FUNDED BY
NIHR | National Institute for Health and Care Research

Full title of trial	Non-randomised feasibility study of point of care natural language processing using the MiADE system to assist structured clinical documentation
Short title	Feasibility study of 'MiADE' point of care natural language processing
Version and date of protocol	Version 1.1 24/07/2023
Sponsor:	University College London (UCL)
Sponsor reference number:	157134
Funder (s):	National Institute of Health Research
IRAS Number:	322887
Clinicaltrials.gov no:	To be registered
UCL Data Protection Number:	Z6364106/2023/03/03 health research
Intervention:	Non-randomised feasibility and mixed methods study

Chief investigator:
Dr Anoop D. Shah
Rom 403, Institute of Health informatics, 222 Euston Road, London, NW1 2DA

Sponsor Representative:
Mr Pushpsen Joshi, uclh.randd@nhs.net
UCLH/UCL Joint Research Office,
4th Floor, West
250 Euston Road
London
NW1 2PG

PROTOCOL VERSION HISTORY

Version Stage	Versions Number	Version Date	Protocol updated & finalised by;	Reasons for Update
Previous	0.1	01/02/2023	Dr Anoop D. Shah	Initial version shared with JRO
Previous	0.2	13/02/2023	Dr Anoop D. Shah	Current version
Previous	0.3	12/04/2023	Dr Jack Ross and Dr Anoop D. Shah	Incorporated feedback from JRO
Previous	1.0	22/05/2023	Anoop D. Shah	Version for initial submission to ethics committee
Current	1.1	24/07/2023	Anoop D. Shah	Amendments as requested by ethics committee: removal of patients who lack capacity from observed consultations

DECLARATIONS

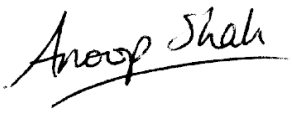
The undersigned confirm that the following protocol has been agreed and accepted and that the investigator agrees to conduct the study in compliance with the approved protocol and will adhere to the U.K. Policy Framework for Health and Social Care Research 2017 (3rd edition) (as amended thereafter), the EU General Data Protection Regulation (2016/679) and the UK Data Protection Act (2018), Sponsor SOPs and applicable Trust policies and legal frameworks.

I, Anoop Shah, agree to ensure that the confidential information contained in this document will not be used for any other purposes other than the evaluation or conduct of the research investigation without the prior written consent of the Sponsor.

I (investigator) agree to ensure that no research activity or recruitment will commence at participating research sites until the appropriate regulatory approvals and NHS confirmations of Capacity and Capability have been issued, and Sponsor green light confirmed.

I (investigator) also confirm that I will make the findings of the study publicly available through publication or other dissemination tools without any unnecessary delay and that an honest, accurate and transparent account of the study will be given. Any deviations from the study as planned in this protocol will be explained and reported accordingly.

Chief Investigator:

Signature:  **Date** 22/05/2023

Print Name (in full): Anoop Dinesh Shah

Position: Associate Professor

On behalf of the Study Sponsor:

Signature:  **Date** 26/05/2023

Print Name (in full): Pushpsen Joshi

Position: Research Governance Manager

STUDY SUMMARY

IDENTIFIERS	
IRAS Number	322887
REC Reference No.	
Sponsor Reference No.	157143
Other research reference number(s) (if applicable)	UCL data protection number: Z6364106/2023/03/03 health research
Full (Scientific) title	Non-randomised feasibility study of point of care natural language processing using the MiADE system to assist structured clinical documentation
Health condition(s) or problem(s) studied	Recording of diagnoses, medications and allergies in electronic health records
Study Type i.e. Cohort etc	Non-randomised intervention study utilising mixed methods
Aim(s):	To clinically evaluate a system to support automatic conversion of clinicians' text into structured data at the point of care.
Objectives:	Primary: To ascertain whether the MiADE point of care NLP system increases the number of diagnoses, medication and allergies recorded by clinicians in a structured way in the electronic health record. Secondary: • To evaluate the usability of the MiADE system • To evaluate the accuracy and technical performance of the MiADE system • To evaluate the effect of using MiADE on patient consultations • To explore how the MiADE system can be improved
Type of trial:	Single centre non-randomised intervention study
Trial design and methods:	The trial consists of two non-randomised intervention substudies with before and after comparison (outpatient and inpatient), with embedded qualitative substudies involving clinician interviews and (in the outpatient setting only) observed consultations and patient interviews. Clinicians will be recruited, consented and provided with training to use the MiADE system. The system will be switched on at a particular time point, and clinical records for patients seen before and after the system was switched on will be compared. The primary outcome will be the comparison of the number of diagnoses, medication and allergies entered in a structured way per

	patient. A rapid ethnography-based approach will be used to observe a subset of patient consultations to find out how clinicians are using the system, and semi-structured interviews will be held with patients and clinicians to understand the effect of social, cultural and behavioural information on human experiences and practices.
Trial duration per participant:	5 months
Key Study milestones	Jul 2023: Clinician recruitment and consent Aug 2023: Training of clinicians, surveys and observed consultations without MiADE Sep 2023: MiADE switch on Oct 2023: Interviews, surveys and observed consultations with MiADE Nov 2023: Study end Dec 2023 – Jan 2024: Data analysis
Estimated total trial duration:	7 months
Planned trial sites:	Single-site, UCLH
Total number of participants planned:	Outpatient study: Clinicians using MiADE: at least 25, target 50 Patients for analysis of pseudonymised health record: estimated 3000 (assuming 25 clinicians each see 5 new patients per week, and 24 weeks of data are analysed) Clinicians for observed consultations and interview: 5 Patients for observed consultation and interview: at least 20 (10 before MiADE, 10 after MiADE) Inpatient study: Clinicians using MiADE: at least 25, target 50 Patients for analysis of pseudonymised health record: estimated 600 (assuming 5 inpatient teams each discharge 5 patients per week, and 24 weeks of data are analysed)
Main inclusion/exclusion criteria:	Electronic health record of patients being seen in clinics under the care of participating clinicians
Statistical methodology	Outpatients: comparison (pre- and post-intervention) of mean

and analysis:	number per patient of problems, medications and allergies recorded in structured data by the clinician. Outpatient observation substudy: comparison (pre- and post-intervention) of proportion of problems, medications and allergies recorded in structured data per patient, with manual review of free text clinical notes considered the 'gold standard'. Descriptive summary statistics will be used to describe the participant demographics for observations and interviews, and qualitative results will be analysed using thematic analysis mapped to the theoretical domains framework. Inpatients: comparison (pre- and post-intervention) of proportion of ICD-10 coded diagnoses for hospital admission which have a corresponding structured problem list entry.
FUNDING & OTHER	
Funding	National Institute of Health Research
Other support	Support from the UCL/UCLH NIHR Biomedical Research Centre and the Clinical and Research Informatics Unit
STORAGE of SAMPLES / DATA (if applicable)	
Human tissue samples	Not applicable
Data collected / Storage	The MiADE system will be implemented wholly within the hospital (UCLH) network. All research data will be stored on UCLH servers or in the UCL Data Safe Haven.
KEY STUDY CONTACTS	
Committees	The overall NIHR funded project of which this study is a part is overseen by the MiADE steering committee. The project co-leads (Anoop D. Shah and Wai Keong Wong) jointly chair the MiADE steering committee.
Sub-contractors	None
Other relevant study personnel	Project co-lead and Chief Research Information Officer, UCLH: Wai Keong Wong Lead for clinician engagement: Jack Ross Software developers: James Brandreth, Jennifer Jiang Natural language processing expert: Richard Dobson Clinical data standards lead: Leilei Zhu Clinical safety and evaluation of digital health interventions: Yogini Jani Human-computer interaction expert: Enrico Costanza

KEY ROLES AND RESPONSIBILITIES

SPONSOR: The sponsor is responsible for ensuring before a study begins that arrangements are in place for the research team to access resources and support to deliver the research as proposed and allocate responsibilities for the management, monitoring and reporting of the research. The Sponsor also must be satisfied there is agreement on appropriate arrangements to record, report and review significant developments as the research proceeds, and approve any modifications to the design.

FUNDER: The funder is the entity that will provide the funds (financial support) for the conduction of the study. Funders are expected to provide assistance to any enquiry, audit or investigation related to the funded work.

CHIEF INVESTIGATOR (CI): The person who takes overall responsibility for the design, conduct and reporting of a study. If the study involves researchers at more than once site, the CI takes on the primary responsibility whether he/she is an investigator at any particular site.

The CI role is to complete and to ensure that all relevant regulatory approvals and confirmations of NHS Capacity and Capability are in place before the study begins. Ensure arrangements are in place for good study conduct, robust monitoring and reporting, including prompt reporting of incidents, this includes putting in place adequate training for study staff to conduct the study as per the protocol and relevant standards.

The Chief Investigator is responsible for submission of annual reports as required. The Chief Investigator will notify the REC and JRO of the end of the study (including the reasons for premature termination, where applicable). Within one year after the end of study, the Chief Investigator will submit a final report with the results, including any publications/abstracts to the REC and JRO.

PRINCIPAL INVESTIGATOR (PI): Individually or as leader of the researchers at a site; ensuring that the study is conducted as per the approved study protocol, and report/notify the relevant parties – this includes the CI of any breaches or incidents related to the study.

TRIAL PERSONNEL

See protocol cover page for Chief Investigator and Sponsor contact details.

