set in the UK (10), Spain (9), Canada (2), Italy (2), New Zealand (2), Australia (1), Ireland (1) and Israel (1). Population sizes ranged from 96 to 5.4 million with a median value of 94 264 (IQR 12 800–434 027). Hospital admission was the most commonly predicted outcome (34), followed by healthcare costs (14), emergency department attendance (9), access to primary care services (8), mortality (5) and readmission (2).

19 studies reported the derivation and internal validation of a risk stratification model with 32 describing validation of a model in a separate population dataset. 10 studies reported the results of implementing PHM measures based on case selection by a risk stratification

model in a real-world clinical pathway. These included five randomised control trials (RCTs), three prospective cohort studies and two retrospective cohort studies. PHM strategies used were case management (8), telemonitoring (4) and care coordination (3).

We identified 28 risk stratification tools across all studies. 42 studies examined a single model, whereas 9 studied the comparative efficacy of several models. Johns-Hopkins ACG was the most studied algorithm (20), followed by the Charlson Comorbidity Index (10), Hierarchical Condition Categories (8), the Chronic Illness and Disability Payment System (3), RxRisk (3), the Elder Risk Assessment Index (2), the Patients At Risk of Rehospitalisation algorithm (2) and QAdmissions (2). Of the remainder, four were proprietary ML algorithms.

Results of internal and external validation studies

A summary of the derivation characteristics of each of the 28 discovered models is compared with the results of subsequent validation studies in online supplemental table 2.[25–84] The results of internal validation studies echoed previous reviews with C-statistics for various outcomes ranging from 0.67 to 0.90. Notably, three of the highest C-statistics within internal validation samples were displayed by models derived using ML techniques—0.84,[67] 0.85[42] and 0.90.[55]

Half (14) of the discovered models underwent external validation. Of these, only the Charlson Comorbidity Index and the Johns Hopkins ACG System were validated internationally. Model performance in external validation studies generally resembled internal validation performance for each model, with C-statistics ranging from 0.53 to 0.88. Accounting for heterogeneity in study design and reporting, there was no evident association between validation context and model discrimination, with models broadly displaying consistent predictive performance when transported to external datasets.

Results of real-world evaluation studies

Two studies reported the implementation of risk stratification tools into care pathways within the same population used for development. The Nairn Case Finder[73] and the Predictive RIsk Stratification Model (PRISM)[25] algorithms were used to identify those that might benefit from case management, both in the hope of reducing hospital admissions. In a prospective stepped-wedge clinical trial conducted across more than 230 000 patients in 32 primary care practices, the practice resource allocation intervention linked to PRISM resulted in significantly increased hospital admissions (OR 1.44 (95% CI 1.39 to 1.50), p<0.001), as well as increased emergency presentations, time in hospital, and primary care workload. The intervention guided by the Nairn Case Finder significantly reduced hospital admissions (AD=42.5%, p=0.002) in a population of 96 high-risk patients from a single locality, when matched 1:1 on risk score to patients in a separate control population.

Eight of the discovered models were deployed as tools for case selection as part of a PHM strategy in a separate context from development. The Johns Hopkins ACG System was deployed in two separate studies, whereas each of the other models was deployed only once. Healthcare utilisation measures were not significantly influenced by interventions guided by the Hierarchical Condition Category[71] and PacifiCare's Medicare Risk Programme[37] models. Similarly equivocal evidence for the efficacy of interventions linked with the Johns Hopkins ACG System was observed, with one study showing no benefit[31] and the other demonstrating benefit in groups selected by the model (OR 0.91 (95% CI 0.86 to 0.96)) but reciprocal harm in non-prioritised groups (OR 1.19 (95% CI 1.09 to 1.30)).[32] Interventions linked with the Elder Risk Assessment Index[30] and QAdmissions[48] algorithms led to significant increases in mortality (AD 10.8%, p=0.008) and hospital admissions (difference in difference 79.8 (95% CI 21.2 to 138.4), p=0.01), respectively.

Significant reductions in hospital admissions were achieved through interventions guided by the combined predictive model (AD=−0.9, p<0.001),[39] Patients At Risk for Rehospitalisation algorithm (AD=−0.3, p<0.001)[39] and SCAN Health Plan Model (AD=11.5%, p=0.02).[51] Figure 3 summarises the main findings of this review, describing only the models that underwent external validation or real-world evaluation.

DISCUSSION

Main results

Our review identifies 28 risk stratification tools designed to predict healthcare utilisation in an unselected primary care population. The discriminatory ability of half of the discovered models was validated in an external cohort. However, only two, the Charlson Comorbidity Index and Johns Hopkins ACG System, were validated in a different country from their derivation dataset. No evident association between validation context and model discrimination was observed. Models derived using ML techniques displayed the best predictive performance, however, none of these models underwent external validation.

The results of real-world evaluation studies present equivocal evidence for the efficacy of these population-level interventions. The majority of publications reported no change, or indeed significant increases, in healthcare utilisation within groups targeted by the intervention, with only one-third of reports demonstrating some benefit.

