Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Protocol for a Multimethod Evaluation of AI-Based Colorectal Cancer Risk Prediction in an Emulated Primary Care Electronic Health Record Environment: Trust, Usability, and Diagnostic Decision-Making

Eliya Abedi

eliya.abedi@ki.se

Karolinska Institutet

Andrew Dwi Permana

KTH Royal Institute of Technology

Elinor Nemlander

Karolinska Institutet

Adam S. Darwich

KTH Royal Institute of Technology

Annika Sjövall

Karolinska University Hospital

Axel C. Carlsson

Karolinska Institutet

Andreas Rosenblad

Regional Cancer Centre Stockholm Gotland

Tina Gustavell

Karolinska Institutet

Marcela Ewing

Regional cancer centre West

Jayanth Raghothama

KTH Royal Institute of Technology

Jan Hasselström

Karolinska Institutet

Study protocol

Keywords: Colorectal cancer, artificial intelligence, primary care, electronic health record, risk prediction, diagnostic decision-making, usability, trust, clinical decision support

Posted Date: October 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10392968/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Colorectal cancer is one of the most frequently missed cancer diagnoses in primary care. Artificial intelligence (AI)-based colorectal cancer risk prediction may support earlier detection, but its influence on general practitioners' (GPs) diagnostic reasoning and decision-making has not been thoroughly evaluated. This study aims to evaluate how GPs interpret and use AI-based colorectal cancer risk predictions in an emulated electronic health record environment.

Methods

This multimethod study is conducted in an emulated primary care electronic health record environment using historical electronic health record data. GPs assess 40–60 de-identified patient cases in a randomised patient–doctor design, with AI-based risk predictions displayed for a randomly allocated subset of assessments. Quantitative outcomes include suspicion of colorectal cancer, strength of suspicion, intended management, and ICD-10 coding. Semi-structured interviews explore GPs' perceptions of trust, usability, interpretability, and the influence of AI-based risk predictions on diagnostic decision-making. At the time of submission, all ten participating GPs had completed the assessments and interviews, and data analyses had commenced but were not yet complete.

Discussion

The study will provide evidence on how GPs interpret and use AI-based colorectal cancer risk predictions within an emulated electronic health record environment, supporting future prospective evaluation in routine primary care. The findings will guide refinement of the intervention and inform the development of a future prospective cluster-randomised trial evaluating AI-based clinical decision support in routine primary care.

Trial registration

Not applicable. This study involves an emulated electronic health record evaluation without patient intervention.

Background

Colorectal cancer is the third most commonly diagnosed malignancy worldwide [1]. Despite screening, most patients are diagnosed following symptomatic presentation [2]. Diagnosing colorectal cancer in primary care remains challenging, as it represents one of the most frequently missed or delayed cancer diagnoses in this setting [3–5].

Several risk assessment tools have been developed to assist general practitioners (GPs) in identifying patients with an increased likelihood of underlying colorectal cancer; however, evidence that they improve clinical outcomes is limited [6]. Research on artificial intelligence (AI)-based cancer prediction is rapidly expanding, yet challenges remain regarding sample sizes, input variables, generalisability, and interpretability [7, 8]. In Sweden, colorectal cancer risk assessment tools mainly rely on diagnostic codes from primary care [9–11], potentially limiting precision and introducing bias [12, 13]. Integrating coded data with free-text electronic health record information and laboratory results may enhance predictive accuracy.

AI-based colorectal cancer risk prediction models have not been evaluated in the clinical context for which they are intended. Robust evaluation within a complex intervention framework is needed to understand their potential impact on clinical decision-making and to support their further development and evaluation. The UK Medical Research Council framework for complex interventions highlights the importance of intervention development and evaluation before definitive testing [14]. The GUIDES checklist emphasizes the importance of considering context, content, system, and implementation factors when developing clinical decision support systems [15].

This study evaluates how general practitioners interpret and use AI-based colorectal cancer risk predictions within an emulated primary care electronic health record environment. By assessing usability, interpretability, and the influence of AI predictions on clinical decision-making, this study aims to inform the further development and evaluation of AI-based colorectal cancer risk prediction in routine primary care.