Study Coordinator	Name: Anoop D. Shah E-mail: a.shah@ucl.ac.uk Tel: 07876767478 fax: N/A
Statistician	Name: Anoop D. Shah E-mail: a.shah@ucl.ac.uk Tel: 07876767478 fax: N/A

KEY WORDS

Natural language processing, electronic health record, diagnosis, allergy, medication, problem list

LIST OF ABBREVIATIONS

AE – Adverse event

CDA – Clinical Document Architecture

CRF – Case report form

EHR – Electronic health record

GOSH – Great Ormond Street Hospital

HL7 CDA – Health Level 7 Clinical Document Architecture

HRA – Health Research Authority

IPR – Intellectual property rights

ISG – Information Security Group

JRO – Joint Research Office

MedCAT – Medical Concept Annotation Tool

MiADE – Medical Information AI Data Extractor

NLP – Natural language processing

PPI – Patient and public involvement

PRSB – Professional Record Standards Body

REC –Research Ethics Committee

SAE – Serious adverse event

TSC – Trial Steering Committee

TMG – Trial Management Group

UCLH – University College London Hospitals

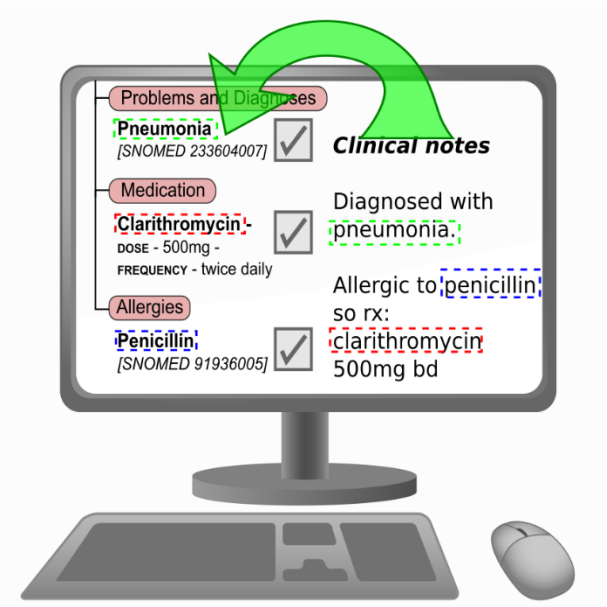
MiADE – Medical Information AI Data Extractor

EHR – Electronic health records

NLP – Natural language processing

Contents

Figure 1: Overview of MiADE system showing suggested structured data items extracted from free text



1. INTRODUCTION

MiADE is a natural language processing system designed to be embedded within electronic health records to enable unstructured text entered by clinicians to be converted into structured data in real time. Structured entries (such as diagnoses, medication and allergies) extracted from the text are presented to the clinician for validation, and can be amended or corrected before being committed to the record. The system is designed to make it easier for clinicians to enter structured data and hence improve the completeness and quality of data in electronic health records (Figure 1).

This study aims to evaluate MiADE to assess its performance and utility, and gather information to inform future improvements and developments. The evaluation consists of two before-and-after studies with qualitative and quantitative assessment. Study 1 will be conducted in the outpatient setting, in which the number of structured entries per patient will be compared before and after switching on the MiADE system, and feedback will be gathered from clinicians. A detailed qualitative substudy will recruit clinicians and patients for observed consultations and semi-structured interviews. Study 2 will take place within the inpatient setting, where inpatient clinical teams will be recruited and surveyed, and electronic health records of patients will be compared before and after MiADE system was switched on.

The findings from this study will be used to support further improvement and roll out of natural language processing at the point of care.

2. BACKGROUND AND RATIONALE

The need for structured data in electronic health records

This NIHR funded project aims to develop a natural language processing (NLP) system to make it easier for clinicians to record structured data in electronic health records (EHRs) for patient care and research. High quality data are essential to realise the benefits of the NHS Long Term Plan [1] such as early diagnosis, better coordination of care and avoidance of unnecessary admissions. The ever-increasing knowledge base in medicine and complexity of treatments makes computer-assisted decision support increasingly important. Detailed recording of patients' clinical characteristics (phenotype) is essential for precision medicine initiatives [2]. In the Chief Medical Officer's 2016 annual report, Dame Sally Davies stated that "'in order to deliver [genomic medicine services] we will also need national standards for defining the symptoms (the phenotype) of patients in a systematic way." [3]

It has long been recognised that well-organised medical records are essential for high quality patient care [4], but today's records are far from ideal. Although patient data are increasingly being collected electronically [5], much of the information is in free text rather than in a structured form [6], and is difficult to use clinically or for research. Although free text is the most faithful way to represent some types of information, such as patient stories, structured recording of diagnoses and other key items of clinical information is recommended by national guidance from the Professional Record Standards Body (PRSB) [7]. Widespread use of controlled clinical terminologies such as SNOMED CT enables clinical concepts to be recorded in a consistent way, but only if clinicians are able to use the system easily. It can be onerous and time-consuming for clinicians to enter detailed structured information in many EHR systems [6], and time spent on data entry can affect the human experience of clinical consultations [8]. There is a need for methods to facilitate data entry without impeding the clinical workflow.

Overall, free text is common in EHRs and structured fields are under-used, making it difficult to retrieve information for clinical care, research or service planning.

One way to reduce the burden on clinicians entering structured data is to minimise the amount of data that needs to be entered, by improving interoperability and the use of shared health records. There are international standards for the design of health records (openEHR) [9] and standards for communication between systems (Fast Healthcare Interoperable Resources, FHIR) [10]. The PRSB publishes professional requirements for health records and interoperability, which are being converted into technical standards for implementation by 'INTEROPen', a collaboration of UK EHR software companies, NHS Digital and other stakeholders [11].

Structured clinical records could be linked directly to audit and registry reporting systems, avoiding the wasteful double entry that currently takes place. They would also enable rapid research using clinical records for emerging diseases such as COVID-19, avoiding the need for wasteful post-hoc duplication of data entry.

The need for NLP at the point of care

Despite major research initiatives to analyse clinical text, NLP technology has been under-utilised in clinical applications [12]. A fundamental problem is that information extracted from text may be accurate enough for population-level inferences (where a small error rate can be accommodated statistically), but may not be safe for making clinical decisions for individual patients without a

clinician in the loop. Even seemingly simple text classification tasks such as negation detection can be prone to error [13]. It is for this reason that it is not possible to use NLP retrospectively to safely populate structured data fields in clinical records. Instead, our proposed NLP solution will analyse free text as soon as it is created, and suggest structured entries which will be verified by the clinician before being committed to the record. These structured data can be used immediately for clinical decision support as well as for secondary uses such as to assist clinical coding, or for research. Immediate user feedback in our design will also enable the system to improve and learn more quickly.

The NHS Common User Interface guidelines [14] contained recommendations for user interfaces incorporating real-time NLP to assist terminology binding. MiADE is an open source system designed around these principles.

Benefits to patients

- Transfers of care – a clear, structured summary of diagnoses and treatments is invaluable as a handover to a new team looking after a patient, whether between shifts in hospital, or when patients are discharged or transferred between care settings.
- Decision support and patient safety – many EHR systems include automatic warnings of medication allergies and interactions, and automated reminders for monitoring of chronic diseases. All these decision support aids rely on accurate, structured data to be present in the EHR. Clinical error is a major source of patient harm.
- Electronic health records are used in a large number of research studies for patient benefit. All these studies rely on high quality data; missing data can introduce bias and might result in inaccurate study outcomes which can lead to patient harm. If clinically-recorded data are not sufficiently complete, time-consuming retrospective data entry may be needed.
- Clinical trials are vital for developing and evaluating new treatments, but many trials fail to recruit an adequate number of participants [26]. Automated algorithms can help to detect patients eligible for certain trials, but only if the EHR contains high quality data [27].

We believe that all patient groups will benefit, but sicker patients or those with more complex clinical histories may benefit more, as they may be at more risk of harm from clinical error due to missing information.

The proposed NLP system will enable the advantages of structured data to be realised while avoiding the disadvantage of the burden on clinicians entering the data. The patient experience of consultations may also improve if clinicians have to spend less time entering data into the computer [8].

Benefits to the NHS and the wider population

This project will make it easier to use data for purposes beyond individual care. Although existing NLP approaches are being applied to health record databases, data need to be validated before they are used for decisions that may impact patients. Our approach enables immediate validation, and the data can be used for operational research, service planning, audit, safety monitoring and clinical coding in near real time. A recent audit in UCLH estimated that only 62% of diagnoses were included in problem lists.

3. OBJECTIVES AND OUTCOME MEASURES/ENDPOINTS

The overall aim of this study is to evaluate the effectiveness of the MiADE point of care natural language processing (NLP) system for improving the recording of diagnoses, medications and allergies in electronic health records (EHR), and to investigate factors affecting the effectiveness of the system.

3.1 Primary Objective

To ascertain whether the MiADE point of care natural language processing system increases the number of diagnoses, medication and allergies recorded by clinicians in a structured way in the electronic health record.

3.2 Secondary Objective(s)

- To evaluate the usability of the MiADE system
- To evaluate the accuracy and technical performance of the MiADE system
- To evaluate the effect of using MiADE on patient consultations
- To explore how the MiADE system can be improved

3.3 Outcome measures/endpoints

Primary outcome measures:

Study 1 (outpatient): Number of structured entries for diagnoses, medication and allergies recorded by the clinician in outpatient consultations.

Study 2 (inpatient): The proportion of ICD-10 coded billing diagnoses for which a similar SNOMED CT concept is present in the problem list for inpatients, evaluated at the point of discharge.

