RESEARCH PROTOCOL

EVALUATING THE IMPACT OF ARTIFICIAL INTELLIGENCE TRIAGE IN ONLINE CONSULTATIONS TO REDUCE DELAYS IN URGENT PRIMARY CARE: INTERRUPTED TIME SERIES ANALYSIS AND QUANTITATIVE PROCESS EVALUATION

(AI TRIAGE IMPACT)

Contents

1) RESEARCH TEAM & KEY CONTACTS

Chief Investigator: Name: Benjamin Brown Address: Centre for Primary Care and Health Services Research Williamson Building University of Manchester M13 9PL Email: benjamin.brown@manchester.ac.uk	***Co-investigators:*** Name: Niels Peek Address: THIS Institute University of Cambridge Strangeways Research Laboratory 2 Worts' Causeway Cambridge CB1 8RN Email: niels.peek@thisinstitute.cam.ac.uk Name: Evan Kontopantelis Address: Centre for Primary Care and Health Services Research Williamson Building University of Manchester M13 9PL Email: evan.kontopantelis@manchester.ac.uk
Sponsor: Name: The University of Manchester Contact: Mohammed Zubair, Research Governance, Ethics, and Integrity Manager Address: Directorate of Research and Business Engagement The University of Manchester Floor 2 Christie Building Oxford Road Manchester M13 9PL Email: medicaldevices@manchester.ac.uk	***Lead R&D Trust contact:*** Name: NIHR Clinical Research Network Greater Manchester Contact: Ms Aleksandra Metryka Address: Citylabs 1.0 Nelson Street Manchester M13 9NQ Email: researchsupport.crngm@nihr.ac.uk

2) PLAIN ENGLISH SUMMARY

Background

Online consultations allow patients to ask for help from their GP practice by completing a form on the internet. They have been available in most English GP practices since May 2020.

GP practices can receive lots of completed online consultation forms at the same time, which means it can be difficult for them to know which patients need urgent or emergency help. This can lead to delays in patients getting the care they need.

We want to test if computers trained to spot urgent and emergency forms (Artificial Intelligence or 'AI') can reduce these delays. We also want to know if AI works in the same way for all patients and whether it is good value for money.

What will we do?

We will study an AI system that is already used in NHS GP practices. We will give it to 20 GP practices not currently using it. We will measure the delays for patients receiving urgent and emergency help for 12 months before and after they start using the AI. We will compare this to 20 other GP practices that will not use the AI. We will also measure whether the AI affects staff workload and whether it works in the same way for patients from different backgrounds.

What difference will we make?

If the AI reduces care delays, patients who need urgent and emergency help will receive it sooner. We will help the NHS and companies that make online consultation systems decide whether they should use AI. We will help members of the public and GP practices understand what AI is and how they can use it to benefit both patients and staff.

3) SCIENTIFIC SUMMARY

Background

Online consultations allow patients to contact their GP practice about their health problems using an online form. Currently, 94% of GP practices in England use online consultations. PATCHS is an online consultation system launched in 2020 by commercial company Spectra Analytics. Approximately 1000 (~20%) GP practices in England currently use PATCHS.

A risk of online consultations is that patients submit forms describing medical emergencies that are not recognised quickly enough by their GP practice. To address this, Spectra Analytics developed artificial intelligence triage (AI Triage) within PATCHS to alert patients and GP practice staff when a patient describes a health problem that may suggest they require urgent or emergency treatment.

PATCHS AI Triage is a Class I (low risk) medical device and has been registered with the MHRA since October 2021. It has NHS approval for use in clinical practice and meets NHS DCB0129 safety standards. The intended purpose of PATCHS AI Triage is to assist patients and GP practice staff in making triage decisions, not to replace human judgment. During this project, PATCHS AI Triage will continue to be used within the scope of its intended purpose.

AI Triage is an optional feature of PATCHS and is currently available on request. GP practices must undertake specific training to have it enabled. Approximately 200 (20%) GP practices using PATCHS (20%) currently have AI Triage enabled – the remaining practices use PATCHS without AI Triage. Spectra Analytics are satisfied with the performance and safety of PATCHS AI Triage and plan to offer it to the

remainder of practices using PATCHS without AI Triage imminently. This presents a unique opportunity to evaluate the use of an AI system in the NHS in a controlled way to generate much-needed high-quality research evidence. To do this, we (The University of Manchester; UoM research team) have partnered with Spectra Analytics.

Methods

There are two parts to this study: an interrupted time series analysis and a quantitative process evaluation. A related qualitative process evaluation is described in a separate protocol (IRAS ID: 335429). GP practices using PATCHS without AI Triage for at least 12 months will be eligible. Practices will be randomised to either intervention (AI Triage now) or control (AI Triage later) groups using an approach based on a Zelen design. We will aim to recruit a minimum of 20 intervention and 20 control GP practices to achieve a sample size of at least 2928 urgent and emergency (combined) online consultations across both intervention and control GP practices in the intervention period. Intervention practices will be contacted by Spectra Analytics using their normal process for enabling AI Triage and will use AI Triage for the 12-month intervention period. Control GP practices will not be contacted by Spectra Analytics until the end of the 12-month intervention period – at which point they will be contacted in the same way. The rationale for this approach is that AI Triage is a selling point of the PATCHS system so if control GP practices are contacted, they may become disappointed and disengage from using PATCHS altogether ('resentful demoralisation'). Any GP practice using PATCHS without AI Triage can still request to use AI Triage at any point during the study including control practices and those outside the study.

The primary outcome measure for the interrupted time series analysis will compare the proportion of delays in completing urgent and emergency online consultations in intervention versus control GP practices. The quantitative process evaluation will measure AI Triage implementation, uptake, and accuracy. Anonymised data from PATCHS will be shared by Spectra Analytics with UoM for independent analysis. When patients and GP practices use PATCHS they are informed their anonymised data may be shared with UoM for research purposes. Patients can opt out of sharing data with UoM at any time using a toggle button in the system without affecting their ability to continue using PATCHS.

Anticipated benefits

If AI Triage is effective, patients in recruited GP practices will experience fewer delays in receiving urgent and emergency care, and this project will provide evidence for the wider adoption of AI Triage and AI interventions in general in the NHS. Regardless of whether AI Triage is effective, evidence generated from this project will be used to create help guides and toolkits on how to use AI Triage safely and effectively.

4) SUMMARY OF MAIN ETHICAL, LEGAL, OR MANAGEMENT ISSUES

Issue 1: Randomisation

Description: Recruited GP practices will be randomised to either intervention or control groups.

Mitigations: The intervention is a UKCA-marked medical device already used in routine clinical practice and will be used within the scope of its intended purpose. As in routine care, GP practices not using the intervention (both control practices and those outside the study) will be able to request to use the intervention at any point. All control GP practices will be offered the intervention at the end of the study.

Issue 2: Zelen design

Description: Control GP practices will not be contacted during the study. The rationale is that the intervention (AI Triage) is a selling point of the PATCHS system. If control GP practices are contacted, they

may become disappointed and disengage from using PATCHS altogether ('resentful demoralisation'), which would threaten study validity.

Mitigations: Control GP practices will continue using the standard PATCHS system and will be contacted at the end of the study to be offered the intervention. They can request to use the intervention at any point during the study. When intervention and control GP practices start using PATCHS without AI and staff create user accounts, they are informed their anonymised data may be shared with UoM for research purposes.

Issue 3: Anonymised data sharing

Description: When patients or their carers and GP practices use PATCHS they are informed their anonymised data may be shared with UoM for research purposes.

Mitigations: Patients or their carers can opt out of sharing their data with UoM at any point using a toggle button in the PATCHS system without affecting their ability to continue to use PATCHS to access GP services. They are clearly informed how to do this when they use PATCHS. A formal risk assessment based on guidance from the Information Commissioner's Office suggests the data are effectively anonymised.

Issue 4: Declaration of interests

Description: Dr Brown is a part-time employee of Spectra Analytics as Chief Medical Officer and is a shareholder in the company. Spectra Analytics develop the AI Triage intervention.

Mitigations: 1) Co-investigators have no conflict of interest and will hold the Chief Investigator to account to ensure that the research is conducted rigorously. 2) We will pre-register our study protocol in advance of undertaking the research and making all details openly available online to prevent outcome switching and promote independent evaluation. Any out-of-protocol analyses will be reported as such in publications. 2) We will appoint an independent Study Steering Committee to provide oversight and ensure the project is conducted rigorously. 3) We will declare all interests in study protocols, reports, and publications. 4) We will make analysis code available when findings are published. 5) We will disseminate our findings regardless of study outcome including submitting papers reporting negative results for publication in peer-reviewed journals.