Methods and analysis

This is a multimethod, exploratory study designed to evaluate how AI-based colorectal cancer risk predictions influence GPs' diagnostic reasoning and intended management in an emulated primary care electronic health record environment.

Objectives

The primary objective is to evaluate whether, and in what ways, access to AI-based colorectal cancer risk prediction models influences GPs' diagnostic reasoning and clinical decision-making within an emulated primary care electronic health record environment. This work aims to inform the future development and evaluation of AI-based colorectal cancer risk prediction in primary care.

The secondary objectives are to:

- Define requirements for the future refinement of the AI-based colorectal cancer risk prediction model and its integration into a clinical decision support system for colorectal cancer in primary care.

- Assess concordance between participants' intended management strategies and actions documented in the original electronic health record notes, with and without access to AI-based risk predictions, and between models using data with or without truncation at three months prior to diagnosis.
- Identify the symptoms and clinical findings GPs rely on in their diagnostic reasoning, and their alignment with AI-based risk predictions.
- Explore GPs' perceptions of the usability, difficulties, and impact on clinical reasoning of the models.
- Compare agreement between GP specialists and residents in diagnostic suspicion, strength of suspicion, intended management strategy, and International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10) coding.

We developed AI-based colorectal cancer risk prediction models using unstructured free-text entries from primary care electronic health records together with structured data. Structured data included consultation patterns, diagnostic and symptom codes, laboratory test results, smoking and alcohol status, age, weight, and height. Two models were trained: one utilising all available data from the year preceding diagnosis (model 1), and another excluding data from the final three months prior to diagnosis (model 2). The full dataset for model development included 2,992 colorectal cancer cases and 11,915 controls, matched on age, sex, and primary care centre. These were split into a training set containing 2,551 cases and their corresponding controls. The remaining 441 cases formed the test set used in this emulated trial. Both models were trained and tested on data based on individual visits.

This multimethod study includes a quantitative assessment of diagnostic reasoning tasks completed by GPs and qualitative semi-structured interviews exploring perceptions of AI-based risk predictions. The study employs a randomised evaluation design, in which de-identified electronic health record entries of patients with colorectal cancer are assessed by GPs. The qualitative assessment will provide an overarching evaluation of the study procedures and the usability, interpretability, and perceived utility of the AI-based models [16].

Study setting and participants

As this exploratory evaluation was not designed for hypothesis testing, a formal power calculation was not undertaken. The sample size of ten GPs was determined based on feasibility considerations and was considered sufficient to evaluate operational procedures, estimate variability in key outcomes, and provide adequate depth for the qualitative analysis, while remaining manageable within the constraints of an intensive emulated review process. Ten doctors, including specialists and residents, were recruited from three university-affiliated primary care centres in Stockholm. All doctors will be referred to as GPs in the study.

Patient and public involvement

Patients and the public were not involved in the design, conduct, or reporting of this study. The research is based on secondary analysis of de-identified electronic health record data within an emulated primary care environment and focuses on GPs' diagnostic reasoning. Patient and public involvement will be considered in later phases of this research programme to support interpretation and future implementation in routine primary care.

Case material and study procedures

Each participating GP will assess 40–60 de-identified, unique colorectal cancer cases. Evaluations will first be conducted without model predictions (control arm), followed by assessments incorporating randomised predictions balanced between model 1 and model 2 (trial arm; Fig. 1).

Prior to testing, the 441 test-set cases will be randomised across 10 GPs, each receiving around 60 cases: 30 in the control arm and 30 in the trial arm. GPs will be blinded to model allocation and will be unaware that all cases included in the study are patients with colorectal cancer and that two distinct models are being evaluated. A user interface will display patient records in a longitudinal format. The interface will highlight key information, including demographic details, smoking status, alcohol use, laboratory test results, diagnostic problems, requested examinations, case history, and doctors' notes, beginning with the earliest entry within the year prior to diagnosis. Extracts will be presented in an emulated electronic health record user interface replicating the primary care electronic health record environment while preserving patient confidentiality (Fig. 2).