Secondary outcome measures:

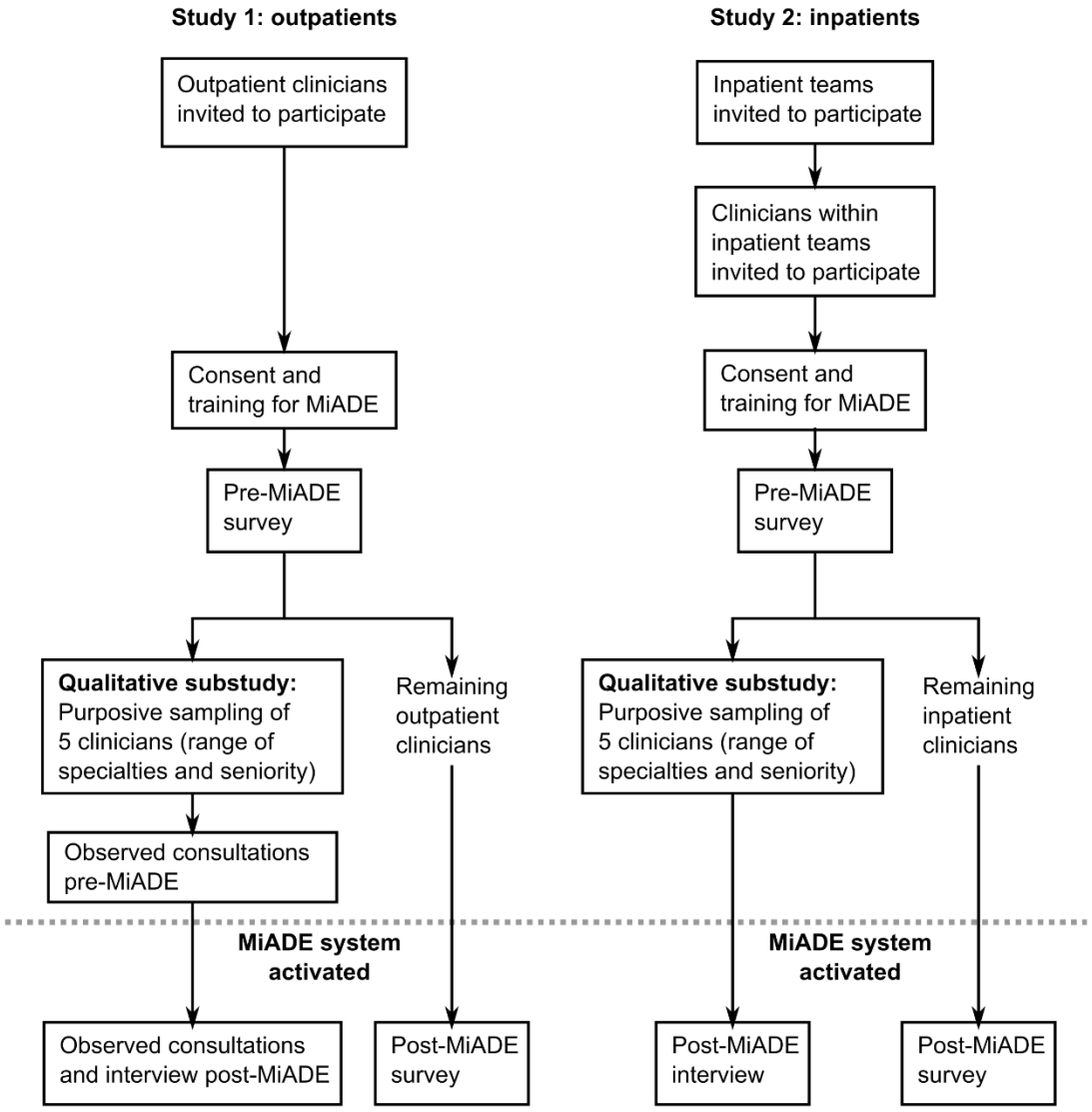
- The proportion of structured data items suggested by MiADE that are accepted by the clinician for entry into the structured record.
- Computing time required per consultation note.
- Clinician and patient perceptions of the MiADE system.
- Changes to consultation workflow or process as a result of using the MiADE systems

4. TRIAL DESIGN

This study consists of two substudies (inpatient and outpatient studies), each of which is a non-randomised intervention study with before and after comparison. The outpatient study also includes an embedded qualitative substudy for a subset of consultations, with patient and clinician

interviews. A subset of clinicians will also be interviewed for the inpatient study. The overall clinician participant flow is illustrated in Figure 2.

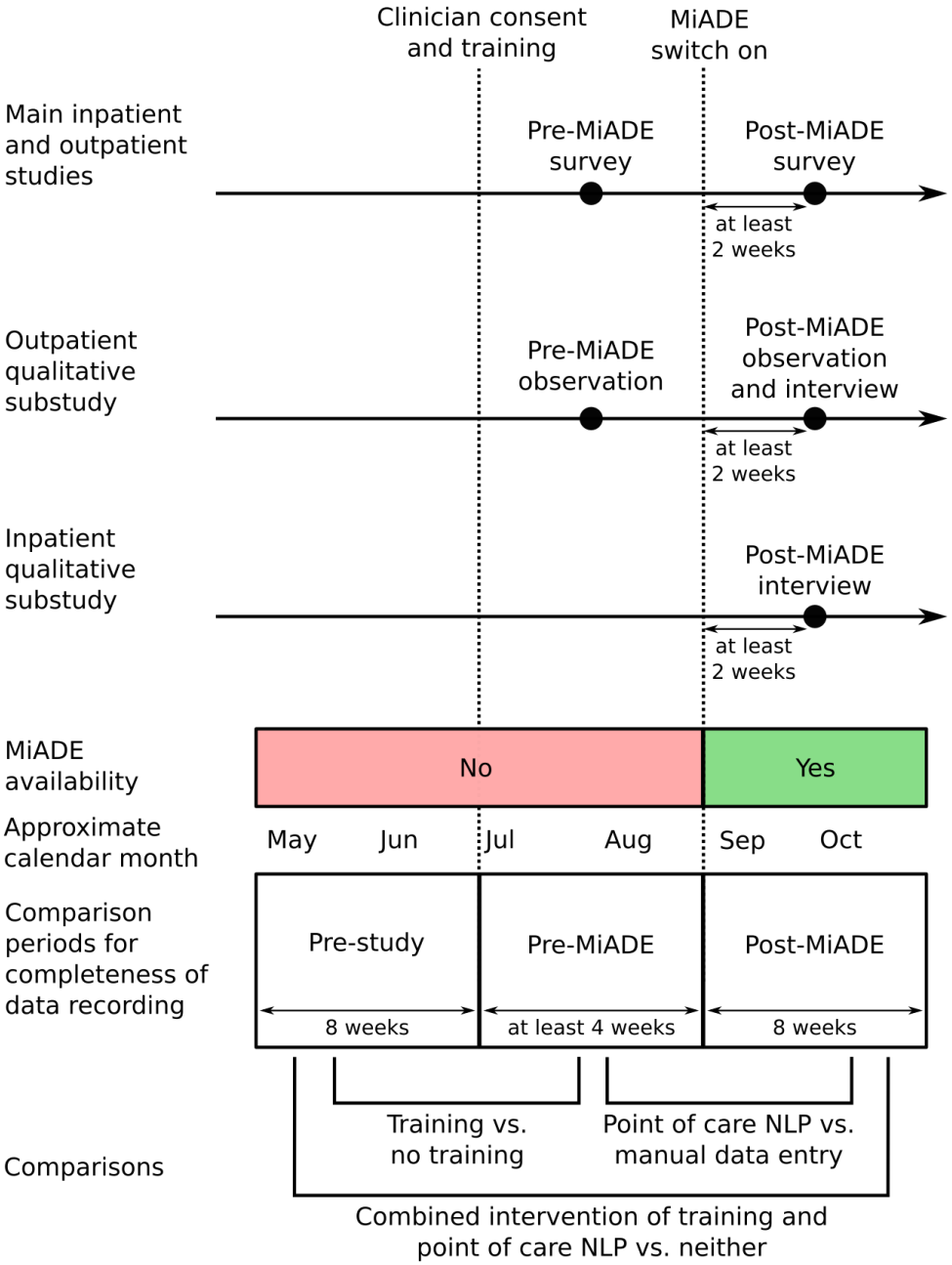
Figure 2: Study flow diagram showing how clinicians will move through the study



Three time periods will be analysed. The primary comparison will be for recording of structured information in patient records between pre-MiADE and post-MiADE periods once clinicians have been consented into the study, as they will have received training and the comparison will show the change in behaviour that results from the system being available. There is a possibility that structured recording may improve because of the study itself (i.e. a Hawthorne effect), so a prior 8-week period will also be studied in order to evaluate this effect. See figure below for the overall trial design and timing of interventions. The overall study timeline is shown in Figure 3.

The MiADE system will be made available according to clinician-level permissions, which can be switched on or off. This will allow clinicians' use of problem lists and other structured data to be compared before and after the system is switched on. There will be a period of at least 4 weeks between clinicians entering the study and having the MiADE system switched on.

Figure 3: Study timeline



Training programme

All clinicians in the study will undergo a standardised training package consisting of viewing a video and reading a user guide, with access to help from the research team by email or phone throughout the study. The training will cover:

- The importance of maintaining problem lists with accurately coded information about a patient's diagnoses
- National and local (UCLH) guidance on the use of problem lists
- How to use the MiADE system
- The aims and methods of the study

Clinicians will not be eligible to enter the study and have the MiADE system switched on for their EHR user profile until they have undertaken the training.

Pre-MiADE questionnaire – questions for all clinicians

- What is your seniority (e.g. FY1, Consultant etc.)?
- What is your specialty?