Issue 5: Potential recruitment challenges

Description: The intervention is currently used in routine clinical practice. Any GP practice not using the intervention can request to use the intervention before and during the study including those from the control group and from outside the study. This reduces the pool of available GP practices to recruit from.

Mitigations: We have agreed progression criteria with the study funder. These include potential mitigation strategies that could be explored in consultation with them, the Study Steering Committee, and NHS Research Ethics Committee and Health Research Authority (if appropriate), if recruitment progress is below target.

5) BACKGROUND

Online consultations and artificial intelligence (AI) triage in primary care

GP practices in England deliver over 30 million patient appointments per month (1). A proportion are for urgent medical conditions that require treatment within 24-48 hours such as infections requiring antibiotics (18%) (2). A smaller proportion are for medical emergencies which require more immediate treatment, including from emergency services, like heart attacks. Delays in urgent or emergency treatment in primary care can lead to patient harm, including hospital admission and death: in a study of

over 300,000 urinary tract infections in elderly patients, 13.4% experienced a delay in treatment, which was associated with a higher chance of hospital admission and sepsis (odds ratio 7.12) (3). Further examples include asthma exacerbations (4), cancer (5), and pulmonary embolism (6). With continued growth in demand for NHS services and capacity remaining the same (and potentially diminishing due to staff leaving the profession), a key challenge in primary care is to identify which patients require urgent or emergency help, and which do not.

Since 2020, most GP practices have moved to online consultations, which enable patients to request help from their healthcare teams by submitting forms over the internet. All English GP practices have been mandated to provide online consultations since April 2020 (7). Their adoption has been further catalysed by the COVID-19 pandemic, and they have been available in 85% of GP practices since May 2020 (8) with latest figures from an NHS England Freedom of Information request suggesting they are available in 96% of all GP practices. In GP practices using online consultations, it is estimated they accounted for 72% of all patient requests for appointments in 2021 (9).

Although online consultations offer many benefits including patient convenience and improved access, they have the potential to exacerbate delays in providing urgent care. Although not intended for urgent or emergency problems, patients have different understandings of these terms to GP practice staff (10). Unlike traditional methods of contacting the GP practice (e.g. telephone), online consultations can be submitted by patients without waiting in a queue or talking to a member of GP practice staff. GP practices can therefore receive many online consultations in short periods of time, including when they are closed, without human filtering. GP practice staff must read each online consultation one by one: urgent and emergency online consultations can be 'hidden' from view and may not be processed in an appropriate timeframe.

One potential solution is for the online consultation system itself to automatically detect and highlight urgent and emergency online consultations as soon as they are submitted ('triage') (11). This is considered 'Artificial Intelligence' (AI) because it automates activities we typically associate with human thinking such as decision-making and problem-solving (12). We recently conducted the largest and most up-to-date systematic review of empirical research up to February 2022 on real-world use of online consultations in primary care (13). Out of 63 papers studying 31 different systems from nine countries, four papers evaluated systems with AI Triage. However, only one of those papers evaluated the AI Triage element of the system (14), which focused on the accuracy of AI Triage to classify patients with COVID-19 symptoms as either urgent, emergency, non-urgent, or self-care. The impacts of AI Triage on delays in urgent and emergency primary care, staff workload, and health inequalities have not yet been studied. Despite this lack of evidence, AI Triage is already used by 5/33 (15%) of NHS online consultation systems (15). These online consultations systems are proprietary and typically do not routinely share their data with research teams; they also do not implement their AI Triage systems in a controlled way to facilitate robust evaluation. We propose to fill this evidence gap by working in collaboration with an online consultation system provider to implement AI Triage in a controlled way to evaluate its impact using an interrupted time series analysis and quantitative process evaluation. A related qualitative process evaluation study for this project is described in a separate protocol (IRAS ID: 335429).

PATCHS online consultation system

Commercial company Spectra Analytics developed an online consultation system (PATCHS, www.patchs.ai) which has been available to GP practices since 2020. It is now used by approximately 1000 GP practices across England (~20% of all GP practices) in rural, urban, and inner-city areas with both high and low levels of deprivation.

PATCHS has the same functionality as all online consultation systems in that patients access the system via their GP practice website and fill out an online form. Patients can choose different types of online

consultation e.g. health problem or administrative. When the form is submitted by the patient it enters an inbox, where GP practice staff can prioritise and respond to online consultations. Patients complete forms in PATCHS by describing their queries in unstructured free text in response to open-ended questions in the system that mimic a typical primary care consultation, as opposed to multiple choice questionnaires. Most (65%) online consultation systems allow patients to describe their queries using unstructured free text [7].

In addition to their query, patients enter their details including sex, ethnicity, home address, email address, and telephone number. Receptionists typically review incoming PATCHS online consultations first, and will deal with any queries they can, and assign those they cannot to other staff within PATCHS, including clinicians, if it requires their input. Patients are then contacted to resolve their query – either by written message or video consultation within PATCHS, telephone or by arranging an in-person appointment. GP practices can assign PATCHS online consultations to future dates to be dealt with by clinicians if deemed necessary to match staffing capacity.

When GP practice staff process patients' online consultations in PATCHS, they record various triage decisions (Appendix 1, Screenshot 1), which have been developed based on qualitative research and workshops conducted with 22 GP practice staff and 37 patients. These triage decisions include whether they believe the online consultation is either:

- **Emergency**: patient could be harmed if not resolved within 24 hours and may require immediate care from emergency services.
- **Urgent**: the patient could be harmed if not resolved within the next 48 hours.
- **Routine**: neither urgent nor emergency.

Analysis of 122,504 PATCHS online consultations submitted to 80 GP practices show that 7922 (6.5%) are urgent or emergency, 25% of which on average are not resolved in the above timelines. We define these as delays in urgent and emergency care: online consultations deemed urgent or emergency by GP practice staff where the difference between the time it was submitted by a patient or their carer in PATCHS and the time it was completed by a staff member is longer than 48 or 24 hours, respectively.

PATCHS AI Triage: development and validation

To attempt to reduce these delays in urgent and emergency care, Spectra Analytics developed an AI Triage add-on system for the standard PATCHS product (called 'PATCHS AI Triage' for this document) in project IRAS ID: 264891. PATCHS AI Triage uses natural language processing (NLP) and deep learning to analyse online consultations to predict whether they are either urgent, emergency, or routine based on the above definitions.

PATCHS AI Triage was initially developed in 2020 using 43,998 online consultations for 37,000 patients to 52 GP practices using PATCHS. The unstructured free-text written by patients in their online consultation, type of online consultation submitted (e.g. new health problem, administrative query), and patient demographics were used as inputs to the model. The manually applied 'urgent' and 'emergency' flags from GP practice staff were used as model outputs. Data were split into training (90%) and test (10%) datasets. AI model training was optimised for a high true positive rate based on findings from co-design workshops where GP practice staff and patients felt it was most important that AI Triage systems missed as few urgent or emergency online consultations as possible. The true positive rate of PATCHS AI Triage to detect urgent and emergency online consultations was 86% in the training dataset and 84% in the test dataset. Subsequently in 2021, PATCHS AI Triage was evaluated in 14 early adopter GP practices covering inner city and urban areas in the north and south of England. Using triage decisions made by GP practice

staff to define urgent and emergency online consultations, its true positive rate was 94% in 19,805 online consultations previously unseen by the AI prospectively for 9,725 patients (21% non-white). Qualitative research with 20 staff at early adopter practices showed they found PATCHS AI Triage both acceptable and useful. Work is currently underway to re-train PATCHS AI Triage to ensure its performance has been maintained over time, and scientific publications reporting its development and validation are being prepared for peer-review.

PATCHS AI Triage has been trained on routinely collected online consultations written by patients, which contain different phrasings and misspellings of symptom descriptions. It can therefore cope with heterogeneity of the language used by patients. For example, when describing shortness of breath, which should usually always be treated promptly, PATCHS AI Triage correctly classifies the following online consultations as either urgent or emergency: “I’m struggling to breathe”, “I’m gasping for air”, “struglin for breth” [sic], “cant breath” [sic], and “its dificult to breth” [sic].

PATCHS AI Triage: regulatory approvals

Both PATCHS (without AI Triage) and PATCHS AI Triage have met all regulatory requirements for use in routine NHS clinical practice. Only online consultation systems on an NHS Buying Catalogue can be used in routine NHS clinical practice (15). NHS Digital approves online consultation systems onto this framework after assessing whether they meet various safety, accessibility, and technical standards. This includes NHS DCB0129 standards (16), which describe an approach to clinical risk management in software design and development. PATCHS has been on the NHS Buying Catalogue since its inception in 2021 (17). The PATCHS AI Triage is a Class I medical device because it offers ‘triage and signposting of next steps based on filters by severity and probability of a match’, but without ‘direct diagnosis’ (18). For Class I medical devices to be used in clinical practice in the UK, they must have a UKCA mark. To obtain a UKCA mark, medical devices must undergo clinical evaluation, post-market surveillance, and registration with the MHRA. PATCHS AI Triage obtained a UKCA mark in October 2021 (MHRA registration number 8387), which means it can be used in routine clinical practice in the UK.