Comparison with the literature

We corroborate the results of previous reviews by observing that the discriminatory power of a variety of risk stratification tools is robust to external validation.[1 2 12 13] We add that the context of model validation appears to have minimal impact on predictive performance and highlight a scarcity of literature appraising the impact of deploying these models to guide PHM strategies despite extensive

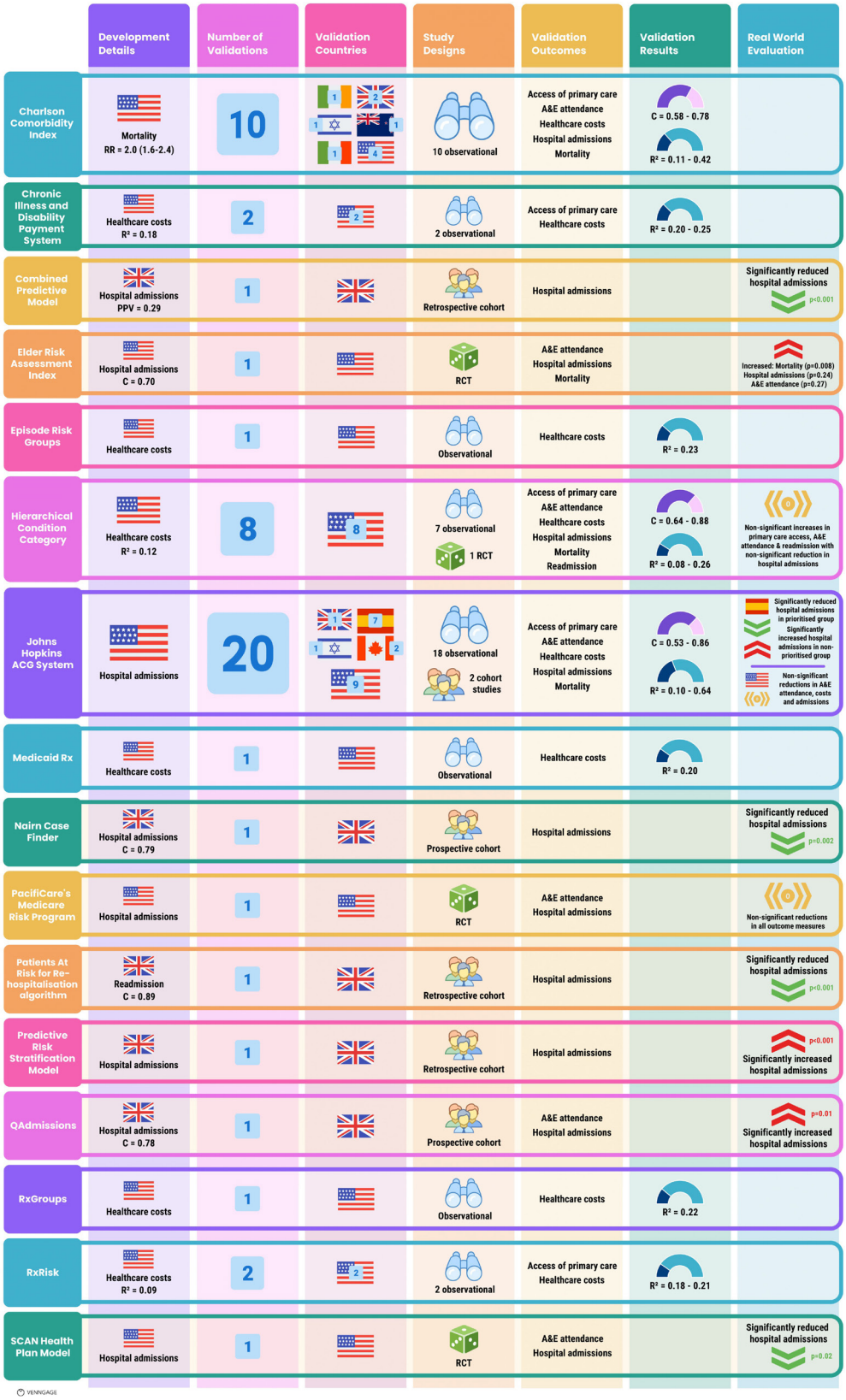

Figure 3 An infographic summarising the validation characteristics of the identified models that underwent external validation or real-world testing. Models that underwent more extensive validation processes are represented by larger boxes. Each box contains aggregated data for all of the external validation and real-world evaluation studies for each model. Validation countries are represented by flags with the number of studies based in each country overlying. R^2 and C-statistics are displayed as ranges for all of the outcome measures tested for each model for illustrative purposes only. A&E, accident and emergency department; PPV, positive predictive value; RCT, randomised controlled trial; RR, risk ratio.

integration of risk stratification into pathways within primary care systems worldwide.[3–6]

Our finding that deployment of these models is not consistently associated with reductions in healthcare utilisation is perhaps unsurprising. PHM strategies applied to unselected primary care cohorts, with case selection achieved through a variety of different means, have frequently been shown to increase costs without an associated reduction in morbidity.[9 85–87] A single 2014 meta-analysis, aggregating a heterogeneous group of strategies as a single intervention, demonstrated marginal reductions in resource use within a relevant cohort.[88] However, these findings were subject to substantial heterogeneity (I^2=58%–85%) and, while ostensibly the target population of this analysis was patients generally at high risk of healthcare resource use, the majority of included studies reported interventions targeted at specific disease cohorts. There is broad consensus that PHM strategies designed specifically for those with certain chronic conditions significantly reduce morbidity.[89–94] Taken with our findings, the available evidence indicates that the success of PHM strategies in specific disease groups may not be generalisable to unselected cohorts, and this remains the case when predictive modelling is employed to augment case selection.

The findings of our analysis of peer-reviewed literature stand in stark contrast to the impact statements of commercial suppliers of care systems that employ risk stratification. One such statement compared resource use statistics of product users to standardised national trends in an unadjusted analysis finding significant reductions in every parameter.[15] However, as is expressly the case for statements within product literature, a lack of transparency relating to the methods of data collection and analysis makes verifying these claims impossible.