Pre-MiADE questionnaire – outpatient study

- Regarding the recording of problems or diagnoses:
    - How frequently do you enter problem list entries or diagnoses when seeing a patient in clinic? [Never – Rarely – Sometimes – Usually – Always]
    - How easy or difficult did you find it to use the Epic interface for entering problems or diagnoses? [Very difficult – Difficult – Neither easy nor difficult – Easy – Very easy]
    - How important do you feel it is to have structured information about problems in the health record? [Very unimportant – Unimportant – Neither important nor unimportant – Important – Very Important]
- Regarding the recording of medications:
    - How frequently do you enter medications in the structured medications section when seeing a patient in clinic? [Never – Rarely – Sometimes – Usually – Always]
    - How easy or difficult do you find it to use the Epic interface for entering medications in the structured medications section? [Very difficult – Difficult – Neither easy nor difficult – Easy – Very easy]
    - How important do you feel it is to have structured information about medications in the health record? [Very unimportant – Unimportant – Neither important nor unimportant – Important – Very Important]
- Regarding the recording of allergies or medication intolerances:
    - How frequently do you enter allergies or medication intolerances seeing a patient in clinic? [Never – Rarely – Sometimes – Usually – Always]
    - How easy or difficult did you find it to use the Epic interface for entering allergies and medication intolerances? [Very difficult – Difficult – Neither easy nor difficult – Easy – Very easy]

- How important do you feel it is to have structured information about allergies in the health record? [Very unimportant – Unimportant – Neither important nor unimportant – Important – Very Important]

Pre-MiADE questionnaire – inpatient study

- Regarding the recording of problems or diagnoses:
  - How frequently do you enter problem list entries or diagnoses when clerking a patient or reviewing a patient on a ward round? [Never – Rarely – Sometimes – Usually – Always]
  - How frequently do you enter problem list entries or diagnoses when completing a discharge summary? [Never – Rarely – Sometimes – Usually – Always]
  - How easy or difficult did you find it to use the Epic interface for entering problems or diagnoses? [Very difficult – Difficult – Neither easy nor difficult – Easy – Very easy]
  - How important do you feel it is to have structured information about problems in the health record? [Very unimportant – Unimportant – Neither important nor unimportant – Important – Very Important]
- Regarding the recording of allergies or medication intolerances:
  - How frequently do you enter allergies or medication intolerances when clerking a patient or reviewing a patient on a ward round? [Never – Rarely – Sometimes – Usually – Always]
  - How frequently do you enter allergies or medication intolerances when completing a discharge summary? [Never – Rarely – Sometimes – Usually – Always]
  - How easy or difficult did you find it to use the Epic interface for entering allergies and medication intolerances? [Very difficult – Difficult – Neither easy nor difficult – Easy – Very easy]
  - How important do you feel it is to have structured information about allergies in the health record? [Very unimportant – Unimportant – Neither important nor unimportant – Important – Very Important]

Post-MiADE questionnaire

After MiADE has been switched on and clinicians have been using it for at least two weeks, they will be asked to fill in a survey questionnaire with the following questions:

- How many weeks have you been using the MiADE system?
- Regarding the recording of problems or diagnoses:
  - How frequently did you use MiADE for entering problem list entries or diagnoses? [Never – Rarely – Sometimes – Usually – Always]
  - When using MiADE for entering problems or diagnoses, were the suggested SNOMED CT concepts correct? [Never – Rarely – Sometimes – Usually – Always]

  - How easy or difficult did you find it to use the MiADE system for entering problems or diagnoses? [Very difficult – Difficult – Neither easy nor difficult – Easy – Very easy]
  - How useful did you find the MiADE system for entering problems or diagnoses? [Not at all useful – Somewhat useful – Useful – Very useful]
- Regarding the recording of allergies and medication intolerances:
  - How frequently did you use MiADE for entering allergies and medication intolerances? [Never – Rarely – Sometimes – Usually – Always]
  - When using MiADE for entering allergies and medication intolerances, were the suggested entries correct? [Never – Rarely – Sometimes – Usually – Always]
  - How easy or difficult did you find it to use the MiADE system for entering allergies and medication intolerances? [Very difficult – Difficult – Neither easy nor difficult – Easy – Very easy]
  - How useful did you find the MiADE system for entering allergies? [Not at all useful – Somewhat useful – Useful – Very useful]
- How do you think MiADE could be improved? [free text]
- How do you think the Epic user interface for entering structured information could be improved? [free text]

Post-MiADE questionnaire – additional questions for clinicians in outpatient study

- How many clinics have you conducted with the MiADE system? (*only for clinicians in outpatient study*) [free text]
- Regarding the recording of medications:
  - How frequently did you use MiADE for entering medications during an outpatient consultation? [Never – Rarely – Sometimes – Usually – Always]
  - When using MiADE for entering problems or diagnoses, were the suggested medications and dosages correct? [Never – Rarely – Sometimes – Usually – Always]
  - How easy or difficult did you find it to use the MiADE system for entering medications? [Very difficult – Difficult – Neither easy nor difficult – Easy – Very easy]
  - How useful did you find the MiADE system for entering medications? [Not at all useful – Somewhat useful – Useful – Very useful]

Post-MiADE interview prompts

- In your experience, how completely are problems, medications and allergies recorded in clinical practice?
- Why do you think this is?
- How effective do you feel MiADE is in increasing the recording of structured information about problems, medications, allergies and medication intolerances?
- Why do you think this is?
- How do you think MiADE could be improved?

- How do you think the Epic user interface for entering structured information could be improved?
- What else needs to change to improve the recording of structured information?

Study 1 (Outpatient study)

Consultants and specialty doctors who see outpatients will be recruited via institutional communications and networks. Interested clinicians will be consented and receive training on the use of the system, as described below. They will be asked to fill in a survey on their level of experience and comfort with EHR systems. They will have the MiADE system switched on within their EHR user profile so that they have access to it from a particular date. Electronic health records of outpatients in the clinics linked to the clinician will be assessed for 3 time periods: pre-study (before the training session), between the training session and MiADE switch-on, and for 8 weeks after MiADE is switched on. The crude number of entries per patient in the problem list, medication record and allergy record will be calculated for outpatients seen by the clinician during each interval.

Embedded qualitative substudy

For each clinician in the qualitative substudy, a sample of at least 2 outpatient consultations will be observed before the MiADE system is switched on for the clinician, and at least 2 outpatient consultations will be observed after the MiADE system is switched on. For each observed consultation, consent will be sought from the patient prior to the consultation. The researcher will sit in the consultation and observe the overall consultation, as well as timing the overall consultation and how long the clinician spends interacting with the EHR. The researcher will ask questions to the clinician after the consultation, if necessary, in order to clarify any behaviours observed.

Patient interviews

A semi-structured interview will be held with the patient after the consultation either in person (if clinic space is available) or by telephone, to explore the patient's perception and experience of the consultation.

The interview guide will include questions such as:

- How did you feel about the way that the doctor interacted with the computer in your consultation?
- How did this compare with your experience of previous consultations in this hospital or elsewhere?
- Use and usefulness of the UCLH patient portal (MyCare): prompts include
    - Access to medical diagnoses, allergies and medications on the UCLH patient portal
    - Usefulness of the features
    - Completeness and accuracy of the information
- Would you be interested in being invited to future workshops or other activities to help improve electronic health records?

Study 2 (Inpatient study)

Inpatient clinical teams will be recruited to the study. Within each team, the clinicians (consultants, specialty and foundation doctors who see inpatients) will be recruited, consented and trained. Doctors rotate through clinical teams every few months, so the MiADE system will be switched on for the relevant clinicians at a time point half-way through the four-month rotation, to allow comparison of the completeness of structured recording of diagnoses on the problem list before and after the switch-on. The 'gold standard' of recording of diagnoses will be the ICD-10 coded billing diagnoses which are assigned by the clinical coding team based on a review of the entire medical record after a patient is discharged. Diagnoses will be aggregated by groups of ICD-10 codes similar to a previous UCLH audit [15], and a diagnosis will be considered as included in the problem list if any SNOMED CT concept which maps to an ICD-10 code in the group was recorded on or prior to the patient's discharge date. This flexibility of mapping will be permitted as different professionals may assign different clinical codes to the same condition based on their interpretation and judgement, especially for cases where an exact code does not exist. The proportion of ICD-10 code groups with a problem list entry will be calculated per patient, and aggregated across the patients under the care of the clinical team during the study period.

For a sample of 5 clinicians, semi-structured interviews will be held with the clinicians after they have been using MiADE for at least two weeks, using the same questions as for outpatient doctors except the question on medications, as MiADE for inpatients does not handle medications.

5. SAMPLING METHODS

Study 1 (outpatient): Clinicians who carry out outpatient clinics will be eligible to take part. Clinicians for the observation substudy will be purposively sampled from the interested clinicians to include a range of specialties and seniorities. One or more face to face clinics will be sampled before and after MiADE, and consecutive patients booked to attend the clinics will be invited to have their consultation observed as part of the study.

Study 2 (inpatient): Inpatient teams will be recruited to the study. All clinicians within interested teams will be eligible to be included. All patients managed by these teams will be included for pseudonymised analysis of their health records.

5.1 Inclusion criteria

Study 1 clinician inclusion criteria: Any interested clinicians who have their own clinic list at UCLH and see outpatients, who consent to be part of the study. A subset of at least 5 clinicians will be purposively sampled to include a range of specialties for the observation substudy, and they will undergo observed consultations.