PATCHS AI Triage: planned roll-out

AI Triage is available to all GP practices using PATCHS as an optional feature. GP practices must request to have it enabled and undertake specific training on how to use it. Approximately 200 GP practices using PATCHS (20%) currently have AI Triage enabled – the remaining practices use PATCHS without AI Triage. Spectra Analytics are satisfied with the accuracy and clinical safety of PATCHS AI Triage, and plan to offer it to the remainder of practices currently using PATCHS without AI Triage imminently. This presents a unique opportunity to evaluate the impact of AI Triage on delays in urgent and emergency primary care, staff workload, and health inequalities, in a controlled way. This project will also help generate much-needed high-quality research evidence on AI use in the NHS and aligns with NHS initiatives such as the NHS AI Lab (19). To do this, we (The University of Manchester; UoM) have partnered with Spectra Analytics to conduct a robust study using an interrupted time series analysis and process evaluation.

6) STUDY OBJECTIVES

6.1 Primary Research Question:

1. What is the impact of AI Triage on delays in completing online consultations defined as urgent and emergency by GP practice staff at the patient-level?

6.2 Secondary Research Questions:

Interrupted Time Series Analysis

2. What is the impact of AI Triage on the total number of appointments provided by GP practices?
3. What is the impact of AI Triage on the number of online consultations submitted by patients?
4. What is the impact of AI Triage on the number of online consultations assigned to clinicians?
5. What is the impact of AI Triage on emergency department attendances and emergency hospital admissions?
6. What are the cost consequences of AI Triage for GP practices and hospitals?

Quantitative Process Evaluation

7. What is the fidelity, dose, and reach of AI Triage and online consultations?
8. What is the accuracy of AI Triage in intervention practices?
9. What is the potential and observed change in triage behaviour?

Both Analyses

10. What is the influence of AI Triage on health inequalities?

7) STUDY DESIGN & PROTOCOL

7.1 Participants

We will aim to recruit a minimum of 20 intervention and 20 control GP practices to obtain a sample size of at least 2928 urgent and emergency (combined) online consultations across both intervention and control GP practices in the intervention period from patients of all sexes, ages, and with any clinical condition.

7.2 Study Intervention and/or Procedures

GP practice recruitment
GP practices are PATCHS customers and view AI Triage as a selling point. In experimental studies of digital health technologies, participation is always driven by an interest in (the benefits of) the technology and it is not possible to blind participants to the exposure (20). We have experienced in multiple randomised studies of digital health technologies that if participants are aware they have been allocated to a control group they become disappointed and disengage (21)(22), thereby threatening the validity of the study – a phenomenon known as 'resentful demoralisation' (23).

One approach is to recruit control practices who will volunteer to defer receiving AI Triage for the 12-month intervention period to allow comparisons with intervention practices. However, GP practices that volunteer to defer the intervention may differ systematically from those not willing to defer thereby introducing bias. We will therefore use an approach based on a Zelen design (24) where GP practices are randomised to intervention or control groups, but only contacted during the study if they are in the intervention group. This therefore minimises the probability of the control group disengaging.

Intervention GP practice recruitment
Intervention GP practices will be 'onboarded' to use AI Triage by Spectra Analytics following the process they currently use in routine clinical practice: GP practices are contacted by email which includes links to associated online help articles (25), optional eLearning modules based on these articles, and a Clinical

Safety Case Report as per NHS clinical risk management standards for software (DCB0129 (16) and DCB0160 (26)).

Control GP practice recruitment
Control GP practices will not be contacted until the end of the 12-month intervention period. At this point they will be offered AI Triage following the same onboarding procedure described above for intervention GP practices. AI Triage is available to all GP practices using PATCHS, therefore any GP practice using PATCHS without AI Triage can request to use it during the study, including control practices and practices outside the study (e.g. that do not meet study inclusion criteria.)

Intervention
The intervention period will last 12 months.

Control GP practices: PATCHS without AI Triage (manual triage)
Control GP practices will continue to use PATCHS without AI Triage as described above and will manually triage patient forms. By default, patient forms are ordered in the inbox by date-time ascending i.e. with the oldest patient forms at the top of the inbox and the newest ones at the bottom. Patient forms are manually triaged by GP practice staff using the categories above, which changes their order in the inbox to aid prioritisation (Appendix 1, Screenshot 2):

- **Emergency**: adds a red flag icon to the message and sorts it to the top of the PATCHS inbox.
- **Urgent**: adds an orange flag icon to the message and sorts it to the top of the PATCHS inbox below 'Emergency' forms.
- **Routine**: neither urgent nor emergency. Grey or purple flag icons are added and the position of the message in the PATCHS inbox does not change.

Patients do not receive any additional messages if they submit an urgent or emergency form.

Intervention GP practices: PATCHS with AI Triage (automated triage)
Intervention GP practices will use PATCHS as described above with assistance from the latest version of AI Triage as an add-on feature (called Urgency and Signpost AI in the PATCHS help documentation (27)(28); currently model version 368d85c790a743928be3dd711f914fcc). AI Triage uses the unstructured free-text written by patients in their online consultation, type of online consultation submitted (e.g. new health problem, administrative query), and patient demographics (age and sex) to predict online consultations as either 'emergency', 'urgent', or 'routine'. AI Triage communicates its predictions with GP practice staff and patients in the following ways:

- **GP practice staff**: The manual triage process described above for control GP practices is automated i.e. when patients submit a form that is predicted as emergency or urgent by AI Triage, a red or orange flag icon is automatically added respectively, and the forms are sorted to the top of the PATCHS inbox (Appendix 1, Screenshot 2). The triage decisions (Appendix 1, Screenshot 1) are pre-populated for GP practice staff, which they can change if they disagree (e.g. upgrade or downgrade the urgency).
- **Patients**: If patients submit an online consultation predicted by AI Triage as emergency at any time or urgent when their GP practice is closed, they immediately receive an automated 'signposting' message advising them to contact emergency services or the NHS 111 'out of hours' service, respectively (Appendix 1, Screenshot 3). They will also be presented with supporting advice from NHS.uk if relevant articles exist on the website. If their online consultation is predicted as 'routine', the patient may be presented with self-help advice from NHS.uk if relevant articles exist on the website (Appendix 1, Screenshot 4). Patients can accept this advice and cancel their online consultation, and the GP practice will not receive the form in their PATCHS inbox but can view it in a

separate 'cancelled' folder if they wish. Alternatively, patients can ignore the advice and continue to submit their online consultation to their GP practice at their own risk.

7.3 End of study

The study will end when the final intervention GP practice has used AI Triage for 12 months. Intervention GP practices will be able to continue using the AI Triage intervention after the study if they choose. All control GP practices will be offered the AI Triage intervention after the study has ended.

8) STUDY PARTICIPANTS

8.1 Inclusion Criteria:

GP practices

GP practices must meet both the following criteria to be eligible for the study:

- Currently actively using PATCHS without AI Triage
- Actively used PATCHS without AI Triage for at least 12 months

Patients

To minimise bias and generate real-world evidence, we will include data from all patients who use PATCHS in both intervention and control GP practices that have not opted out of sharing their anonymised data for research purposes with UoM. Patients can only use PATCHS if they are at least 16 years old, though carers can use PATCHS on behalf of patients under the age of 16 if they have guardianship or parental responsibility (as manually verified by the GP practice in PATCHS (29)).

Data

All data inputted by patients and GP practice staff will be included. There are no minimum requirements in terms of data quality or missingness. There are minimum character input requirements for patients before they can submit an online consultation in PATCHS.

8.2 Exclusion Criteria:

GP practices

GP practices who do not meet all inclusion criteria above.

Patients

Patients who have opted out of sharing their anonymised data for research purposes with UoM will be excluded.

Data

Not applicable.

8.3 Recruitment:

GP practices

A list of GP practices meeting the inclusion criteria in the form of Organisation Data Service codes (freely available online from NHS Digital (30)) will be produced by Spectra Analytics. This will be shared with the UoM research team who will use publicly available datasets from the Office for National Statistics and

NHS Digital to identify practices on the list that are representative of England in terms of patient population size (40), geographic region, and deprivation (34) to maximise generalisability and ensure our research is relevant to underserved groups in clinical research.