Interpretation

We propose that the discouraging results of studies deploying risk stratification tools to guide PHM strategies primarily result from a mismatch between theoretical model development and complexities of real-world pathways. Risk stratifying patients by their likelihood of resource use alone almost invariably leads to the creation of a diverse intervention cohort, where individual clinical need is likely to be heterogeneous. This is likely the reason that population-level interventions have failed to replicate the results of successful programmes targeting specific chronic conditions. Presently, there is a paucity of evidence to guide best practice once high-risk users are identified, and no recommendations can be made about the efficacy of any single intervention over another. Results of real-world evaluation studies, therefore, present a cautionary tale of designing clinical pathways based on the principle of simply flagging high-risk patients without a concrete understanding of how this translates into practice.

We did not observe an effect of validation context on algorithmic performance. This is most likely due to the low number of comparable values obtained for each model, the heterogeneity of the study design, and a predictably small absolute effect size. Diminished performance when algorithms are deployed in new environments is a highly replicable finding, and our results should not be interpreted to contradict this established premise. However, this finding does imply that poor predictive performance is unlikely to be the primary reason for the failure of these algorithms to produce consistent results.

Limitations

It is important to put these findings within the context of our methodological constraints. Primarily, our analysis was limited by the heterogeneity of the included studies. Model performance was variably reported in terms of C-statistics and R^2 values which cannot be directly compared. Real-world evaluation studies suffered from a lack of uniformity of intervention as many reported the results of a bespoke system designed by the study authors. This prevented direct comparison of the efficacy of particular intervention categories within our study cohort as their results could not be appropriately aggregated. While our analysis identified several models with sufficient diversity of validation to demonstrate robust performance in a variety of contexts, this sample was small, and no strong conclusions can be drawn about the scale of algorithmic drift when such models are transported to new datasets. Finally, the majority of included publications were observational or cohort studies, with only a small number of RCTs identified.

Implications

The integration of risk stratification into pathways that define care decisions for millions of individuals around the world is already well established. Our findings suggest an absence of clinical impact, and indeed a signal of harm in a third of cases, raising several important considerations. First, this presents clear implications for patient safety, particularly in the absence of regular independent appraisal of the personal and system-wide effects. In addition to aggregate population health impacts, this includes the impact on individuals of incorrect stratification, and of negative biases through poorly calibrated algorithms. Second, the effects on provider workload of instituting and enacting these often time-consuming PHM interventions must be considered in the calculation of risk versus benefit. Finally, the absence of established benefits calls into question the cost-effectiveness of these programmes, particularly when used in healthcare systems where resources are constrained.

We therefore propose the following recommendations:

1. Deployment of individual-level risk prediction, with impact on clinical care pathways, must be subject to the same controls as other medical technologies. This would require matching their use to a responsible lifecycle of evidence generation, impact evaluation and monitoring for negative consequences. Such a lifecycle should include pre hoc evaluation, in the form of

local testing, and controlled trials for integrated pathways, as well as post hoc analyses of economic impact and healthcare outcomes in targeted and non-targeted groups. The first step in this process may be agreement on an auditable validation framework, such as BS 30440 developed by the British Standards Institution, to permit a more systematic approach to evaluation of such products.

2. National bodies involved in the procurement of commercial risk stratification services must review the cost-effectiveness and systemic implications of adjusting the likelihood of individuals within the population they serve accessing care based on personal predicted risk.
3. Regulatory bodies, including the Medicines and Healthcare products Regulatory Agency and the US Food and Drug Administration, must either confirm that risk stratification algorithms fall within their purview and are thus subject to the same regulation as other technologies defined as a 'Software as a Medical Device', or clarify why these algorithms do not fall into this category.

CONCLUSION

While model performance appears to generalise in most evaluations, there is little evidence to suggest that the identification of high-risk individuals can be translated to improvements in service delivery or morbidity. The available evidence does not support further integration of these types of risk prediction into population healthcare pathways. There is an urgent need to independently appraise the safety, efficacy and cost-effectiveness of risk prediction systems that are already widely deployed within primary care.

X Christopher Oddy @_chrisoddy_

Acknowledgements JZ acknowledges funding from the Wellcome Trust (203928/Z/16/Z) and support from the National Institute for Health Research Biomedical Research Centre based at Imperial College NHS Trust and Imperial College London. JM is a Wellcome Trust Doctoral Fellow.

Contributors CO and JZ were jointly responsible for study conception and design. CO and JZ performed the process of abstract screening and study selection, with all conflicts resolved by JM. CO was responsible for the curation and reporting of the data, the creation of figures and tables and primary manuscript drafting. All authors contributed to manuscript drafting and revision.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests HA is the chief scientific officer of Preemptive Health and Medicine and Flagship Pioneering. JM was paid directly for giving a lecture at Health Education England on the topic of artificial intelligence in the NHS. JM has been a member of the INSIGHT DataTAB for HDR UK. JZ is employed as a Senior Informatician by Arcturis Data, a Real-World Data company.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