Study 1 patient inclusion criteria: All outpatients seen by the clinician from 2 months before being consented until 2 months after the MiADE system is switched on for included clinicians. For clinicians selected for observation (at least 5), we will observe at least 2 consultations from one or more clinics before and 2 consultations from one or more clinics after the MiADE system is switched on will be observed. Hence the minimum number of observed consultations will be 20.

Study 2 clinician inclusion criteria: All clinicians within interested inpatient teams at UCLH.

Study 2 patient inclusion criteria: All inpatients managed under the care of a participating team during the study period.

5.2 Exclusion criteria

Patients who lack capacity to consent will not be eligible for inclusion in the observed consultations and patient interviews.

5.3 Recruitment

Clinicians will be recruited via clinical networks and clinical operational structures within the hospital such as Divisional meetings and the Epic Expert Community.

Consecutive patients from selected clinics will be recruited for the outpatient qualitative substudy.

Participant recruitment at a site will only commence when the trial has:

- Been confirmed by the Sponsor (or its delegated representative), and
- Been issued with Confirmation of Capacity and Capability from each participating site (where applicable).

5.4 Informed Consent

Clinician participants

Informed consent will be sought from all clinician participants. Consent will be sought from patients whose consultation is being observed as part of the qualitative substudy. For patients whose de-identified data are being used for analysis, an opt-out model will be used, in common with other studies using de-identified patient data at UCLH.

It is the responsibility of the Investigator, or a person delegated by the Investigator to obtain written informed consent from each participant prior to participation in the trial, following adequate explanation of the aims, methods, anticipated benefits and potential hazards of the trial.

The person taking consent will be suitably qualified and experienced, and will have been delegated this duty by the CI/ PI on the Staff Signature and Delegation of Tasks.

Patient participants

For patients whose pseudonymised data are included in the study to assess recording of structured data, consent will not be sought.

For patients included in the outpatient observation substudy, written consent will be sought. Information about the study will be provided by the researcher prior to the clinic, and if the patient agrees to participate they will be asked to give written consent prior to their consultation. Adequate time will be given for consideration by the participant before taking part. Patients who lack capacity will not be eligible to participate in the observed consultation.

The researcher will explain that participants are under no obligation to enter the trial and that they can withdraw at any time during the trial, without having to give a reason and without prejudicing his/her further treatment. Data collected up to the point of withdrawal can only be used after withdrawal if the participant has consented for this, and this will be outlined in the consent form. Where a participant is required to re-consent or new information is required to be provided to a participant, the researcher will ensure that this is done in a timely manner.

No study procedures will be conducted prior to the participant giving consent by signing the Consent form.

A copy of the signed Informed Consent form will be given to the participant. The original signed form will be retained in the Investigator Site File and a copy placed in the medical notes.

The PIS and consent form will be reviewed and updated if necessary throughout the study and participants will be re-consented as appropriate.

5.4.1 Additional consent provisions for collection and use of participant data and human tissue in ancillary studies (if applicable)

Not applicable for this study.

6. PRODUCT/INTERVENTIONS

6.1 Name and description of intervention under investigation

The digital health technology being investigated in this study is MiADE, a natural language processing software system embedded within the electronic health record to facilitate the point of care recording of structured information in electronic health records. It is a Tier B technology according to the NICE evidence standards framework for digital technologies.

System overview

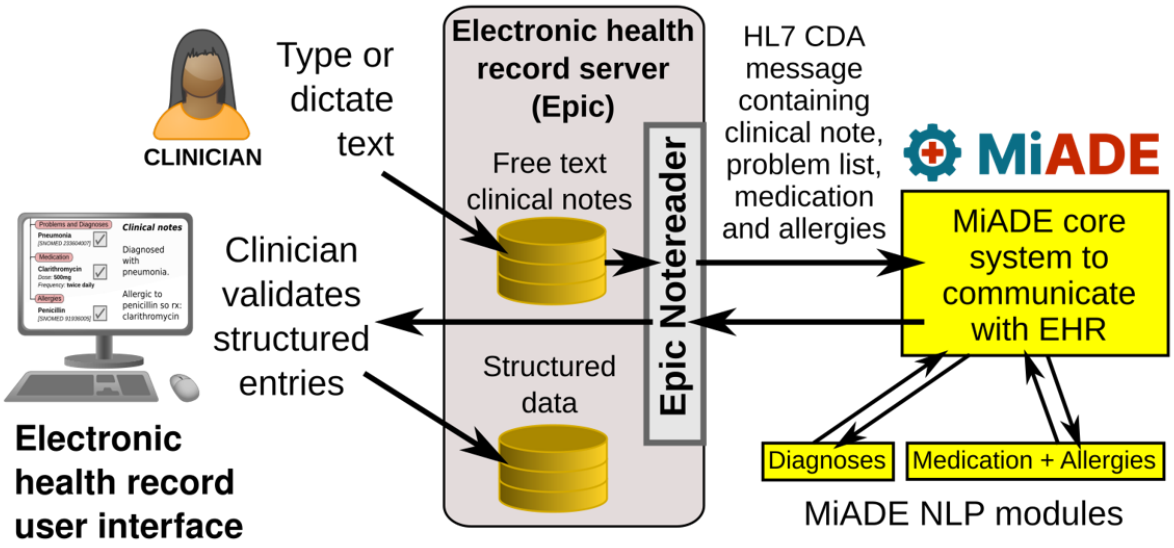
MiADE will interact with the 'Epic' electronic health record system via a user interface component called 'NoteReader'. NoteReader enables newly created text notes to be analysed by a third party NLP system, and presents structured data to the clinician for verification and entry into the record. NoteReader supports data entry for problem list entries (diagnoses), medication (including dosage), allergies and other custom data fields. The structured data entries suggested by the NLP system are displayed alongside the draft text note, and are discarded when the note is saved or 'committed' to the EHR. Thus the NLP suggestions do not form part of the medicolegal record, but exist temporarily only in the draft form of the document.

Epic NoteReader is already used by a number of US hospitals with third party commercial NLP systems, as well as one UK hospital (Great Ormond Street Hospital which is using 3M M*Modal Fluency as the third-party NLP system).

MiADE is an open source software system created by the research team which will be evaluated in this study. It is based on an information extraction and linking engine called MedCAT (Medical

Concept Annotation Tool) [16], with additional modules for dosage extraction, context detection, deduplication and integration with an EHR system.

Figure 4: Data flows between MiADE and Epic



MiADE will be run as a service on UCLH servers within the NHS network, so no patient information will leave the Trust. The service will communicate with Epic using HL7 CDA (Health Level 7 Clinical Document Architecture) format messaging (Figure 4). This is an XML format containing patient demographics, existing structured problem list entries, allergies, medication and the free text clinical note. MiADE will process the CDA sent by Epic and return a new CDA document containing information extracted from the text. Epic NoteReader compares the new entries in the CDA with those already in the record and displays them to the user split by category and new / existing. The clinician can click a button to add each entry to the patient's structured data record.

Evidence standards for digital health technology

We have sought advice from NICE regarding their evidence standards framework for digital health technologies, which was published in 2019 and updated in August 2022 (https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies). It sets out the appropriate evidence standards for different types of digital health technologies, with evidence standards defined by the intended use and potential impact of the technology.

MiADE is a documentation aid, similar to medical voice recognition systems. Voice recognition systems convert spoken words into typed text, which is then reviewed by the clinician before being saved as part of the patient's medicolegal record. MiADE converts typed text into structured data, which is reviewed by a clinician before being committed to the record. We propose that it is a Tier B digital health technology (communication between health professionals and services), and would not be considered a medical device by MHRA. A similar commercial clinical NLP software system (3M

M*Modal, as used in Great Ormond Street Hospital) and clinical voice recognition systems are not registered as medical devices on the MHRA database (https://pard.mhra.gov.uk/).

There is low risk to patients if the algorithm does not perform as expected, because the results are reviewed for accuracy by a clinician before being committed to the record. The clinician can choose not to use MiADE or ignore the results, in which case the documentation would be carried out just as if MiADE was not in place.

If the technology's intended use is to support the recording of data in patients' electronic health record, which informs future clinical decision and management, it could fall into Tier C: informing clinical decisions. This would be the case particularly if there are high risks associated with the algorithm not performing as expected (such as potential danger to patients). However, this is not the case with MiADE; the system itself merely suggests structured data for information that is already in the health record in a free text form.

6.2 Storage and handling of drugs at site (if applicable)

Not applicable for this study.

6.3 Accountability of drug (if applicable)

Not applicable for this study.

6.4 Concomitant medication (if applicable)

Not applicable for this study.

6.5 Dosages, modifications, and methods of administration (if applicable)

Not applicable for this study.

7. TRIAL PROCEDURES

7.1 Pre-intervention assessments

There are no trial specific procedures to be carried out after consent to assess patient or clinician eligibility.

7.2 Registration/Randomisation Procedures (delete as appropriate) (if applicable)

Clinician participant registration will be undertaken centrally by the coordinating trial team. Clinician login names for the electronic health record system will be recorded on the study database within the UCLH system. For patients for observed consultations, the clinic code of the clinic they attend

and the time and date of their appointment will be recorded on the study database. All data will be stored on UCLH servers.

There are no randomisation procedures as this is a single-arm intervention study with before and after comparison.

7.3 Baseline and Intervention Procedures

7.3.1 Baseline data

Clinician participants: A baseline questionnaire will be administered to gather information about specialty, seniority, and self-rated level of expertise with the electronic health record system.

Patient participants: All data on patients will be collected as part of the interview after the observed consultation. This will include questions on patient characteristics including age group, gender and ethnicity. This information will be collected to ensure the diversity of the patient sample.