Intervention GP practices will be contacted by Spectra Analytics as they normally would in routine clinical practice using their standard 'onboarding' process for AI Triage described above. They are contacted by email which includes links to associated online help articles (25), optional eLearning modules based on these articles, and a Clinical Safety Case Report as per NHS clinical risk management standards for software (DCB0129 (16) and DCB0160 (26)). If no response is received from the GP practice within two weeks, a follow-up email is sent. One final email is sent one to two weeks later if still no response has been received. Control GP practices will be contacted at the end of the 12-month intervention period using the same 'onboarding' process as intervention GP practices.

To become intervention GP practices, practice managers must reply to emails from Spectra Analytics sent as part of the 'onboarding' process to confirm they consent to using AI Triage. When intervention and control GP practices start using PATCHS without AI and staff create user accounts, they are informed their anonymised data may be shared with UoM for research purposes.

GP practices and staff are not under any time pressure to decide whether they want to use PATCHS without AI in their practice. Intervention GP practices (and control practices at the end of the study period) will be given approximately one month to decide whether to use AI Triage.

Patients

Patients meeting the study inclusion criteria will be identified from the PATCHS database by Spectra Analytics. Patients will not be contacted by the UoM research team or Spectra Analytics.

When patients or their carers (verified by the GP practice (29)) use PATCHS they are informed their anonymised data may be shared with UoM for research purposes (Appendix 1, Screenshot 5). They can opt out of sharing their anonymised data at any point using a toggle button in the PATCHS system without affecting their ability to continue to use PATCHS to access GP services (Appendix 1, Screenshot 6).

Patients are not under any time pressure to decide whether they want to use PATCHS. They can also choose to not use PATCHS and contact their GP practice by telephone or in-person.

8.4 Randomisation:

GP practices will be randomly allocated to intervention (AI Triage now) or control (AI Triage later) groups in blocks of 5 practices in each arm (control and intervention) using the *runiform* command in Stata 18 with a target of 20 in each group. Depending on the response rate we may increase or decrease block size, and through a combination of randomisation and algorithmic approaches (31) using appropriate statistical software (32) we will attempt to match the two groups as closely as possible on patient population size (40), Index of Multiple Deprivation (34), monthly volume of online consultations per 1000 patients, baseline levels of the primary outcome measure, and baseline prevalence of urgent and emergency online consultations. We will also explore the possibility of matching GP practices on geographic region, rurality (36), number of whole-time equivalent GPs per 1000 patients (39), mix of patient ethnicities, and levels of patient morbidity (40), though given the target sample size is relatively small this may not be possible. Where we cannot match characteristics, we will attempt to adjust for them in our statistical models.

This is a single-blind study. The UoM research team will not know which GP practices are allocated to intervention or control groups. Due to the nature of the intervention GP practices cannot be blinded. The

UoM research team will randomly allocate GP practices to "0" or "1" groups using the approach described above. The allocation sequence will be provided to Spectra Analytics who will use statistical software to randomly allocate "0" or "1" to be the intervention group. They will then approach GP practices in the intervention group to offer them AI Triage as described elsewhere. Spectra Analytics will retain the mapping key and not share it with the UoM research team until after data analysis is finalised. The interrupted time series analysis (including analysis of the primary outcome measure) will not be undertaken until the end of the intervention period. Analyses will be undertaken by the UoM research team only not Spectra Analytics.

8.5 Participants who withdraw consent or lose capacity to consent:

If a patient opts out of sharing their anonymised data for research purposes with UoM (Appendix 1, Screenshot 6), we will withdraw their data. When a patient has used PATCHS it is impractical to find out whether they subsequently lose capacity as all registered patients at the GP practices can use PATCHS. If a GP practice staff member uses PATCHS then subsequently loses capacity, we will continue to use the data they provided before losing capacity. If data have already been extracted, we will be unable to withdraw participants' data because the data will have been anonymised and we will be unable to identify the participant.

9) OUTCOME MEASURES

Primary outcome measure

Our primary outcome measure is the proportion of delays in completing urgent and emergency online consultations at the patient-level:

$$(\text{delayed urgent online consultations} + \text{delayed emergency online consultations}) \div (\text{total urgent patient online consultations} + \text{total emergency online consultations})$$

It was chosen because patients and clinicians we consulted during our PPI work felt it was the most clinically important outcome measure. Delays will be measured by the difference between the date-time that an urgent or emergency online consultation was submitted by a patient in PATCHS and the date-time the consultation was completed in PATCHS by GP practice staff. Based on our PPI work, we will consider an urgent online consultation delayed if it is not completed within 48 hours and an emergency online consultation delayed if it is not completed within 24 hours. Online consultations that have not yet been completed but were submitted by patients more than 48 hours previously at the point of data extraction will be included. In intervention practices we define online consultations as urgent or emergency where a GP practice staff member has applied a triage decision as either 'urgent' or 'emergency', or where AI Triage predicts an online consultation as 'urgent' or 'emergency' which is unchanged by staff. We will assume that patients who receive a signpost message and cancel their online consultation will have sought care from other services and not experienced a care delay. In control practices we define online consultations as urgent and emergency only if GP practice staff apply a triage decision as either 'urgent' or 'emergency'.

Secondary outcome measures

Interrupted Time Series Analysis

Patient-level – binary

- Proportion of delayed emergency online consultations at the patient-level (disaggregated primary outcome measure; denominator=number of emergency online consultations)

- Proportion of delayed urgent online consultations at the patient-level (disaggregated primary outcome measure; denominator=number of urgent online consultations)
- Proportion of online consultations cancelled by patients at the patient-level (denominator=number of online consultations)
- Proportion of online consultations assigned to clinicians (denominator=total number of online consultations)

Patient-level – continuous

- Absolute time to completion for urgent and emergency online consultations (combined and separate)
- Absolute time to completion for routine online consultations
- Absolute time to first staff user action in online consultation system for urgent and emergency online consultations (combined and separate)
- Absolute time to first staff user action in online consultation system for routine online consultations

GP practice-level – counts (workload measures)

- Proportion of total appointments provided by GP practices (denominator=GP practice population size (1))
- Proportion of online consultations submitted by patients (denominator= GP practice population size)
- Proportion of patients with emergency department attendances (denominator=GP practice population size (37))
- Proportion of online consultations with emergency department attendances (denominator=total number of online consultations (37))
- Proportion of patients with emergency hospital admissions (denominator=GP practice population size (37))
- Proportion of online consultations with emergency hospital admissions (denominator=total number of online consultations (37))

GP practice-level – continuous (cost measures)

- Cost of clinicians processing online consultations within the GP practice per 1000 patients
- Cost of emergency department attendances and emergency hospital admissions per 1000 patients

Quantitative process evaluation

For readability, these are described in more detail in the statistical analysis section.

- Fidelity, dose, and reach of AI Triage and online consultations
- AI Triage accuracy, true positive rate, true negative rate, positive predictive value, and negative predictive value for predicting urgent and emergency online consultations (combined and separate) in intervention GP practices
- Potential triage behaviour change in control GP practices
- Observed triage behaviour change in intervention and control GP practices

10) DATA COLLECTION, SOURCE DATA, AND CONFIDENTIALITY

Patient-level data

We will use anonymised routinely collected PATCHS data held by Spectra Analytics. Data are collected automatically each time both patients and staff interact with PATCHS. This includes data collected 'passively', including when users visit certain pages within the system, in addition to data collected intentionally inputted into the system such as triage decisions made by staff or when patients enter their

age, sex, and ethnicity. Each patient will be assigned a randomly generated identification number in the PATCHS database before sharing with the UoM research team. The mapping key will then be deleted using hard drive eraser software to anonymise the data. The following anonymised patient-level data will be shared by Spectra Analytics with the UoM research team:

- Patient randomly generated identification number
- Patient year of birth
- Patient sex
- Patient ethnicity according to methods mandated for NHS organisations and response codes set out in the NHS data dictionary (33)
- GP practice Organisation Data Service code (30)
- Index of Multiple Deprivation (34)
- Date-time online consultation was submitted by the patient
- Date-time online consultation was completed by GP practice staff
- Whether online consultation was submitted by patient or someone else (carer or staff member)
- Type of online consultation chosen by patient e.g. health problem or administrative request
- Whether the patient submitted their online consultation in a non-English language
- Triage predictions made by AI Triage regarding the online consultation
- Triage decisions made by GP practice staff regarding the online consultation
- Signposting messages presented to patient
- Whether the patient or carer cancelled the online consultation
- How the online consultation was processed by GP practice staff e.g. if it was assigned to other staff users in PATCHS, if messages were sent to the patient in response
- Role of staff who processed the online consultation e.g. GP, nurse, receptionist

A formal risk assessment by Spectra Analytics based on guidance from the Information Commissioner's Office (35) concludes there is a low risk of individuals being identified from the data through 'singling out', 'data linkages', or 'inference', and the data are therefore effectively anonymised (Appendix 2).