ORCID iD

Christopher Oddy http://orcid.org/0000-0002-2311-9261

REFERENCES

1 Alonso-Morán E, Nuño-Solinis R, Onder G, *et al*. Multimorbidity in risk stratification tools to predict negative outcomes in adult population. *Eur J Intern Med* 2015;26:182–9.
2 Girwar S-AM, Jabroer R, Fiocco M, *et al*. A systematic review of risk stratification tools internationally used in primary care settings. *Health Sci Rep* 2021;4:e329.
3 Kingston M, Griffiths R, Hutchings H, *et al*. Emergency admission risk stratification tools in UK primary care: a cross-sectional survey of availability and use. *Br J Gen Pract* 2020;70:e740–8.
4 NHS England. Risk stratification. 2023. Available: https://www.england.nhs.uk/ig/risk-stratification/ [Accessed 28 Nov 2023].
5 Mora J, Iturralde MD, Prieto L, *et al*. Key aspects related to implementation of risk stratification in health care systems-the ASSEHS study. *BMC Health Serv Res* 2017;17:331.
6 Reddy A, Sessums L, Gupta R, *et al*. Risk stratification methods and provision of care management services in comprehensive primary care initiative practices. *Ann Fam Med* 2017;15:451–4.
7 Swarthout M, Bishop MA. Population health management: review of concepts and definitions. *Am J Health Syst Pharm* 2017;74:1405–11.
8 Hudon C, Chouinard M-C, Lambert M, *et al*. Effectiveness of case management interventions for frequent users of healthcare services: a scoping review. *BMJ Open* 2016;6:e012353.
9 Stokes J, Panagioti M, Alam R, *et al*. Effectiveness of case management for 'at risk' patients in primary care: a systematic review and meta-analysis. *PLoS ONE* 2015;10:e0132340.
10 Albertson EM, Chuang E, O'Masta B, *et al*. Systematic review of care coordination interventions linking health and social services for high-utilizing patient populations. *Popul Health Manag* 2022;25:73–85.
11 Paré G, Jaana M, Sicotte C. Systematic review of home telemonitoring for chronic diseases: the evidence base. *J Am Med Inform Assoc* 2007;14:269–77.
12 O'Caoimh R, Cornally N, Weathers E, *et al*. Risk prediction in the community: a systematic review of case-finding instruments that predict adverse Healthcare outcomes in community-dwelling older adults. *Maturitas* 2015;82:3–21.
13 Wallace E, Stuart E, Vaughan N, *et al*. Risk prediction models to predict emergency hospital admission in community-dwelling adults: a systematic review. *Med Care* 2014;52:751–65.
14 Graphnet Health. Carecentric: shared care record solution. 2023. Available: https://www.graphnethealth.com/solutions/integrated-digital-care-record/carecentric/ [Accessed 28 Nov 2023].
15 Prescribing Serives Ltd. Impact assessment 2015. 2015. Available: https://www.eclipsesolutions.org/impactassessment/ [Accessed 28 Nov 2023].
16 Tal E. Target specification bias, counterfactual prediction, and Algorithmic fairness in Healthcare. AIES 2023 - Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society; Montreal QC Canada, 2023
17 DeCamp M, Lindvall C. Latent bias and the implementation of artificial intelligence in medicine. *J Am Med Inform Assoc* 2020;27:2020–3.
18 Flores L, Kim S, Young SD. Addressing bias in artificial intelligence for public health surveillance. *J Med Ethics* 2024;50:190–4.
19 Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. *J Glob Health* 2019;9:020318.
20 Goddard K, Roudsari A, Wyatt JC. Automation bias: empirical results assessing influencing factors. *Int J Med Inform* 2014;83:368–75.
21 Liberati A, Altman DG, Tetzlaff J, *et al*. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate Healthcare interventions: explanation and elaboration. *BMJ* 2009;339:b2700.
22 Zhang J, Symons J, Agapow P, *et al*. Best practices in the real-world data life cycle. *PLOS Digit Health* 2022;1:e0000003.