Randomisation will be applied at the patient-doctor level, such that the same patient may be reviewed with an AI-based prediction by one GP and without by another (Fig. 3).

Predictions will be presented as percentages indicating risk of colorectal cancer, followed by the features that were incorporated into the risk assessment. Responses will be stored in a secure database. GPs will provide demographic and professional information, including sex, age, years of GP experience, any non-routine clinical or other professional duties, and specialist or resident status. Each GP will be assigned a pseudonymised assessor-identifier (ID). All GPs will first assess two example cases not included in the study for each arm, to familiarise themselves with the interface.

For each electronic health record entry across the timeline of the individual patient record, in regard to suspicion of colorectal cancer, GPs will be asked to:

1. Indicate whether they suspect colorectal cancer (yes/no). If yes, rate the level of suspicion (very low, low, moderate, high, or very high).
2. Select their intended management strategy by rating as: No action, Further investigation in primary care, Referral for radiology, Referral for colonoscopy, or Referral for specialist investigation
3. Record the action documented by the GP in the original electronic health record note (using the same categories as in b, or "I don't know").

4. Identify the symptoms and findings underpinning their decisions by writing a free-text response.
5. Assign an ICD-10 code they would link to the note.

All responses will be recorded and linked to the GP's assessor-ID via the interface and stored in the secure database. GPs will provide their assessments based on the displayed information, and all responses from their evaluations will be recorded. Patient records will be presented sequentially, preserving previous visits to allow GPs to review them if necessary.

For patients with multiple visits in the electronic health record, the interface will display a "Next Visit" button that allows navigation to the next entry for the same patient. If a patient has only one entry or the final entry for a given patient has been reached, a "Next Patient" button will appear to proceed to the next case. GPs must submit their assessment before moving to the next entry/patient by clicking the "Submit Assessment" button. If an attempt is made to proceed before submitting the assessment, a warning message will pop up on the interface, reminding the user to complete the assessment.

Individual semi-structured interviews will be conducted with GPs immediately after completing the emulated electronic health record tasks. Guided by a predefined interview guide with open-ended and probing questions, interviews will capture contextual factors and non-verbal cues. Sessions will be held privately, audio-recorded with informed consent, and accompanied by field notes. Interviews are expected to last 30–60 minutes and will be transcribed verbatim, with transcripts pseudonymised using the same assessor-ID as the quantitative data.

Statistical analyses

The analyses will compare outcomes with and without access to AI-based risk predictions in relation to the true colorectal cancer status of the cases, and will also compare outcomes between model 1 and model 2. Outcomes of main interest include suspicion of colorectal cancer, strength of suspicion, intended management, and ICD-10 coding. Comparisons with original electronic health record actions and inter-rater differences due to GP characteristics will also be performed. Analyses will be performed using univariate methods, relative risk of marginal effects [17], and regression methods such as generalised estimating equations. Inferential analyses will account for the inherent dependency in the data arising because each GP reviews multiple patients and each patient is reviewed by multiple GPs.

Qualitative analyses

The interviews will be analysed using conventional content analysis [18]. All interviews will be read to gain an overall understanding, followed by coding of key concepts. Codes will be organised into categories and refined into broader themes. Each theme will be defined and illustrated with representative quotes.

Data management and rigour

All data will be stored on secure, access-controlled servers in accordance with Swedish data protection regulations. Participants will be assigned assessor-IDs, and no personal identifiers will be stored. Rigour for quantitative data will be ensured through standardised data collection procedures, automated data entry checks, and validation of responses to minimise errors. Rigour for qualitative data will be ensured by independent coding by more than one researcher, regular team discussions to resolve discrepancies, and maintaining an audit trail of analytic decisions. Credibility will be strengthened by using illustrative data extracts.

Study progress

At the time of manuscript submission, all ten GPs had completed the assessments and interviews according to the protocol without major issues (Fig. 4). Data analyses had commenced but were not yet complete. Submission was intentionally deferred until all GPs had completed the protocol to avoid jeopardising the study's blinding procedures. No study results have been submitted for publication or published.