GP practice-level data

In addition to PATCHS data, we will also use the following GP practice-level data that we will link to PATCHS data via the GP practice Organisation Data Service (30) code:

- Built Up Areas from the Office for National Statistics (freely available online) (36)
- Emergency Care Dataset from NHS Digital Data Access Request Service (application required and cost associated) (37)
- National General Practice Profiles from Public Health England (freely available online) (38)
- GP Practice Workforce from NHS Digital (freely available online) (39)
- Total appointment counts from NHS Digital (freely available online) (1)
- Index of Multiple Deprivation from Office for National Statistics (freely available online) (34)
- Quality and Outcomes Framework performance from NHS Digital (freely available online) (40)

Data transfer and storage

Data sharing agreements will be signed between UoM and Spectra Analytics, and UoM and NHS Digital. Data will be transferred from Spectra Analytics and NHS Digital to UoM as password-protected comma separated value files using industry standard Hypertext Transfer Protocol Secure over a Transport Layer Security connection encrypted both in transit and at rest.

Once transferred to UoM, data will be stored on secure UoM Research Data Storage servers (41). These servers are only accessible by specific UoM users on the UoM network via multi-factor authentication. Only members of the UoM research team will be given permissions to access the data.

Data for the interrupted time series analysis (including the primary and secondary outcome measures – section 9) will not be transferred from Spectra Analytics and NHS Digital to UoM until after the intervention period has finished. Data for the quantitative process evaluation (section 9) will be transferred monthly from Spectra Analytics to UoM after GP practice recruitment has finished.

11) STATISTICAL CONSIDERATIONS

11.1 Statistical Analysis

Interrupted Time Series Analysis

Design

Each intervention GP practice will be matched one-to-one with a control practice according to the characteristics described above. At least 12 months of outcome data before, and following, the index date will be available for analysis.

Descriptive statistics

Characteristics of intervention and control practices and their patients, and those that crossed over or declined the intervention, in terms of variables used for matching and all outcome measures, will be compared descriptively and inferentially in the pre-intervention period using t-tests for continuous and Fisher's exact tests for categorical variables. Outcome measures will be plotted as monthly time series of events across the pre- and intervention periods.

Modelling

To analyse binary patient-level outcomes, including our primary outcome measure (proportion of delayed urgent and emergency online consultations; proportion of delayed urgent online consultations; proportion of delayed emergency online consultations; proportion of online consultations cancelled by patients; proportion of online consultations assigned to clinicians), we will use mixed-effects logistic regression models with appropriate offset terms (number of urgent and emergency online consultations). We will initially analyse data as a time series with a minimum of 12 time points (months) pre-intervention and 12 intervention. The main exposure of interest will be membership to the intervention or control group modelled as binary (0/1). Models will be adjusted for practice characteristics that we have been unable to match during randomisation described elsewhere (e.g. patient population size). Where possible, we will also adjust models for the following practice characteristics that may influence the primary outcome measure: length of time using PATCHS without AI Triage, and other features enabled in PATCHS (the system has several configurable features such as different AI modules). Time will be modelled as continuous (1 to 24) to account for trends in the pre-intervention period. We will also attempt to include month as a categorical variable to account for seasonality. The main parameter of interest will be the interaction term between practice group (intervention vs control) and study period (pre- vs intervention); post-estimation commands will be used to obtain estimates for each study period by practice group.

To analyse continuous patient-level outcome measures (absolute time to completion or staff user action for urgent and emergency online consultations – combined and separate; absolute time to completion or staff user action for routine online consultations) we will use mixed-effects linear regression models with all other aspects of the analysis remaining the same.

If multiple online consultations are submitted by the same patient, we will randomly sample one per patient for analysis. Multiple online consultations from the same patient are expected to be infrequent, though if this approach adversely affects reaching our sample size target we will instead include all online consultations with a patient-level variable in our models.

To analyse count GP practice-level outcome measures (proportion of total appointments provided by GP practices; proportion of online consultations submitted by patients; proportion of patients with emergency department attendances; proportion of online consultations with emergency department attendances; proportion of patients with emergency hospital admissions; proportion of online consultations with emergency hospital admissions) we will use negative binomial regression models with appropriate offset terms for the denominators. For example, GP practice population size (40) or number of whole-time equivalent GPs per 1000 patients (39).

Cost consequences of clinician, emergency department, and hospital admission contacts will be estimated by multiplying counts by the relevant weighted average unit cost (42). The timeframe for the cost consequences analyses will be limited to the 12-month period of implementation in the study. Impacts on health and wellbeing outcomes will not be evaluated due to the variety of different health-related reasons patients may present with, resources required to collect health-related quality of life measures via primary data collection, and the retrospective nature of the pre-intervention period evaluation. PATCHS is delivered across both comparator and treated practices so the cost of PATCHS itself will not feature in the economic analyses. The costs of the intervention include the AI element of PATCHS and training of this element to practices (funded centrally). We will explore the identification of these costs and, where feasible, the apportioning of these costs to practices. To analyse continuous GP practice-level outcome measures (cost of processing online consultations within the GP practice per 1000 patients; cost of emergency department attendances and emergency hospital admissions per 1000 patients) we will use linear regression models with all other aspects of the analysis remaining the same.

Sensitivity analyses

Limitations with the above approach include that: staff triage decisions could be applied by non-clinicians which could be systematically different to those applied by clinicians (43); we assume that patients who receive a signpost message and cancel their online consultation have not experienced a care delay; and the triage decisions made by GP practice staff may be highly variable. We will therefore conduct sensitivity analyses where: we restrict triage decisions to those only made by clinicians; we exclude patients who receive a signpost message and cancel their online consultation; we define online consultations as urgent and emergency in intervention and control practices if they are predicted as either ‘urgent’ or ‘emergency’ by AI Triage. We will also conduct sensitivity analyses where we sub-sample practices and patients with similar: baseline levels of the outcome measure, monthly volume of online consultations per 1000 patients, prevalence of urgent and emergency online consultations, agreement with AI Triage predictions, contributions to AI Triage model training, and other variables matched during randomisation and model adjustment where appropriate. Further sensitivity analyses may be undertaken based on findings from the process evaluation (44), for example, we may find a cohort of practices that did not engage with the AI Triage training, and we may test the hypothesis that AI Triage was less effective at reducing delays in urgent and emergency online consultations for them. We will also explore using an alternative analysis approach, interacting practice group with time and period to estimate the adjusted intervention time series. For example, a difference-in-difference method pooling the outcome (and the offset) in the pre- and intervention periods.

Health inequalities

We will use the pre-intervention period to assess for inequalities in the influence of patient characteristics on experiencing urgent and emergency care delays in both the pre- and intervention periods. We will use

patient age, sex, ethnicity, socioeconomic deprivation, and non-English language usage as predictors in our regression model to compare the probability of experiencing a care delay (45). We will also explore the possibility of adding data on patient multimorbidity and frailty if available. If there is a main effect in the outcome analyses, we will also use sub-group models for appropriate interaction terms in the main models to explore the effectiveness of the intervention on population strata of interest described above. We appreciate power will be lower for these investigations so these approaches will be exploratory, and this approach assumes that there is a main effect for the primary and / or secondary outcomes.

Quantitative Process Evaluation

Fidelity, dose, and reach of AI Triage and online consultations

'Fidelity' is whether the intervention is delivered as intended (44). We will evaluate fidelity through counts of how many practices have AI Triage switched on and the number and proportions of staff in each practice that access online learning materials.

'Dose' is how much intervention is delivered (44). We will evaluate dose through descriptive analyses of counts of online consultation usage submissions in both intervention and control practices in terms of overall numbers and specific types of online consultations (for example, health problems or administrative requests). In intervention practices, we will undertake descriptive analyses of counts of predictions made by AI Triage (urgent, emergency, or routine).

'Reach' is the extent a target audience encounters the intervention (44). We will evaluate reach in both intervention and control practices through descriptive analyses of counts of patients that submit online consultations, how many each they submit, counts of staff that process them (including comparisons between clinical and non-clinical staff), and how many each they process. In intervention practices we will include separate counts of staff that process online consultations that have been predicted by AI Triage as urgent or emergency, and how many patients are presented with signpost messages.