23 McGinn TG, Guyatt GH, Wyer PC, *et al*. Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-based medicine working group. *JAMA* 2000;284:79–84.
24 Wolff RF, Moons KGM, Riley RD, *et al*. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med* 2019;170:51–8.
25 Snooks H, Bailey-Jones K, Burge-Jones D, *et al*. Predictive risk stratification model: a randomised stepped-wedge trial in primary care (PRISMATIC). *Health Serv Deliv Res* 2018;6:1–164.
26 Wallace E, McDowell R, Bennett K, *et al*. Comparison of count-based multimorbidity measures in predicting emergency admission and functional decline in older community-dwelling adults: a prospective cohort study. *BMJ Open* 2016;6:e013089.
27 Wahls TL, Barnett MJ, Rosenthal GE. Predicting resource utilization in a veterans health administration primary care population comparison of methods based on diagnoses and medications. *Med Care* 2004;42:123–8.
28 Van Houtte C, Gellen C, Ranchhod D. Acute admission risk stratification of New Zealand primary care patients using demographic, Multimorbidity, service usage and modifiable variables. *J Prim Health Care* 2022;14:116–23.
29 Tomlin AM, Lloyd HS, Tilyard MW. Risk stratification of New Zealand general practice patients for emergency admissions in the next year: adapting the Peony model for use in New Zealand. *J Prim Health Care* 2016;8:227–37.
30 Takahashi PY, Pecina JL, Upatising B, *et al*. A randomized controlled trial of telemonitoring in older adults with multiple health issues to prevent hospitalizations and emergency department visits. *Arch Intern Med* 2012;172:773–9.
31 Sylvia ML, Griswold M, Dunbar L, *et al*. Guided care: cost and utilization outcomes in a pilot study. *Dis Manag* 2008;11:29–36.
32 Soto-Gordoa M, de Manuel E, Fullaondo A, *et al*. Impact of stratification on the effectiveness of a comprehensive patient-centered strategy for multimorbid patients. *Health Serv Res* 2019;54:466–73.
33 Sicras-Mainar A, Velasco-Velasco S, Navarro-Artieda R, *et al*. Obtaining the mean relative weights of the cost of care in Catalonia (Spain): retrospective application of the adjusted clinical groups case-mix system in primary health care. *J Eval Clin Pract* 2013;19:267–76.
34 Sicras-Mainar A, Navarro-Artieda R, ACG-BSA RGE. Validating the adjusted clinical groups ACG case-mix system in a Spanish population setting: a multicenter study. *Gac Sanit* 2009;23:228–31.
35 Sicras-Mainar A. Retrospective application of adjusted clinical groups (ACGs) at a primary care centre. *Aten Primaria* 2006;37:439–45.
36 Sibley LM, Moineddin R, Agha MM, *et al*. Risk adjustment using administrative data-based and survey-derived methods for explaining physician utilization. *Med Care* 2010;48:175–82.
37 Shannon GR, Wilber KH, Allen D. Reductions in costly healthcare service utilization: findings from the care advocate program. *J Am Geriatr Soc* 2006;54:1102–7.
38 Shadmi E, Balicer RD, Kinder K, *et al*. Assessing socioeconomic health care utilization inequity in Israel: impact of alternative approaches to morbidity adjustment. *BMC Public Health* 2011;11:609.
39 Reilly S, Abell J, Brand C, *et al*. Case management for people with long-term conditions: impact upon emergency admissions and associated length of stay. *Prim Health Care Res Dev* 2011;12:223–36.
40 Reid RJ, MacWilliam L, Verhulst L, *et al*. Performance of the ACG case-mix system in two Canadian provinces. *Med Care* 2001;39:86–99.
41 Rea F, Corrao G, Ludergnani M, *et al*. A new population-based risk stratification tool was developed and validated for predicting mortality, hospital admissions, and health care costs. *J Clin Epidemiol* 2019;116:62–71.
42 Rahimian F, Salimi-Khorshidi G, Payberah AH, *et al*. Predicting the risk of emergency admission with machine learning: development and validation using linked electronic health records. *PLoS Med* 2018;15:e1002695.
43 Petersen LA, Pietz K, Woodard LD, *et al*. Comparison of the predictive validity of diagnosis-based risk adjusters for clinical outcomes. *Med Care* 2005;43:61–7.
44 Orueta JF, García-Alvarez A, Aurrekoetxea JJ, *et al*. FINGER (forming and identifying new groups of expected risks): developing and validating a new predictive model to identify patients with high Healthcare cost and at risk of admission. *BMJ Open* 2018;8:e019830.
45 Orueta J-F, Urraca J, Berraondo I, *et al*. Adjusted clinical groups (Acgs) explain the utilization of primary care in Spain based on information registered in the medical records: a cross-sectional study. *Health Policy* 2006;76:38–48.
46 Mosley DG, Peterson E, Martin DC. Do Hierarchical condition category model scores predict hospitalization risk in newly enrolled Medicare advantage participants as well as probability of repeated admission scores *J Am Geriatr Soc* 2009;57:2306–10.
47 Maltenfort MG, Chen Y, Forrest CB. Prediction of 30-day pediatric unplanned hospitalizations using the Johns Hopkins adjusted clinical groups risk adjustment system. *PLoS One* 2019;14:e0221233.
48 Lugo-Palacios DG, Hammond J, Allen T, *et al*. The impact of a combinatorial digital and organisational intervention on the management of long-term conditions in UK primary care: a non-randomised evaluation. *BMC Health Serv Res* 2019;19:159.
49 López-Aguilà S, Contel JC, Farré J, *et al*. Predictive model for emergency hospital admission and 6-month readmission. *Am J Manag Care* 2011;17:e348–57.
50 Liu C-F, Sales AE, Sharp ND, *et al*. Case-mix adjusting performance measures in a veteran population: pharmacy-and diagnosis-based approaches. *Health Serv Res* 2003;38:1319–37.
51 Levine S, Steinman BA, Attaway K, *et al*. Home care program for patients at high risk of hospitalization. *Am J Manag Care* 2012;18:e269–76.
52 Lemke KW, Weiner JP, Clark JM. Development and validation of a model for predicting inpatient hospitalization. *Med Care* 2012;50:131–9.
53 Kronick R, Gilmer T, Dreyfus T, *et al*. Improving health-based payment for medicaid beneficiaries: CDPS. *Health Care Financ Rev* 2000;21:29–64.
54 Khanna S, Rolls DA, Boyle J, *et al*. A risk stratification tool for Hospitalisation in Australia using primary care data. *Sci Rep* 2019;9:5011.
55 Jung D, Pollack HA, Konetzka RT. Predicting hospitalization among medicaid Home- and community-based services users using machine learning methods. *J Appl Gerontol* 2023;42:241–51.
56 Juncosa S, Bolíbar B, Roset M. Performance of an ambulatory Casemix measurement system in primary care in Spain: ambulatory care groups (ACGs). *The European Journal of Public Health* 1999;9:27–35.
57 Inouye SK, Zhang Y, Jones RN, *et al*. Risk factors for hospitalization among community-dwelling primary care older patients: development and validation of a predictive model. *Med Care* 2008;46:726–31.
58 Hippisley-Cox J, Coupland C. Predicting risk of emergency admission to hospital using primary care data: derivation and validation of Qadmissions score. *BMJ Open* 2013;3:e003482.
59 Haas LR, Takahashi PY, Shah ND, *et al*. Risk-stratification methods for identifying patients for care coordination. *Am J Manag Care* 2013;19:725–32.
60 Greenwald S, Chamoun GF, Chamoun NG, *et al*. Risk stratification index 3.0, a broad set of models for predicting adverse events during and after hospital admission. *Anesthesiology* 2022;137:673–86.
61 Fishman PA, Goodman MJ, Hornbrook MC, *et al*. Risk adjustment using automated ambulatory Pharmacy data the RxRisk model. *Med Care* 2003;41:84–99.
62 Falasca P, Berardo A, Di Tommaso F. Development and validation of predictive Mosaico (Modello Statistico Combinato) on emergency admissions: can it also identify patients at high risk of frailty? *Ann Ist Super Sanita* 2011;47:220–8.
63 Dudley RA, Medlin CA, Hammann LB, *et al*. The best of both worlds?: potential of hybrid prospective/concurrent risk adjustment. *Med Care* 2003;41:56–69.
64 Donnan PT, Dorward DWT, Mutch B, *et al*. Development and validation of a model for predicting emergency admissions over the next year (PEONY): A UK historical cohort study. *Arch Intern Med* 2008;168:1416–22.
65 Cumming RB, Knutson DJ, Cameron BA, *et al*. A comparative analysis of claims-based methods of health risk assessment for commercial populations. *Society of Actuaries* 2002. Available: https://web.archive.org/web/20201107012013id_/https://www.soa.org/globalassets/assets/Files/Research/Projects/2005-comp-analysis-methods-commercial-populations.pdf
66 Crane SJ, Tung EE, Hanson GJ, *et al*. Use of an electronic administrative database to identify older community dwelling adults at high-risk for hospitalization or emergency department visits: the elders risk assessment index. *BMC Health Serv Res* 2010;10:338.
67 Chen S, Bergman D, Miller K, *et al*. Using applied machine learning to predict healthcare utilization based on socioeconomic determinants of care. *Am J Manag Care* 2020;26:26–31.
68 Charlson ME, Charlson RE, Peterson JC, *et al*. The Charlson comorbidity index is adapted to predict costs of chronic disease in primary care patients. *J Clin Epidemiol* 2008;61:1234–40.