Discussion

This multimethod study has been designed to evaluate how general practitioners interpret and use AI-based colorectal cancer risk predictions in an emulated primary care electronic health record environment. Beyond examining effects on diagnostic reasoning, the study also evaluates the usability, interpretability, and acceptability of AI-based risk predictions, together with the practicality of the study procedures.

Several practical considerations have already been identified during the conduct of the study. The use of an emulated electronic health record environment based on historical real-world cases allows standardised exposure to comparable clinical materials across participants while avoiding risks to patient safety.

The organisation of study visits has also provided important insights for future study design. Individual and paired scheduling of participants has been feasible; however, paired sessions required access to separate assigned workstations and independent case sets in order to maintain assessor independence. Conducting the semi-structured interviews immediately after case assessments proved practical and reduced participant burden. However, careful time planning was required to ensure that both the assessment and interview components could be completed within a single session.

Another important consideration relates to the integration of AI-based risk predictions into clinical workflow. In this study, AI-based risk predictions are presented in a controlled and standardised format. This will introduce additional practical considerations, including integration into existing electronic health record systems, alert presentation, consultation time constraints, and competing clinical priorities.

Insights from the current study regarding how GPs interpret and use the AI-based risk predictions will inform interface design and implementation strategies for the real-world trial.

The present study is not designed to assess effectiveness or patient outcomes. Instead, its purpose is to generate knowledge about how clinicians interpret and use AI-based risk predictions, and to support refinement of the intervention for future clinical evaluation. The knowledge gained from this study will inform the design of a prospective cluster-randomised trial evaluating the intervention in routine primary care. This includes refining study procedures, identifying appropriate outcome measures, and understanding potential barriers and facilitators to implementation. In this future trial, the intervention will be implemented in a real-world primary care setting using prospective patient encounters.

A limitation of the current feasibility design is that the emulated environment cannot fully replicate real clinical workflow, and repeated exposure to colorectal cancer cases does not reflect real-world prevalence. However, this design enables controlled comparison of AI-assisted and non-AI-assisted decision-making and ensures that all participants encounter diagnostically relevant scenarios.

Overall, this study will provide essential information on the acceptability, usability, and practical requirements of implementing AI-based colorectal cancer risk prediction in primary care. The findings will directly inform the design, logistics, and outcome selection of a subsequent full-scale cluster-randomised trial in real-world clinical practice. They will also contribute to the safe and effective introduction of AI-based clinical decision support for earlier colorectal cancer detection.

Abbreviations

AI
artificial intelligence
GP
general practitioner
ICD-10
International Statistical Classification of Diseases and Related Health Problems,10th Revision
ID
identifier
RCT
randomised controlled trial

Declarations

Ethics approval and consent to participate

Ethical approval has been obtained from the Swedish Ethical Review Authority (2020–06037,

Consent for publication

Not applicable.

Funding

This study was supported by the Sjöberg and Belvén Foundations, the NSV grant (FoUI-977323) from Region Stockholm, and the Digital Futures grant at KTH Royal Institute of Technology. In addition, Eliya Abedi received a grant from Region Stockholm supporting PhD research. The funders had no role in the design, conduct, analysis, or reporting of the study. All authors are affiliated with publicly funded institutions (Academic Primary Health Care Center Stockholm, Karolinska Institutet, KTH, Regional Cancer Centre West, Region Stockholm, University of Gothenburg and Uppsala University).

Author Contribution

EA, ADP, and AR drafted the manuscript, while EN, ASD, AS, ACC, TG, ME, JR, and JH critically revised it. EA, AR, EN, AS, ACC, ME, and JH contributed to data collection, and ADP, AR, ASD, and JR managed the data. All authors contributed to the study design.

Acknowledgement

We thank the participating primary care centres and GPs for collaboration.

Data Availability

Due to data protection regulations, the de-identified electronic health record dataset is not publicly available but may be accessible on reasonable request and subject to ethical approval.