AI Triage accuracy in intervention GP practices

We will calculate the overall accuracy (proportion of 'correct' predictions), true positive rate, true negative rate, positive predictive value, and negative predictive value of AI Triage in intervention practices. Our primary measure of accuracy will be for urgent and emergency online consultations combined; secondary measures will assess urgent and emergency online consultations separately. There is no gold standard test to decide the 'correct' triage decision for online consultations written by patients in their own words. We will therefore use the triage decisions made by GP practice staff when processing online consultations and consider their triage decision as 'correct'. Where AI Triage predicts an online consultation as 'urgent' or 'emergency' we will consider it a:

- **True positive** if the triage decision is not changed by staff, or if staff change the triage decision and the highest triage decision applied by staff is 'urgent' or 'emergency'.
- **False positive** if staff change the triage decision and the highest triage decision applied by staff is 'routine'.

Where AI Triage predicts an online consultation as 'routine', we will consider it a:

- **True negative** if the triage decision is not changed by staff, or if staff change the triage decision and the highest triage decision applied by staff is 'routine'.
- **False negative** if staff change the triage decision and the highest triage decision applied by staff is 'urgent' or 'emergency'.

We will quantify cases where emergency online consultations have been predicted as 'routine' as these represent the highest risk misclassifications. The Spectra Analytics Clinical Safety team will investigate these and other patient safety incidents reported by GP practices as per their internal processes to comply with MHRA and NHS DCB0129 standards (16). We will review their findings to understand if there are patient groups or online consultation topics that are at higher risk of misclassification by AI Triage.

Potential triage behaviour change in control GP practices

We will obtain AI Triage predictions (urgent and emergency – both separate and combined) for control practices for each online consultation submitted during the intervention period and compare them to the actual triage decisions made by GP practice staff using the same definitions of true/false positives/negatives above. These predictions will differ from those in intervention GP practices because they will not have been presented to GP practice staff whilst they processed the online consultations. It identifies a group of patients for whom potentially different triage decisions would have been made if AI Triage had been enabled in those practices. We will then estimate the potential impact of these different triage decisions.

Observed triage behaviour change in intervention and control GP practices

To evaluate observed changes in triage behaviour, we will compare monthly and weekly time series counts of urgent and emergency triage decisions (both separate and combined) made by intervention and control GP practices. In intervention practices we will consider a triage decision as urgent or emergency if the highest triage applied by staff is 'urgent' or 'emergency' or an AI Triage prediction of urgent or emergency is left unchanged (same definition as 'true positives' above). In control practices, we will consider a triage decision as urgent or emergency if the highest triage applied by staff is 'urgent' or 'emergency'. We will also measure counts of patients cancelling their online consultations, and map whether patients cancel their request or receive input from a clinician following a signpost message and/or emergency prediction.

Sensitivity analyses

Similar to the interrupted time series analysis, limitations with the above approach include that: staff triage decisions could be applied by non-clinicians; the true urgency of an online consultation may only be apparent when further information has been obtained from the patient (e.g. over the telephone); we assume AI Triage predictions for patients who receive a signpost message and cancel their online consultation is correct. We will therefore conduct sensitivity analyses where: we restrict triage decisions to those only made by clinicians; the final triage decision when the online consultation is resolved is used; patients who receive a signpost message and cancel their online consultation are excluded. As mentioned above, multiple online consultations from the same patient are expected to be infrequent, though will be assessed by using a patient-level variable in models and by re-running analyses after randomly sampling one online consultation per patient.

Health inequalities

We will assess the potential influence of AI Triage on health inequalities across different age groups, sexes, ethnic backgrounds, Index of Multiple Deprivation quintiles, and non-English language usage. In dose and reach metrics we will compare the proportion of each sub-group in the population using online consultations in intervention and control practices in the pre- and intervention periods (44), and to the characteristics of the wider practice populations in National General Practice Profiles (38).

Each accuracy metric will be tested for differences in performance between patient sub-groups. We will also undertake a failure case analysis to explore factors why AI Triage may have predicted incorrectly (46)

by comparing the characteristics of patients who have submitted online consultations classified as false positive and negatives to those of true positive and negative predictions ('error auditing') (47). Additional factors to test will include those related to the online consultation (including type of online consultation, time and day of submission), GP practice staff who applied the triage decision (including role, experience using PATCHS), and GP practice (including size, geographic location, experience using PATCHS).

In potential behaviour change analyses, we will find patients for whom the AI Triage prediction differs from the actual decision made by GP staff (potential false positives and negatives) and test for differences in characteristics of the patient, online consultation, GP practice staff, and GP practice using the error auditing approach described above (47).

Allocation non-adherence

In a traditional Zelen design, data are analysed on an intention-to-allocate basis rather than whether or not the intervention was actually received (24). This approach is practical when the intervention is not widely available and when participants can provide immediate consent, which is not the case in this study. Firstly, AI Triage has been available to all GP practices using PATCHS since October 2021, therefore control GP practices can start using AI Triage at any time during the study. Because it is considered a selling point, GP practices may also receive communications encouraging them to use AI Triage outside the study e.g. from their NHS commissioning organisation. Secondly, GP practices may take several weeks to reply following an invitation to use AI Triage. Thirdly, GP practices not yet using AI Triage may have explicitly chosen not to use it, meaning they may be more likely to decline an invitation to use it or drop out of the study later. These factors combine to exacerbate our anticipated recruitment challenges described above. Therefore, we will take the following approach:

- **Control GP practices that cross over to the intervention group:** If we have sufficient follow-up data we will treat them as intervention GP practices in the main analysis, otherwise we will use their data as controls up to the point they cross over. If numbers allow, we will analyse them separately in an uncontrolled interrupted time series analysis. We will conduct sensitivity analyses where we treat them control GP practices for the entire study and where we exclude them entirely from the analysis.
- **GP practices that decline to use the intervention or do not respond to recruitment communications after being allocated to the intervention group:** We will treat them as control GP practices in the main analysis. We will conduct sensitivity analyses where we treat them as intervention GP practices and where we exclude them entirely from the analysis.
- **Intervention GP practices that cross over to the control group (i.e. that stop using AI Triage) or stop using PATCHS altogether:** We will include them in the main analysis in the intervention group. We will undertake sensitivity analyses where we use their data as intervention GP practices up to the point they cross over and where we exclude them entirely from the analysis.

11.2 Sample Size:

Our sample size calculation is based on our primary outcome measure. Patients and clinicians we consulted during our PPI work felt an absolute reduction of 5% in the proportion of urgent and emergency online consultation delays would be meaningful. Assuming a baseline delay of 25% from our prior research and a simple before-after design, we estimate a minimum of 2928 urgent and emergency (combined) online consultations across both intervention and control GP practices during the intervention period are required to detect a minimum absolute reduction of 5% to 20% with 90% power and 5% alpha. Further assuming 50 urgent and emergency online consultations (combined) on average for a GP practice per month and a 12-month intervention period, this translates to a minimum of three intervention and three control practices. However, this does not account for between-practice variability. Consequently, we will aim to recruit 20 intervention and 20 control practices. Between-practice variability can be difficult to

predict, so we have estimated the power of this sample size for different assumptions of between-practice variance using ipdpower for Stata (48), which calculates power for mixed-effects models using simulations (Table 1). To minimise between-practice variance in our sample, we will recruit intervention and control GP practices in blocks (described in section 8.4). If we find that between-practice variance is high, we will target recruitment of practices with similar baseline characteristics for the outcome measure in subsequent stages.

Table 1: Study power simulations

Between-practice variance	Power (%)	95% confidence interval (%)
0	100	96.4-100.0
0.1	72	62.1-80.5
0.2	53	42.8-63.1
0.825	21	13.5-30.3

12) DATA AND INTERVENTION ACCESS POST-STUDY

Data will be stored on secure UoM Research Data Storage servers for 5 years after publication of our results per UoM's Record Retention Schedule (49). Only the UoM research team will have access to these data during this time. Following consideration of all legal and ethical perspectives, interests and contractual stipulations of third-party funders and other stakeholders, as well as aspects of confidentiality and security, the data will then be deleted using hard drive eraser software. We will document any deletion and destruction of data and make it accessible for possible future audit.

Following the study intervention GP practices can continue to use AI Triage after the study if they wish, and control GP practices will be offered AI Triage. PATCHS (without AI Triage) and PATCHS AI Triage are available for free to all NHS GP practices via the NHS Buying Catalogue (15). All online consultation software solutions on this framework are centrally funded and do not incur any direct cost to individual GP practices or NHS commissioning organisations. In addition, Spectra Analytics will explore the feasibility of creating an application programming interface that enables external organisations to receive predictions from the AI Triage model to validate its outputs using their own datasets.