69 Brilleman SL, Gravelle H, Hollinghurst S, *et al*. Keep it simple? Predicting primary health care costs with clinical morbidity measures. *J Health Econ* 2014;35:109–22.
70 Brilleman SL, Salisbury C. Comparing measures of Multimorbidity to predict outcomes in primary care: a cross sectional study. *Fam Pract* 2013;30:172–8.
71 Boult C, Reider L, Leff B, *et al*. The effect of guided care teams on the use of health services: results from a cluster-randomized controlled trial. *Arch Intern Med* 2011;171:460–6.
72 Billings J, Dixon J, Mijanovich T, *et al*. Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients. *BMJ* 2006;333:327.
73 Baker A, Leak P, Ritchie LD, *et al*. Anticipatory care planning and integration: a primary care pilot study aimed at reducing unplanned Hospitalisation. *Br J Gen Pract* 2012;62:e113–20.
74 Ash AS, Ellis RP. Risk-adjusted payment and performance assessment for primary care. *Med Care* 2012;50:643–53.
75 Aguado A, Guinó E, Mukherjee B, *et al*. Variability in prescription drug expenditures explained by adjusted clinical groups (ACG) case-mix: a cross-sectional study of patient electronic records in primary care. *BMC Health Serv Res* 2008;8.
76 Levine S, Adams J, Attaway K, *et al*. Predicting the financial risks of seriously ill patients. 2011. Available: https://www.hmprg.org/wp-content/themes/HMPRG/backup/GOHIT/Work%20Groups/IDSR/August%2027%20-%20Predicting%20the%20Financial%20Risks%20of%20Seriously%20Ill%20Patients.pdf [Accessed 28 Nov 2023].
77 Hutchings HA, Evans BA, Fitzsimmons D, *et al*. Predictive risk stratification model: a progressive cluster-randomised trial in chronic conditions management (PRISMATIC) research protocol. *Trials* 2013;14:301.
78 Billings J, Mijanovich T, Dixon J, *et al*. Case finding Algorithms for patients at risk of re-Hospitalisation PARR1 and PARR2. 2006. Available: https://www.kingsfund.org.uk/sites/default/files/field/field_document/PARR-case-finding-algorithms-feb06.pdf [Accessed 28 Nov 2023].
79 Wilber KH, Allen D, Shannon GR, *et al*. Partnering managed care and community-based services for frail elders: the care advocate program. *J Am Geriatr Soc* 2003;51:807–12.
80 Johns Hopkins Medicine. Johns Hopkins ACG system. 2023. Available: https://www.hopkinsacg.org/ [Accessed 28 Nov 2023].
81 Pope G, Ellis R, Ash A, *et al*. Diagnostic cost group hierarchical condition category models for Medicare risk adjustment. Health Care Financing Administration. 2000. Available: https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/reports/downloads/pope_2000_2.pdf [Accessed 28 Nov 2023].
82 Optum. Symmetry episode risk groups. 2022.
83 Wennberg D, Dixon J, Billings J, *et al*. Combined predictive model final report. 2006. Available: https://www.kingsfund.org.uk/sites/default/files/field/field_document/PARR-combined-predictive-model-final-report-dec06.pdf [Accessed 28 Nov 2023].
84 Charlson ME, Pompei P, Ales KL, *et al*. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373–83.
85 Exley J, Abel GA, Fernandez J-L, *et al*. Impact of the Southwark and Lambeth integrated care older people's programme on hospital utilisation and costs: controlled time series and cost-consequence analysis. *BMJ Open* 2019;9:e024220.
86 Gravelle H, Dusheiko M, Sheaff R, *et al*. Impact of case management (Evercare) on frail elderly patients: controlled before and after analysis of quantitative outcome data. *BMJ* 2007;334:31.
87 Stokes J, Kristensen SR, Checkland K, *et al*. BMJ open effectiveness of multidisciplinary team case management: difference-Indifferences analysis. *BMJ Open* 2016;6:e010468.
88 Tricco AC, Antony J, Ivers NM, *et al*. Effectiveness of quality improvement strategies for coordination of care to reduce use of health care services: a systematic review and meta-analysis. *CMAJ* 2014;186:E568–78.
89 Vanderplasschen W, Rapp RC, De Maeyer J, *et al*. A meta-analysis of the efficacy of case management for substance use disorders: a recovery perspective. *Front Psychiatry* 2019;10:186.
90 Backhouse A, Ukoumunne OC, Richards DA, *et al*. The effectiveness of community-based coordinating interventions in dementia care: a meta-analysis and subgroup analysis of intervention components. *BMC Health Serv Res* 2017;17:717.
91 Saragih ID, Saragih IS, Tarihoran DETAU, *et al*. A meta-analysis of studies of the effects of case management intervention for stroke survivors across three countries. *J Nurs Scholarsh* 2023;55:345–55.
92 Gorin SS, Haggstrom D, Han PKJ, *et al*. Cancer care coordination: a systematic review and meta-analysis of over 30 years of empirical studies. *Ann Behav Med* 2017;51:532–46.
93 Scholte NTB, Gürgöze MT, Aydin D, *et al*. Telemonitoring for heart failure: a meta-analysis. *Eur Heart J* 2023;44:2911–26.
94 Umeh CA, Torbela A, Saigal S, *et al*. Telemonitoring in heart failure patients: systematic review and meta-analysis of randomized controlled trials. *World J Cardiol* 2022;14:640–56.