7.3.2 Intervention Procedures

Clinician participants: Clinicians will be consented and asked to watch a training video on the need for structured data recording and how to use the MiADE system. They will then be asked to fill in a pre-MiADE questionnaire survey on their experience of using the electronic health record system.

Clinicians in the qualitative substudy will be observed for at least 4 face to face consultations (2 before MiADE and 2 after) for which patients have given consent. They will also be interviewed after using MiADE. The time period between consent and having the MiADE system switched on will be at least 4 weeks. Patient participants in the qualitative substudy will be consented prior to their consultation and the researcher will sit in the consultation to observe the interactions between doctor, patient and the computer. A purposive sample of 5 clinicians (covering a range of inpatient teams and levels of seniority) from the inpatient substudy will also be interviewed post MiADE switch on, around the time of the questionnaire survey.

During the observed consultations, the researcher will make notes but no recording devices will be used. The researcher will leave the room or stay outside the curtains if the patient is being examined. After the consultation, the researcher may ask questions of the clinician in order to clarify any behaviours or events that were observed.

Patients will be invited to participate in an interview following the consultation. This may be immediately after the consultation or by telephone or virtually at a convenient time for them. Interviews will explore individuals' experience, understanding and perception of the MiADE system. The interview questions will be informed technology evaluation and behavioural change theories and developed from relevant literature and the experience of the research team. The interview schedule will mainly comprise open questions to allow patients to provide their own perspectives,

be iterated in consultation with our patient and public partners and piloted before use to ensure face validity.

We anticipate a total of 10-15 interviews for patients for each phase, although this will be subject to saturation checking. Through targeted patient recruitment we will ensure, as far as possible, that the interview sample reflects diversity with respect to gender, age, ethnicity and any other characteristics identified as important during the data collection period.

Audio data collected from patient and clinician interviews will be transferred from a password-protected digital recording device directly to a secure computing environment. It will then be transcribed and identifiable features removed by a member of the research team. De-identified transcripts will be used for the purposes of conducting the qualitative analysis. Anonymised quotations will be presented in the published study results.

After the MiADE system has been switched on, clinicians will be given a post-MiADE survey on their experience of using the system. Clinicians in the outpatient qualitative substudy will also be observed for at least 2 consultations post-MiADE, for which patients will be asked to consent. Clinicians should have had access to the MiADE system for at least 2 weeks before the post-MiADE survey or observation.

Patients who are under the care of a clinician participant but not part of the qualitative substudy will not experience any difference in the care they receive.

7.4 Subsequent assessments and procedures

A schedule of all trial assessments and procedures is set out in Appendix 1.

7.5 Samples (if applicable)

Not applicable for this study.

7.5.1 Laboratory assessments

Not applicable for this study.

7.5.2 Translational research samples (if applicable)

Not applicable for this study.

7.5.3 Sample storage and transfer

No samples will be collected from patients as part of this study.

7.6 Discontinuation/withdrawal of participants

In consenting to participate in the trial, patient participants are consenting to intervention, assessments, follow-up and data collection.

It is always within the remit of the physician responsible for a patient to withdraw the patient from a trial (or certain aspects of the trial) for appropriate medical reasons, adverse events or new information gained about an intervention. A participant may be withdrawn from trial whenever continued participation is no longer in the participant's best interests, but the reasons for doing so must be recorded. Reasons for discontinuing the trial may include:

- Breakdown of the doctor-patient relationship
- The topic of the consultation is particularly distressing
- Patient withdraws consent
- Patient is clinically unwell and requires emergency treatment

The decision to withdraw a participant from the trial will be recorded in the CRF and medical notes/electronic health record system. If a participant explicitly states they do not wish to contribute further data to the trial their decision must be respected and recorded in the CRF and medical notes. Any inclusion of collected data needs to be used in accordance with GDPR and HRA guidance.

7.7 Definition of End of Trial

The expected duration of the trial is 5 months from recruitment of the first participant.

The planned end of the trial is 8 weeks after the MiADE system has been switched on for the last clinician participant to enter the trial.

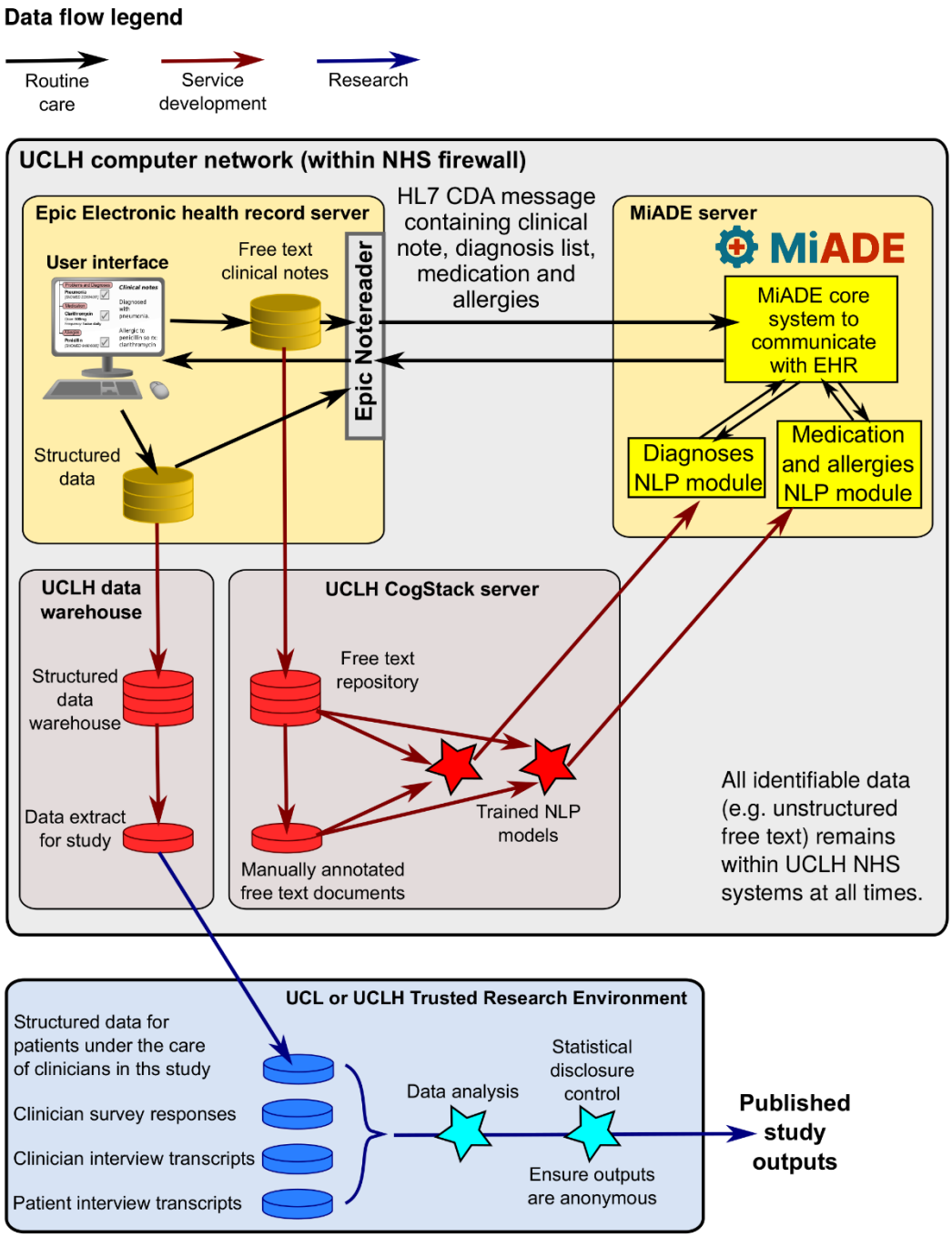
8. FINANCE AND SUPPLY OF EQUIPMENT

The research costs for the study have been supported by NIHR AI award (grant number AI_AWARD01864, amount £644k, awarded August 2020). This funding covers the programming and installation of the MiADE system for evaluation, researcher salaries, installation of computer hardware, consultancy fees and contingency.

9. DATA MANAGEMENT

The overall data flows for the MiADE evaluation study are illustrated in Figure 5.

Figure 5: Data flow diagram for MiADE evaluation study



9.1 Confidentiality

The study is compliant with the requirements of the General Data Protection Regulation (2016/679) and the UK Data Protection Act (2018). All Investigators and study site staff will comply with the requirements of the General Data Protection Regulation (2016/679) with regards to the collection, storage, processing and disclosure of personal information, and will uphold the Act's core principles.

UCL is the data controller; the UCL Data Protection Officer is data-protection@ucl.ac.uk. The data processors are Anoop Shah (PI), Jennifer Jiang (software developer), James Brandreth (software developer), Jack Ross (researcher) and other members of the Trial Management Group. The study will be collecting the following personal data:

- Consent forms from patients and clinicians
- Survey responses from clinician participants, collected electronically
- Researcher notes observing consultations
- Notes, transcripts and recordings of interviews with patients and clinicians
- Patient information from patients under the care of the clinic code or under the care of an inpatient team participating in the study will be extracted for analysis of completeness of data. The extract will be pseudonymised and will contain minimal data that is required to answer the research questions.

Paper records will be stored in a secure filing cabinet. Electronic study records will be stored on UCLH servers or in the UCL data safe haven, which has been certified to the ISO27001 information security standard and conforms to NHS Digital's Information Governance Toolkit.

The Case Report Forms (CRFs) will not bear the participant's name or other personal identifiable data. For patients in the qualitative study, the patient's clinic code, appointment time and trial identification number will be used for identification and this will be clearly explained to the patient in the patient information sheet. Patient consent for this will be sought.