13) MONITORING AND QUALITY ASSURANCE

The study will be subject to the audit and monitoring regime of The University of Manchester. An independent Study Steering Committee will be appointed to ensure that the project is conducted to rigorous standards. The committee will include academics with appropriate clinical and methodological expertise, and one member of the public. The committee will meet at least three times during the project. As a UKCA-marked medical device already used in routine clinical practice, any safety issues associated with the AI Triage intervention will be reported and managed in the usual way by Spectra Analytics as per UKCA and NHS DCB0129 standards.

The project will be monitored against progression criteria (Table 2) covering recruitment, protocol non-adherence, and outcome data using a traffic light system (9): red/stop (intractable issues that cannot be remedied), amber/amend (remediable issues) or green/go (no concerning issues). Progression criteria will be judged at four months (halfway through) into our planned recruitment period. We anticipate that recruitment will be most problematic. Progress will be reported to the Study Steering Committee. If red or amber criteria are met for recruitment, the following mitigation strategies will be considered and

discussed with the: 1) Study Steering Committee, 2) funder, and 3) NHS Research Ethics Committee and Health Research Authority (if appropriate): approaching more GP practices to aid recruitment; working with industry partners to offer control practices an incentive for deferring that will not impact study outcomes; conduct a prospective uncontrolled interrupted time series analysis; conduct a retrospective controlled or uncontrolled interrupted time series analysis using data from GP practices that previously met the inclusion criteria; or relaxing inclusion criteria.

Table 2: Progression criteria judged at four months into recruitment

Study aspect	Red	Amber	Green
Recruitment	• <10% of total intervention patients recruited* • <10% of total intervention GP practices recruited*	• 10-50% of total intervention or control patients recruited • 10-50% of total intervention or control GP practices recruited	• >50% of total intervention and control patients recruited • >50% of total intervention and control GP practices recruited
Protocol non-adherence	• >80% of patients have withdrawn consent to share data • >80% of intervention GP practices no longer using the intervention*	• 40-80% of patients have withdrawn consent to share data • 40-80% of intervention GP practices no longer using the intervention • 40-80% of control GP practices are using the intervention	• <40% of patients have withdrawn consent to share data • <40% of intervention GP practices no longer using the intervention • <40% of control GP practices are now using the intervention
Outcome data	• Outcome data available for <40% of recruited GP practices • Outcome data available for <40% of recruited patients	• Outcome data available for 40-80% of recruited GP practices • Outcome data available for 40-80% of recruited patients	• Outcome data available for >80% of recruited GP practices • Outcome data available for >80% of recruited patients

*Control patients nor control practices are not included because one mitigation strategy could be to conduct an uncontrolled time series analysis.

14) PEER REVIEW

This protocol has been peer-reviewed as part of the application for funding to NIHR HSDR programme.

15) PATIENT AND PUBLIC INVOLVEMENT (PPI) AND STAKEHOLDER ENGAGEMENT

PPI input to this plan

We conducted interviews with seven patients and held workshops with 10 GP practice staff to gather feedback on our study plans. PPI co-applicants on the funding grant reviewed and edited the grant application including the plain English summary. Changes to our approach based on our PPI work included adding objectives to evaluate the impact of AI Triage on health inequalities and health system resources, selection of the primary outcome measure, and the magnitude of its reduction that would be clinically meaningful.

PPI input during the study

We will recruit a PPI group of six people who will be members of the public who use GP services. The group will meet up to 10 times during the project every 2-3 months. Meetings will be face-to-face or via video conference, though members will also be able to contribute by other means (e.g. email, phone, or post), depending on individual circumstances. In addition to the PPI group, we will also recruit up to six members of staff (GPs, practice managers etc) from GP practices to be part of a separate stakeholder group. This group will meet in a similar way, at a similar frequency. Both groups will contribute to data analysis and project outputs.

16) ETHICAL AND REGULATORY CONSIDERATIONS

16.1 Approvals

NHS Research Ethics Committee and Health Research Authority approval will be obtained before commencing research. Submission to the MHRA is not required for this study because the intervention is a UKCA-marked medical device already used in routine clinical practice and will be used within the scope of its intended purpose. The study will be conducted in full conformance with all relevant legal requirements and the principles of the Declaration of Helsinki, Good Clinical Practice and the UK Policy Framework for Health and Social Care Research 2017.

16.2 Risks

Risks to participants

This study uses routinely collected data to evaluate a UKCA-marked medical device already used in routine clinical practice. Consequently, risks for participants are minimal. The only risk we have identified related to study procedures is the risk of identifying participants from the research data. A formal risk assessment (Appendix 2) based on guidance from the Information Commissioner's Office (35) has been undertaken as described above. It concludes the data are effectively anonymised and the risk of a motivated intruder identifying a participant is low. Security arrangements for protecting the data include: 1) Using industry standard practices for transferring data as password-protected files using Hypertext Transfer Protocol Secure over a Transport Layer Security connection encrypted both in transit and at rest, 2) Storing the data on secure UoM Research Data Storage servers.

Risks to researchers

Researchers will analyse routinely collected data only. We have identified no additional risks to them in conducting the study.

16.3 Benefits

There may be benefits associated with using AI Triage that impact delays in urgent and emergency primary care, staff workload, and health inequalities. GP practices in the study (both intervention and control) will be offered the intervention before GP practices outside the study; they and their patients may therefore experience these benefits sooner.

17) STATEMENT OF INDEMNITY

The University has insurance available regarding research involving human subjects that provides cover for legal liabilities arising from its actions or those of its staff or supervised students. The University also has insurance available that provides compensation for non-negligent harm to research subjects occasioned in circumstances that are under the control of the University.

18) FUNDING AND RESOURCES

This project is funded by the NIHR HSDR programme, reference number NIHR153121.

19) PUBLICATION POLICY

This research protocol will be publicly registered on the ISRCTN (International Standard Randomised Controlled Trial Number) registry (51). Findings will be published in open-access peer-reviewed scientific journals. Analysis code will be made available. We will also produce short evidence summaries communicating key findings in an accessible way, which will be hosted on publicly available websites (e.g. www.patchs.ai) and disseminated to participating GP practices and patients by Spectra Analytics via email. GP practices will be encouraged to share these findings with their patients, for example by publishing them on their website.

20) DECLARATION OF INTERESTS

Dr Brown is a part-time employee of Spectra Analytics as Chief Medical Officer and is a shareholder in the company. Spectra Analytics develop the AI Triage intervention. Co-Investigators have no relevant interests to declare.

21) REFERENCES

1. NHS Digital [Internet]. [cited 2023 Dec 13]. Appointments in General Practice. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/appointments-in-general-practice

2. Van Staa T, Li Y, Gold N, Chadborn T, Welfare W, Palin V, et al. Comparing antibiotic prescribing between clinicians in UK primary care: an analysis in a cohort study of eight different measures of antibiotic prescribing. BMJ Qual Saf. 2022 Mar 3;bmjqs-2020-012108.

3. Gharbi M, Drysdale JH, Lishman H, Goudie R, Molokhia M, Johnson AP, et al. Antibiotic management of urinary tract infection in elderly patients in primary care and its association with bloodstream infections and all cause mortality: population based cohort study. BMJ. 2019 Feb 27;364:l525.

4. Royal College of Physicians. Why asthma still kills: the National Review ofAsthma Deaths (NRAD) Confidential Enquiry report. London: RCP; 2014.

5. Arhi CS, Burns EM, Bottle A, Bouras G, Aylin P, Ziprin P, et al. Delays in referral from primary care worsen survival for patients with colorectal cancer: a retrospective cohort study. Br J Gen Pract. 2020 Jul;70(696):e463–71.

6. Walen S, Damoiseaux RAMJ, Uil SM, van den Berg JW. Diagnostic delay of pulmonary embolism in primary and secondary care: a retrospective cohort study. Br J Gen Pract. 2016 Jun;66(647):e444–50.

7. NHS England. GP Contract. 2019 [cited 2022 May 9]. GP Contract documentation 2019/20. Available from: www.england.nhs.uk/gp/investment/gp-contract/gp-contract-documentation-2019-20

8. Bakhai M. NHSX. 2020 [cited 2022 May 9]. The use of online and video consultations during the COVID-19 pandemic - delivering the best care to patients. Available from: https://www.nhsx.nhs.uk/blogs/use-online-and-video-consultations-during-covid-19-pandemic-delivering-best-care-patients/

9. Clarke GM, Dias A, Wolters A. Access to and delivery of general practice services: a study of patients at practices using digital and online tools [Internet]. London: The Health Foundation; 2022. Available from: https://www.health.org.uk/publications/access-to-and-delivery-of-general-practice-services

10. MacKichan F, Brangan E, Wye L, Checkland K, Lasserson D, Huntley A, et al. Why do patients seek primary medical care in emergency departments? An ethnographic exploration of access to general practice. BMJ Open. 2017 May 4;7(4):e013816.