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A (2024) Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 74(3):229–263
2. Cardoso R, Guo F, Heisser T, De Schutter H, Van Damme N, Nilbert MC, Tybjerg AJ, Bouvier AM, Bouvier V, Launoy G et al (2022) Proportion and stage distribution of screen-detected and non-screen-detected colorectal cancer in nine European countries: an international, population-based study. Lancet Gastroenterol Hepatol 7(8):711–723
3. Fernholm R, Pukk Härenstam K, Wachtler C, Nilsson GH, Holzmann MJ, Carlsson AC (2019) Diagnostic errors reported in primary healthcare and emergency departments: A retrospective and

descriptive cohort study of 4830 reported cases of preventable harm in Sweden. Eur J Gen Pract 25(3):128–135

4. Singh H, Schiff GD, Graber ML, Onakpoya I, Thompson MJ (2017) The global burden of diagnostic errors in primary care. BMJ Qual Saf 26(6):484–494
5. Aaronson EL, Quinn GR, Wong CI, Murray AM, Petty CR, Einbinder J, Schiff GD (2019) Missed diagnosis of cancer in primary care: Insights from malpractice claims data. J Healthc Risk Manag 39(2):19–29
6. Medina-Lara A, Grigore B, Lewis R, Peters J, Price S, Landa P, Robinson S, Neal R, Hamilton W, Spencer AE (2020) Cancer diagnostic tools to aid decision-making in primary care: mixed-methods systematic reviews and cost-effectiveness analysis. Health Technol Assess (Winchester Eng) 24(66):1–332
7. Felici A, Peduzzi G, Pellungrini R, Campa D (2025) Artificial intelligence to predict cancer risk, are we there yet? A comprehensive review across cancer types. Eur J Cancer 222:115440
8. Nemlander E, Abedi E, Hasselström J, Carlsson AC, Rosenblad A (2025) Letter RE: artificial intelligence to predict cancer risk, are we there yet? a comprehensive review across cancer types. Eur J Cancer 228:115715
9. Ewing M, Naredi P, Zhang C, Månsson J (2016) Identification of patients with non-metastatic colorectal cancer in primary care: a case-control study. Br J Gen Practice: J Royal Coll Gen Practitioners 66(653):e880–e886
10. Abedi E, Ewing M, Nemlander E, Hasselström J, Sjövall A, Carlsson AC, Rosenblad A A machine learning tool for identifying metastatic colorectal cancer in primary care. Scand J Prim Health Care 2025:1–9
11. Nemlander E, Ewing M, Abedi E, Hasselström J, Sjövall A, Carlsson AC, Rosenblad A (2023) A machine learning tool for identifying non-metastatic colorectal cancer in primary care. Eur J Cancer 0(0)
12. Nemlander E, Ewing M, Carlsson AC, Rosenblad A (2023) Transforming early cancer detection in primary care: harnessing the power of machine learning. Oncoscience 10:20–21
13. Al-Sahab B, Leviton A, Loddenkemper T, Paneth N, Zhang B (2024) Biases in Electronic Health Records Data for Generating Real-World Evidence: An Overview. J Healthc Inf Res 8(1):121–139
14. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, Boyd KA, Craig N, French DP, McIntosh E et al (2024) A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. Int J Nurs Stud 154:104705
15. Van de Velde S, Kunnamo I, Roshanov P, Kortteisto T, Aertgeerts B, Vandvik PO, Flottorp S (2018) The GUIDES checklist: development of a tool to improve the successful use of guideline-based computerised clinical decision support. Implement Sci 13(1):86
16. Eldridge SM, Lancaster GA, Campbell MJ, Thabane L, Hopewell S, Coleman CL, Bond CM (2016) Defining Feasibility and Pilot Studies in Preparation for Randomised Controlled Trials: Development of a Conceptual Framework. PLoS ONE 11(3):e0150205

17. Rosenblad AK (2025) A note on two novel easy-to-interpret feature effect measures for partial dependence plots in a classification setting. J Applied Statistics 0(0):1–13
18. Hsieh HF, Shannon SE (2005) Three approaches to qualitative content analysis. Qual Health Res 15(9):1277–1288