9.2 Data collection tools and source document identification

Data from clinician surveys will be collected electronically using a survey tool. Data from clinician and patient interviews and patient observation will be collected on trial specific case report forms. For patients in the qualitative substudy, information about diagnoses, medication and allergies in their electronic health record (whether in free text or in structured form) will be manually abstracted into an electronic database.

It is the responsibility of the investigator to ensure the accuracy of all data entered in the CRFs. The delegation log will identify all those personnel with responsibilities for data collection and handling, including those who have access to the trial database.

9.3 Completing Case Report Forms

Once completed the original CRFs must be sent to the principal investigator.

9.4 Data Handling

In the study, observations of consultations, interview responses, and medical information abstracted from their health record will be collected from patients in the qualitative substudy in accordance with the patient consent form, patient information sheet and sections 4 and 7.3.2 of this protocol.

With the consent of the patient, interviews will be audio recorded, anonymised by the researcher and then transcribed verbatim by a professional transcribing service as soon as possible after conducting the interviews.

All qualitative data will be entered into NVivo software (QSR International (UK) Limited, Southport, UK), a data management and analysis programme to enable the application of qualitative analytical procedures which employ a system of coding and memoing developed by Lofland and Lofland (1995).

For patients who are not consented to be in the qualitative substudy but are treated by clinicians in the study, pseudonymised information about the medication, allergies and diagnoses recorded in their health record will be extracted and analysed to assess the performance of the MiADE system. All data for this study will be stored and analysed on UCLH servers. UCLH will act as the data controller of such data for the study.

UCLH will process, store and dispose of patient data in accordance with all applicable legal and regulatory requirements, including the Data Protection Act 2018 and any amendments thereto.

The patient data will not be transferred to any party not identified in this protocol and are not to be processed and/or transferred other than in accordance with the patients' consent.

Direct access to the data will be granted to authorised representatives from the Sponsor, host institution and the regulatory authorities to permit study-related monitoring, audits and inspections, in line with participant consent.

Personal Data breaches

Personal data breaches will be immediately reported to the UCL Information Security Group (ISG) and the UCL Data Protection Officer (data-protection@ucl.ac.uk), and to the Sponsor via the UCL JRO research incident reporting form (as per form and guidance: https://www.ucl.ac.uk/legal-services/guidance/reporting-loss-personal-data). The following information will be provided: full details as to the nature of the breach, an indication as to the volume of material involved, and the sensitivity of the breach (and any timeframes that apply). Sites will additionally follow their Trust incident reporting mechanisms, and will document this within their TMF/ISFs.

10.STATISTICAL CONSIDERATIONS

10.1 Primary outcome

Study 1 (outpatient): The primary outcome is the difference in the mean number of structured problem list, medication or allergy entries, comparing the time periods with and without MiADE switched on.

Study 2 (inpatient): The primary outcome is the difference in the proportion of ICD-10 discharge diagnoses with a corresponding structured problem list entry, comparing the time periods with and without MiADE switched on.

10.2 Secondary outcome(s)

To estimate the proportion of structured data items suggested by MiADE that are accepted by the clinician for entry into the structured record.

To estimate and summarise (mean, standard deviation, median, interquartile range) the distribution of computing time required per consultation note.

Clinician and patient perceptions of structured data recording in electronic health records and the MiADE system.

10.3 Sample size calculation

Based on a prior audit [15], 62% of problem list entries were entered as structured data and there were mean 8.8 problems per patient. In order to detect an increase from 62% to 70%, 549 problems are required in each group, which is 62 patients (assuming 8.8 problems per patient). This study aims to recruit at least 5 inpatient teams who would each discharge at least one inpatient per day, giving 100 patients in a 4-week period.

For outpatients, assuming each clinician sees 5 new patients each week, they will see 20 patients in a 4 week period. Using the same calculation as for inpatients, a sample size of only 5 clinicians is required to investigate the primary endpoint. However, a larger sample (target = 25) will be used in order to assess variability between clinicians.

For the patients in the qualitative substudies, there is no technique for formal sample size estimation. The number of observed clinicians (5 for inpatients and 5 for outpatient) and number of outpatient observations pre and post MiADE (2 per clinician per phase) has been selected to have a high probability of demonstrating potential themes. Using the method of Potts et al. [17], with a sample size of 5, assuming random sampling, there is 83% probability of surfacing a theme exhibited by 70% of the population.

The maximum number of clinicians to be recruited has been set at 100 according to the specification of the NLP server intended to be used for this trial.

10.4 Planned recruitment rate

The target minimum sample sizes are as follows:

Clinicians for outpatient study: minimum 25, target 50

Clinicians for observation substudy: 5

Patient consultations for observation substudy (at least 2 pre and 2 post for each clinician, total 20)

Clinicians for inpatient study: minimum 25, target 50

Inpatient clinicians for interview: 5

The clinician participants will be recruited over a short period of time (a few weeks) via clinical meetings. As per the study timeline diagram, patient participants would be recruited from eligible clinics. As this is a qualitative study only a small sample is required (5 clinicians for each substudy for

interview, and 2 outpatient consultations per clinician pre and post). Patients are seen on a single visit so there should be minimal loss to follow up, and clinicians will be easy to reach as they will most likely still be working at the Trust.

10.5 Randomisation methods

Not applicable - randomisation will not be used.

10.6 Statistical analysis

10.6.1 Summary of baseline data and flow of participants

Descriptive statistics on clinician characteristics (e.g. seniority, specialty) and patient characteristics (e.g. age distribution, gender, ethnicity) will be calculated.

10.6.2 Primary outcome analysis

For the outpatient study, multilevel Poisson regression will be used to calculate the ratio of counts of structured data entries between pre-MiADE and post-MiADE, with the clinician (clinic code) as a random effect to account for clustering by clinician. This analysis will also yield an estimate of the variability of the effect by clinician.

For the inpatient study, the primary outcome will be the difference in proportion of ICD-10 coded diagnoses recorded in the problem list comparing pre-MiADE and post-MiADE. This measure will be calculated separately for each inpatient team and then summarised as an overall difference in proportions, combining data from the different teams. For the inpatient study it will not be possible to differentiate results from patients seen by different clinicians within a team.

10.6.3 Secondary outcome analysis

The electronic health record system audit logs will be used to assess whether each structured data item suggested by MiADE is accepted by the clinician. Descriptive statistics will be calculated for this proportion among subgroups such as inpatients and outpatients, and the distribution among clinicians in the study will be calculated.

The distribution of computing times per consultation note will be calculated and summarised (mean, standard deviation, median and interquartile range).

Descriptive statistics will be calculated for each of the survey questions.

A full descriptive analysis of qualitative data will be conducted to meet study objectives. Interview transcripts and observation notes will be analysed using thematic coding mapped to the theoretical domains framework to understand the cognitive, affective, social, environmental, organisational and professional influences on behaviours relating to the use of MiADE. A researcher will code emerging themes drawing on the theoretical frameworks that underpin the interview schedule and topic

guide. An iterative approach using constant comparison will be employed in the development of coding frames and coding of data. A second researcher will read all the transcripts and code a sample to ensure reliability. Emerging themes from patient interviews will be discussed at team meetings and shared with our patient representatives to confirm that interpretations made by researchers stay close to the direct experience of patients

Interview findings from patient and clinician interviews will be transcribed, coded and findings aggregated into themes.

10.6.4 Sensitivity and other planned analyses

Subgroup analyses for the completeness of structured data recording will be carried out by clinical specialty and seniority of doctor. Correlations between clinician characteristics in the pre-MiADE survey and the difference in mean number of structured entries with MiADE for outpatients will be explored.

Secondary analyses will be carried out to compare recording of structured data between the following time periods:

1. Before consent versus before MiADE. This will assess the Hawthorne effect, i.e. if there is any change in clinician behaviour once they know they are in the study. This will provide useful information for future interventions to improve structured data recording.
2. Before consent versus after MiADE. This will assess the overall effect of the complex intervention (i.e. the whole programme of training, institutional buy-in for quality improvement, as well as the tool to assist clinicians).

There are no relevant stopping criteria.

11.ASSESSMENT AND MANAGEMENT OF RISK

The table below summarises the risks and mitigations of all interventions above standard care that are being performed:

Intervention	Potential risk	Risk Management
Entering clinical data using MiADE	MiADE may return an inaccurate suggestion (possible)	The clinician has to actively review a suggestion in order to add it to the record, so there is minimal risk that an incorrect entry will be added to the patient record.
Entering clinical data using MiADE	Clinicians may not use the system appropriately (possible)	Training will be provided and access will be limited to clinicians who have completed training.
Entering clinical data	The MiADE system may malfunction	The MiADE system is separate from Epic and works in the

using MiADE	(unlikely)	background, so users can continue with their Epic work regardless of the activity of MiADE. If Epic does not receive a reply from MiADE in a reasonable time it will timeout the request. If the MiADE system is completely down, clinicians will be able to use Epic just as before, without MiADE.
Entering clinical data using MiADE	The MiADE system may perform poorly and be a hindrance to clinicians (unlikely)	Clinicians can choose to ignore the recommendations from MiADE or switch the system off at any time.