BMJ Health & Care Informatics

Artificial intelligence in healthcare: Opportunities come with landmines

Usman Iqbal ,[1,2,3] Yi-Hsin Elsa Hsu,[4,5,6,7] Leo Anthony Celi ,[8,9,10] Yu-Chuan (Jack) Li [3,11,12,13]

To cite: Iqbal U, Hsu Y-HE, Celi LA, *et al*. Artificial intelligence in healthcare: Opportunities come with landmines. *BMJ Health Care Inform* 2024;**31**:e101086. doi:10.1136/bmjhci-2024-101086

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjhci-2024-101086).

Received 03 April 2024
Accepted 02 May 2024

For numbered affiliations see end of article.

Correspondence to
Prof Yu-Chuan (Jack) Li; jack@tmu.edu.tw

In the ever-evolving landscape of healthcare, the convergence of artificial intelligence (AI) within breast cancer screening and the transformative potential of natural language processing (NLP) in ensuring patient safety stands as a testament to ground-breaking progress.[1 2] The seamless integration of AI technologies in radiology is reshaping diagnostic precision while NLP's capacity to decipher and enhance safety protocols heralds a new era in healthcare innovation.[3 4]

Contrary to common assumptions, the presence of AI does not necessarily guarantee improved efficiency or accuracy in interpreting medical images. It is concerning that AI's identification of potential errors can paradoxically lead some radiologists to make more mistakes and spend more time analysing images, highlighting the dangers of developing AI systems in isolation.[5] This underscores the crucial need for designing collaborative human-AI systems rather than standalone AI solutions, as the full extent of AI's influence on human behaviour remains unpredictable. Moreover, there is also a critical concern regarding patient safety as a matter of health equity, shedding light on the disparities in medical errors and treatment injuries exacerbated by social determinants of care. It calls for a holistic approach to healthcare delivery that prioritises equity and inclusivity, ensuring that all patients receive the highest standard of care irrespective of their social circumstances.

The two 'editor's choice' articles highlight how crucial it is to embrace AI in breast cancer screening and NLP in enhancing patient safety in healthcare's dynamic landscape. Högberg *et al*[6] studied an insightful exploration into the potential and challenges associated with AI in breast radiology. The Swedish breast radiologists' perspective on AI in mammography screening revealed an overwhelmingly positive attitude towards its incorporation, highlighting the potential to enhance efficiency in diagnostic processes. However, alongside this optimism, the study uncovered a labyrinth of uncertainties and diverse viewpoints. Concerns loomed over potential risks ranging from medical outcomes to the reshaping of working conditions and crucial uncertainties regarding the assignment of responsibility in AI-mediated medical decision-making.[7] The complexity of delineating accountability between AI systems, radiologists and healthcare providers emerged as a pivotal issue demanding resolution.

Addressing these intricacies is paramount for harnessing AI's potential while upholding the integrity of patient care and professional practice in the evolving landscape of breast radiology.[8–10] Most professionals favoured AI as a supportive tool, but divergent opinions arose regarding its optimal integration into the screening workflow. The authors delineated varied views on AI's impact within the profession, stressing the absence of consensus on the extent of change and the consequent transformation of breast radiologists' roles.[6] Collaboration between human radiologists and AI assistance in radiology, expected to heavily impact the field, is under investigation. While AI tools show promise, biases in human use of AI limit potential gains. Radiologists should either solely rely on AI or work independently, rather than collaboratively.[5] Additionally, optimal delegation policies are proposed, considering time costs and suboptimal use of AI information. Future research should explore AI-specific training for radiologists and organisational factors influencing human–AI collaboration. A pressing need exists to address multifaceted challenges, particularly in establishing clear ethical, legal and social frameworks governing AI integration in radiology.

In the second study by Tabaie *et al*,[11] uncovered crucial contributing factors from patient safety event reports, showcasing the

transformative potential of NLP algorithms in healthcare insights. This study's findings involved identifying and categorising contributing factors within a decade's worth of self-reported patient safety events from a multihospital healthcare system. These contributing factors pivotal in precipitating or permitting patient safety incidents, often remain concealed within the intricate narratives of these reports. The authors introduced a method to extract 'information-rich sentences' from reports, unveiling hidden contributing factors and refining their categorisation using NLP, leveraging unstructured data in patient safety event reports to isolate crucial sentences defining contributing factors.[11] Automating the identification and categorisation of contributing factors empowers healthcare systems to proactively address safety concerns, fostering quicker responses and continuous improvement. However, the study's reliance on data from a singular health system prompts inquiries about its generalisability. As healthcare increasingly embraces data-driven decision-making, harnessing NLP emerges as a pivotal strategy in safeguarding patient well-being.[12–14] The findings call for further exploration and adoption of NLP-driven approaches to enhance patient safety initiatives globally.