Figures

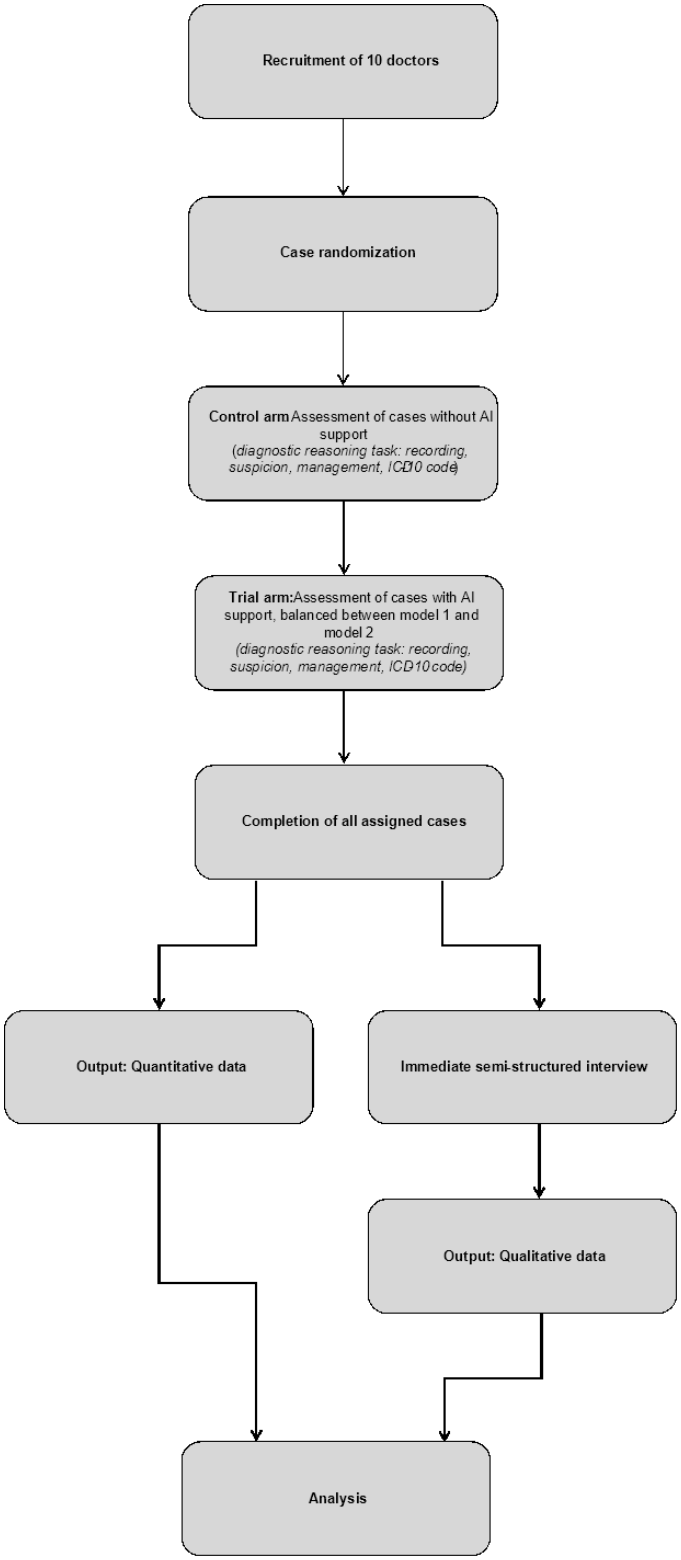

Figure 1

Flowchart illustrating the study procedure.

