STATISTICAL ANALYSIS PLAN

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) STUDY OBJECTIVES

4.1 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?

4.2 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?

5) 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

6) STATISTICAL CONSIDERATIONS

6.1 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.

6.2 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

7) 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/