11. Berikol GB, Berikol G. Use of artificial intelligence in emergency medicine. In: Artificial Intelligence in Precision Health. Elsevier; 2020. p. 405–13.

12. Bellman R. An Introduction to Artificial Intelligence: Can Computers Think? Boyd & Fraser Publishing Company; 1978. 168 p.

13. Darley S, Coulson T, Peek N, Moschogianis S, van der Veer SN, Wong DC, et al. Understanding how the design and implementation of online consultations impact primary care quality: Systematic review of evidence with recommendations for designers, providers, and researchers (Preprint). J Med Internet Res. 2022 Feb 22;

14. Judson TJ, Odisho AY, Neinstein AB, Chao J, Williams A, Miller C, et al. Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for COVID-19. J Am Med Inform Assoc. 2020;27(6):860–6.

15. NHS Digital. Buying Catalogue. 2022 [cited 2022 May 9]. Find Buying Catalogue Solutions. Available from: https://buyingcatalogue.digital.nhs.uk/catalogue-solutions

16. NHS Digital [Internet]. [cited 2023 Dec 13]. DCB0129: Clinical Risk Management: its Application in the Manufacture of Health IT Systems. Available from: https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/dcb0129-clinical-risk-management-its-application-in-the-manufacture-of-health-it-systems

17. NHS Digital. Buying Catalogue. 2021 [cited 2022 May 9]. PATCHS Online Consultation. Available from: https://buyingcatalogue.digital.nhs.uk/catalogue-solutions/10046-006/features

18. MHRA. Guidance: Medical device stand-alone software including apps (including IVDMDs) v1.08. 2021.

19. NHS AI Lab. NHS Transformation Directorate. 2021 [cited 2022 Jul 13]. The National Strategy for AI in Health and Social Care. Available from: https://www.nhsx.nhs.uk/ai-lab/ai-lab-programmes/the-national-strategy-for-ai-in-health-and-social-care/

20. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating Digital Health Interventions: Key Questions and Approaches. Am J Prev Med. 2016;51(5):843–51.

21. Goud R, De Keizer NF, Ter Riet G, Wyatt JC, Hasman A, Hellemans IM, et al. Effect of guideline based computerised decision support on decision making of multidisciplinary teams: Cluster randomised trial in cardiac rehabilitation. BMJ Online. 2009 May 9;338(7703):1132.

22. Kraal JJ, Elske Van Den Akker-Van Marle M, Abu-Hanna A, Stut W, Peek N, Kemps HMC. Clinical and cost-effectiveness of home-based cardiac rehabilitation compared to conventional, centre-based cardiac rehabilitation: Results of the FIT@Home study. Eur J Prev Cardiol. 2017 Aug 1;24(12):1260–73.

23. Onghena P. Resentful Demoralization. In: Balakrishnan N, Colton T, Everitt B, Piegorsch W, Ruggeri F, Teugels JL, editors. Wiley StatsRef: Statistics Reference Online [Internet]. 1st ed. Wiley; 2014 [cited 2023 Aug 17]. Available from: https://onlinelibrary.wiley.com/doi/10.1002/9781118445112.stat06754

24. Simon GE, Shortreed SM, DeBar LL. Zelen design clinical trials: why, when, and how. Trials. 2021 Dec 1;22(1).

25. Patchs AI (Artificial Intelligence) – PATCHS Support [Internet]. [cited 2023 Dec 13]. Available from: https://help.patchs.ai/hc/en-gb/sections/360012017013-Patchs-AI-Artificial-Intelligence-

26. NHS Digital. DCB0160: Clinical Risk Management: its Application in the Deployment and Use of Health IT Systems - NHS Digital. NHS Digit [Internet]. 2018; Available from: https://digital.nhs.uk/data-and-information/information-standards/information-standards-and-data-collections-including-extractions/publications-and-notifications/standards-and-collections/dcb0160-clinical-risk-management-its-application-in-the-deployment-

27. Brown B. PATCHS Help Centre. 2022 [cited 2022 May 9]. Urgency AI. Available from: https://help.patchs.ai/hc/en-gb/articles/360058979713-Urgency-AI

28. Brown B. PATCHS Help Centre. 2022 [cited 2022 May 9]. Signpost AI. Available from: https://help.patchs.ai/hc/en-gb/articles/4410509916183-Signpost-AI

29. PATCHS Support [Internet]. 2023 [cited 2023 Dec 15]. How to verify carer relationship. Available from: https://help.patchs.ai/hc/en-gb/articles/1500004882962-How-to-verify-carer-relationship

30. NHS Digital [Internet]. [cited 2023 Dec 13]. About the Organisation Data Service. Available from: https://digital.nhs.uk/services/organisation-data-service/about-the-organisation-data-service

31. Berger VW, Bour LJ, Carter K, Chipman JJ, Everett CC, Heussen N, et al. A roadmap to using randomization in clinical trials. BMC Med Res Methodol. 2021 Dec;21(1):168.

32. Kontopantelis E. A Greedy Algorithm for Representative Sampling: repsample in Stata. J Stat Softw [Internet]. 2013;55(Code Snippet 1). Available from: http://www.jstatsoft.org/v55/c01/

33. NHS Digital [Internet]. [cited 2023 Dec 13]. Ethnicity. Available from: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/mental-health-services-data-set/submit-data/data-quality-of-protected-characteristics-and-other-vulnerable-groups/ethnicity

34. Office for National Statistics. Community and society. 2019 [cited 2022 Sep 30]. English indices of deprivation 2019. Available from: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019

35. Information Commissioner's Office. How do we ensure anonymisation is effective? [Internet]. Information Commissioner's Office; 2021 Oct. Available from: https://ico.org.uk/media/about-the-ico/documents/4018606/chapter-2-anonymisation-draft.pdf

36. Built Up Area to Region (December 2022) Lookup in Great Britain [Internet]. [cited 2023 Dec 13]. Available from: https://geoportal.statistics.gov.uk/search?collection=Dataset

37. NHS Digital [Internet]. [cited 2023 Dec 13]. Emergency Care Data Set (ECDS). Available from: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/emergency-care-data-set-ecds

38. Public Health England. National General Practice Profiles. Crown Copyr. 2023;1–14.

39. NHS Digital [Internet]. [cited 2023 Dec 13]. General Practice Workforce. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/general-and-personal-medical-services

40. NHS Digital [Internet]. [cited 2023 Dec 13]. Quality and Outcomes Framework. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/quality-and-outcomes-framework-achievement-prevalence-and-exceptions-data

41. University of Manchester. Research Data Storage Documentation [Internet]. Available from: https://ri.itservices.manchester.ac.uk/rds/

42. Jones K, Burns A. Costs of Health and Social Care 2021 [Internet]. Personal Social Services Research Unit, Kent, UK,; 2021. Available from: http://www.pssru.ac.uk

43. Sexton V, Atherton H, Dale J, Abel G. Clinician-led secondary triage in England's urgent care delivery: a cross-sectional study. Br J Gen Pract. 2023 Jun;73(731):e427–34.

44. Moore G, Audrey S, Barker M, Bonell C, Hardeman W, Moore L, et al. Process evaluation of complex interventions: UK Medical Research Council (MRC) guidance. 2014.

45. Kontopantelis E, Doran T, Springate DA, Buchan I, Reeves D. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. BMJ. 2015;350(4):h2750.

46. Rivera SC, Liu X, Chan A wen, Denniston AK, Calvert MJ. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. BMJ. 2020 Sep 9;m3210.

47. Oakden-Rayner L, Dunnmon J, Carneiro G, Re C. Hidden stratification causes clinically meaningful failures in machine learning for medical imaging. In: ACM CHIL 2020 - Proceedings of the 2020 ACM Conference on Health, Inference, and Learning. Association for Computing Machinery, Inc; 2020. p. 151–9.

48. Kontopantelis E, Springate DA, Parisi R, Reeves D. Simulation-Based Power Calculations for Mixed Effects Modeling: **ipdpower** in *Stata*. J Stat Softw [Internet]. 2016 [cited 2023 Jun 14];74(12). Available from: http://www.jstatsoft.org/v74/i12/

49. The University of Manchester [Internet]. [cited 2023 Dec 13]. Retention of records. Available from: https://www.manchester.ac.uk/discover/privacy-information/freedom-information/record-retention/

50. Avery KNL, Williamson PR, Gamble C, O'Connell Francischetto E, Metcalfe C, Davidson P, et al. Informing efficient randomised controlled trials: exploration of challenges in developing progression criteria for internal pilot studies. BMJ Open. 2017 Feb;7(2):e013537.

51. ISRCTN Registry [Internet]. [cited 2023 Dec 13]. Available from: https://www.isrctn.com/