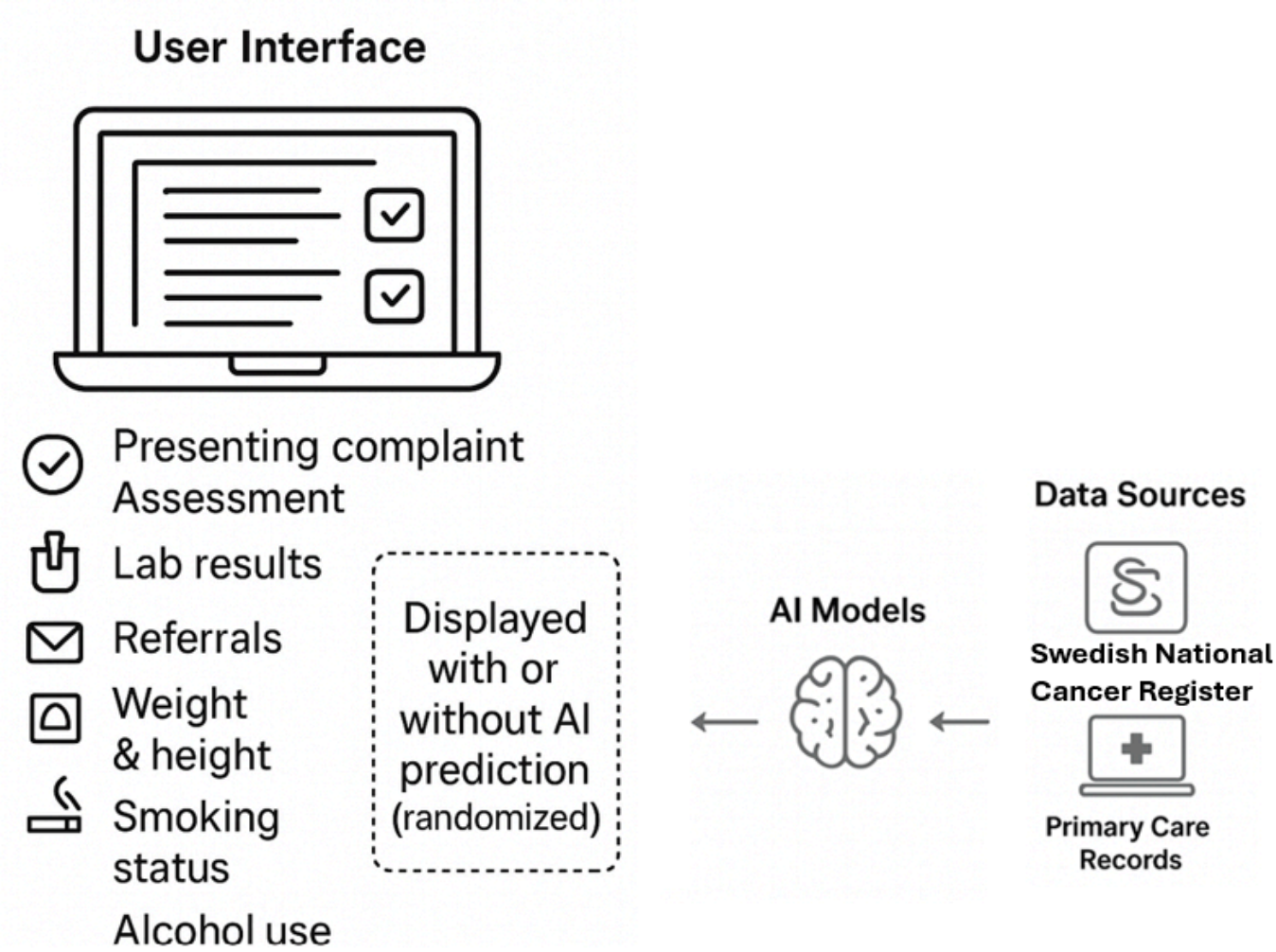

Figure 2

Illustration of what will be displayed in the interface emulating the electronic health record.