While both studies mark significant strides in healthcare, certain considerations arise.[6 11] The study on AI integration in breast radiology highlights uncertainties and the need for collaborative efforts in establishing clear governance frameworks. The retrospective nature of the NLP study calls for real-time validation and raises concerns about generalisability beyond a singular healthcare system.

Nonetheless, these studies underscore the transformative potential of technology in reshaping healthcare paradigms. Embracing AI in breast cancer screening and leveraging NLP for patient safety initiatives open avenues for proactive, data-driven decision-making. Further evaluation, exploration and widespread adoption of these technologies throughout their life cycle are pivotal in promoting patient safety and elevating healthcare quality, emphasising the central focus on integrating fairness and equity globally within healthcare.[15 16]

Author affiliations
[1]School of Population Health, Faculty of Medicine and Health, University of New South Wales (UNSW), Sydney, NSW, Australia
[2]Global Health and Health Security Department, College of Public Health, Taipei Medical University, Taipei, Taiwan
[3]International Center for Health Information and Technology, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan
[4]Biotechnology Executive Master's Degree in Business Administration (BioTech EMBA), Taipei Medical University, Taipei, Taiwan
[5]School of Healthcare Administration, College of Management, Taipei Medical University, Taipei, Taiwan
[6]International Ph.D. Program in BioTech and Healthcare Management, College of Management, Taipei Medical University, Taipei, Taiwan
[7]Department of Humanities in Medicine, College of Medicine, School of Medicine, Taipei Medical University, Taipei, Taiwan
[8]Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[9]Division of Pulmonary, Critical Care and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA
[10]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
[11]Graduate Institute of Biomedical Informatics, College of Medical Science & Technology, Taipei Medical University, Taipei, Taiwan
[12]Department of Dermatology, Taipei Municipal Wanfang Hospital, Taipei, Taiwan
[13]The International Medical Informatics Association (IMIA), Zürich, Switzerland

X Leo Anthony Celi @MITCriticalData

Contributors UI drafted the initial manuscript. Supervision was provided by Y-HEH, LAC and Y-CL. All authors approved the final manuscript as submitted and agreed to be accountable for all aspects of the work.

Funding LAC is funded by the National Institute of Health through R01 EB017205, DS-I Africa U54 TW012043-01 and Bridge2AI OT2OD032701, and the National Science Foundation through ITEST #2148451.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Usman Iqbal http://orcid.org/0000-0002-0614-123X
Leo Anthony Celi http://orcid.org/0000-0001-6712-6626
Yu-Chuan (Jack) Li http://orcid.org/0000-0001-6497-4232

REFERENCES

1 Dembrower K, Crippa A, Colón E, *et al*. Artificial intelligence for breast cancer detection in screening Mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study. *Lancet Digit Health* 2023;5:e703–11.
2 Boxley C, Fujimoto M, Ratwani RM, *et al*. A text mining approach to categorize patient safety event reports by medication error type. *Sci Rep* 2023;13:18354.
3 Zheng C, Duffy J, Liu I-LA, *et al*. Identifying cases of shoulder injury related to vaccine Administration (SIRVA) in the United States: development and validation of a natural language processing method. *JMIR Public Health Surveill* 2022;8:e30426.
4 Bitencourt A, Daimiel Naranjo I, Lo Gullo R, *et al*. AI-enhanced breast imaging: where are we and where are we heading. *Eur J Radiol* 2021;142:109882.
5 Agarwal N, Moehring A, Rajpurkar P, *et al*. Combining human expertise with artificial intelligence: experimental evidence from Radiology. *National Bureau of Economic Research* 2023.
6 Högberg C, Larsson S, Lång K. Anticipating artificial intelligence in mammography screening: views of Swedish breast Radiologists. *BMJ Health Care Inform* 2023;30:e100712.
7 Lokaj B, Pugliese M-T, Kinkel K, *et al*. Barriers and Facilitators of artificial intelligence conception and implementation for breast imaging diagnosis in clinical practice: a Scoping review. *Eur Radiol* 2024;34:2096–109.
8 Hickman SE, Baxter GC, Gilbert FJ. Adoption of artificial intelligence in breast imaging: evaluation, ethical constraints and limitations. *Br J Cancer* 2021;125:15–22.
9 Eriksson M, Román M, Gräwingholt A, *et al*. European validation of an image-derived AI-based short-term risk model for individualized breast cancer screening—a nested case-control study. The Lancet Regional Health – Europe,
10 Gichoya JW, Thomas K, Celi LA, *et al*. AI pitfalls and what not to do: mitigating bias in AI. *Br J Radiol* 2023;96:20230023.
11 Tabaie A, Sengupta S, Pruitt ZM, *et al*. A natural language processing approach to Categorise contributing factors from patient safety event reports. *BMJ Health Care Inform* 2023;30:e100731.
12 Fong A. Realizing the power of text mining and natural language processing for analyzing patient safety event narratives: the challenges and path forward. *J Patient Saf* 2021;17:e834–6.

13 Melton GB, Hripcsak G. Automated detection of adverse events using natural language processing of discharge summaries. *J Am Med Inform Assoc* 2005;12:448–57.
14 Ozonoff A, Milliren CE, Fournier K, *et al*. Electronic surveillance of patient safety events using natural language processing. *Health Informatics J* 2022;28.
15 Wawira Gichoya J, McCoy LG, Celi LA, *et al*. Equity in essence: a call for operationalising fairness in machine learning for healthcare. *BMJ Health Care Inform* 2021;28:e100289.
16 Shortliffe EH. Role of evaluation throughout the life cycle of biomedical and health AI applications. *BMJ Health Care Inform* 2023;30:e100925.