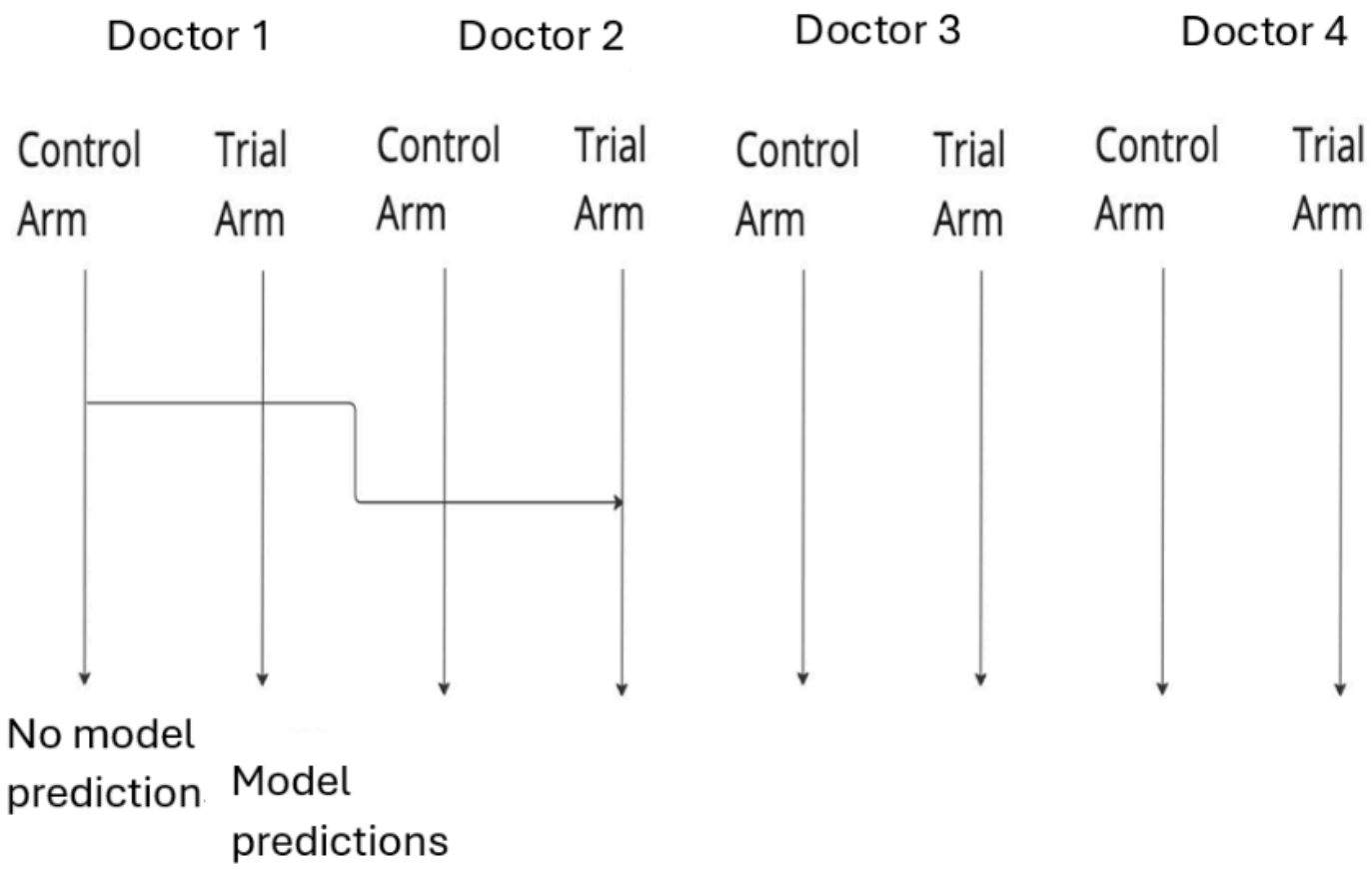

Figure 3

Illustration of how cases can appear in the control arm (without model prediction) for one doctor, and in the trial arm (with model prediction) of another.

Study Visit	1	2	3	4	5	6	7
	GP 1	GP 2	GP 3	GP 4	GPs 5 & 6	GPs 7 & 8	GPs 9 & 10
EHR assesment	■	■	■	■	■ ■	■ ■	■ ■
Interview	■	■	■	■	■ ■	■ ■	■ ■
Analysis	◊	◊	◊	◊	◊ ◊	◊ ◊	◊ ◊

Legend: ■ Completed ◊Pending

Figure 4

Study visits and GP participation in the feasibility study. All electronic health record (EHR) assessments were conducted in an emulated primary care environment. Interviews were conducted immediately following each GP's assessment.