12.RECORDING AND REPORTING OF ADVERSE EVENTS

12.1 Definitions

Term	**Definition**
Adverse Event (AE)	Any untoward medical occurrence in a patient or trial participant, which does not necessarily have a causal relationship with the intervention involved.
Serious Adverse Event (SAE).	Any adverse event that: • results in death, • is life-threatening*, • requires hospitalisation or prolongation of existing hospitalisation**, • results in persistent or significant disability or incapacity, or • consists of a congenital anomaly or birth defect. • Other 'important medical events' may also be considered serious if they jeopardise the participant or require an intervention to prevent one of the above consequences
* A life- threatening event, this refers to an event in which the participant was at risk of death at the time of the event; it does not refer to an event which hypothetically might have caused death if it were more severe. ** Hospitalisation is defined as an in-patient admission, regardless of length of stay. Hospitalisation for pre-existing conditions, including elective procedures do not constitute an SAE.	

12.2 Assessments of Adverse Events

Each adverse event (AEs) will be assessed for severity, causality, seriousness and expectedness as described below. Adverse events will be reported and recorded for the period that MiADE is switched on. MiADE does not affect a patient's care or treatment directly so any adverse events that are not related to the electronic health record system do not need to be reported.

12.2.1 Severity

The generic categories below will be used for assessing severity.

Category	**Definition**
Mild	The adverse event does not interfere with the participant's daily routine, and does not require further intervention; it causes slight discomfort
Moderate	The adverse event interferes with some aspects of the participant's routine, or requires further intervention, but is not damaging to health; it causes moderate discomfort
Severe	The adverse event results in alteration, discomfort or disability which is clearly damaging to health

12.2.2 Causality

It is of particular importance in this trial to capture events related to the functioning of electronic systems and recording of clinical information in the electronic health records. The assessment of relationship of an adverse event to this/these additional safety issue(s) will also be carried out as part of the trial.

The differentiated causality assessments will be captured in a trial specific adverse event log.

The following categories will be used to define the causality of the adverse event:

Category	**Definition**
Related	A causal relationship between the intervention and an adverse event is at least a reasonable possibility, i.e., the relationship cannot be ruled out.
Not related	There is no reasonable possibility of a causal relationship between the intervention and an adverse event.
Not Assessable	Unable to assess on information available.

12.2.3 Expectedness

All SAEs assigned by the Investigator or delegate as suspected to be related to the intervention (I.e. related to the function of the electronic health record system for recording structured information) will be assessed for expectedness as defined in this protocol.

Category	Definition
Expected	An adverse event which is consistent with the information about the intervention defined in this protocol.
Unexpected	An adverse event which is not consistent with the information about the intervention defined in this protocol.

** This includes listed events that are more frequently reported or more severe than previously reported.*

The reference document to be used to assess expectedness against the Intervention is this protocol. There are no expected procedural/disease related AEs related to MiADE.

12.2.4 Recording of Adverse Events

All adverse events will be recorded in the medical records in the first instance.

AEs will not be collected in the CRFs for this trial because there are no patient-specific CRFs for the use of the MiADE system. AEs or SAEs will not require reporting to the sponsor unless they are related the functioning of the MiADE system, but this is unlikely.

12.3 Procedures for recording and reporting Serious Adverse Events (SAEs)

All serious adverse events will be recorded in the medical records.

All SAEs (except those specified in the protocol as not requiring reporting to the Sponsor) will be reported to the Sponsor within 24 hours of becoming aware. The CI/PI or designated individual will complete the Sponsor's online Research Incident Reporting Form (https://redcap.slms.ucl.ac.uk/surveys/?s=NE5dypTdFo) within 24 hours of becoming aware of the event. The Chief or Principal Investigator will respond to any SAE queries raised by the Sponsor as soon as possible.

12.4 Managing serious adverse events across research sites

SAEs will be reported to the Sponsor until the end of the trial.

Follow-up SAE reports (clearly marked as follow-up) should be completed via https://redcap.slms.ucl.ac.uk/surveys/?s=NE5dypTdFo and submitted to the JRO as further information becomes available.

12.5 Serious Adverse Events (SAEs) that do not require reporting

SAEs will not require reporting if they do not relate to the functioning of the MiADE system for recording structured information in electronic health records.

12.6 Incidental Findings in Research

Incidental fundings are not applicable to this study, as it concerns only the documentation of information that is already available in the course of usual care.

12.7 Unblinding (if applicable)

Unblinding is not relevant to this study.

12.8 Reporting Urgent Safety Measures

If any urgent safety measures are taken the CI/ PI shall immediately and in any event no later than 3 days from the date the measures are taken, give written notice in the form of a substantial amendment to the relevant REC and Sponsor of the measures taken and the circumstances giving rise to those measures.

12.9 Protocol Deviations and Violations

The Sponsor will be notified immediately of any protocol violations during the trial conduct phase by completion of the online JRO Research Incident Reporting Form: https://redcap.slms.ucl.ac.uk/surveys/?s=NE5dypTdFo. All protocol violations must be recorded on the Protocol Violation Log and filed in the site file.

Protocol deviations are **minor** unintended departures from the expected conduct of the study protocol/SOPs, which **do not impact** participants' safety or compromises the integrity of the study data. For example, an observed consultation being outside the time window defined in the protocol. Protocol deviations do not need to be reported to the Sponsor, but should be recorded in the Protocol Deviation Log and filed in the site file.

12.10 Reporting incidents involving a medical device

Any adverse incident involving a medical device will be reported to the manufacturer of the device.

All adverse incidents must be reported to the study team (contact uclh.miade@nhs.net or anoop.shah3@nhs.net). Incidents should be reported as soon as possible (usually within 24 hours).

Local trust reporting procedures for medical device events will also need to be followed. It is the responsibility of the PI and study site team to ensure they are aware of any specific local requirements for reporting device incidents.

12.11 NHS Serious Incidents and Near Misses (if applicable)

A serious incident or near miss is any unintended or unexpected event that could have or did lead to harm, loss or damage that contains one or more of the following components:

a. It is an accident or other incident which results in injury or ill health.
b. It is contrary to specified or expected standard of patient care or service.
c. It places patients, staff members, visitors, contractors or members of the public at unnecessary risk.
d. It puts the Trust in an adverse position with potential loss of reputation.
e. It puts Trust property or assets in an adverse position or at risk.

Serious Incidents and near misses will be reported to the Sponsor and Trust Quality & Safety department as soon as the study team becomes aware of them.

12.12 Complaints from research participants

In the first instance, research participant complaints (patients or health volunteers) will be reported to the CI/PI to investigate, as documented in the patient information sheet(s), and to the Sponsor via research-incidents@ucl.ac.uk, following the *UCL Complaints from Research Subjects about UCL Sponsored Studies and Trials* policy. For participants who are NHS patients, complaints will be reported to the NHS Complaints Manager at the Trust where the recruitment and study procedures were undertaken. Complaints from NHS patients are handled under NHS complaints policies and procedures, with involvement from PALS and the Sponsor where necessary.

13.OVERSIGHT COMMITTEES

This trial will have a Trial Management Group (consisting of Chief Investigator and trial staff) and will be overseen by the MiADE project Steering Committee which will function as the Trial Steering Committee (TSC).

13.1 Trial Management Group (TMG)

The TMG will include the Chief Investigator and trial staff. The TMG will be responsible for overseeing the trial. The TMG will review recruitment figures, SAEs and substantial amendments to the protocol prior to submission to the REC.

13.2 Other committees

The project will be overseen by the MiADE project Steering Committee which will function as a Trial Steering Committee (TSC) for this trial. The role of the TSC is to provide overall supervision of the trial. The TSC will recommend any appropriate amendments/actions for the trial as necessary. The TSC acts on behalf of the funder(s) and Sponsor.

An independent data monitoring committee is not required for this trial.

14. REGULATORY REVIEW AND PATIENT AND PUBLIC INVOLVEMENT

14.1 Regulatory Review

The Sponsor will ensure that the trial protocol, participant information sheet, consent form, GP letter and submitted supporting documents have been approved by the appropriate research ethics committee, prior to any participant recruitment. The protocol, all other supporting documents including and agreed amendments, will be documented and submitted for ethical and regulatory approval as required. Amendments will not be implemented prior to receipt of the required approval(s).

The study was deemed to require regulatory approval from the following bodies: NHS REC and HRA. Before any site can enrol patients into the study, the Chief Investigator/Principal Investigator or designee will ensure that the appropriate regulatory approvals have been issued, and NHS Confirmations of Capacity and Capability and Sponsor green lights are in place.

For any amendments to the study, the Chief Investigator or designee, in agreement with the Sponsor, will submit information to the appropriate body in order for them to issue approval for the amendment. The Chief Investigator or designee will work with sites (R&D departments as well as the study delivery team) to confirm ongoing Capacity and Capability for the study.

All correspondence with the Sponsor, REC and HRA will be retained. The Chief Investigator will notify the Sponsor and REC of the end of the study.

It is the Chief Investigator's responsibility to produce the annual progress reports when required; an annual progress report (APR) will be submitted to the Sponsor and REC within 30 days of the anniversary date on which the favourable opinion was issued, and annually until the study is declared ended.

Within 90 days after the end of the trial, the CI will ensure that the main REC is notified that the trial has finished. If the trial is terminated prematurely, those reports will be made within 15 days after the end of the trial.

Within one year after the end of the study, the Chief Investigator will submit a final report with the results, including any publications/abstracts, to the Sponsor and to the REC and HRA.

14.2 Peer Review

The study has been peer reviewed in accordance with the requirements outlined by UCL.

14.3 Patient and public involvement (PPI)

The project Steering Committee includes three lay members who have provided input into the design of the study, have reviewed study materials such as the patient information form and will assist with the dissemination of findings.

15. MONITORING AND AUDITING

A trial specific oversight and monitoring plan will be established for studies. The trial will be monitored in accordance with the agreed plan. The degree of monitoring will be proportionate to the risks associated with the trial. Risk will be assessed on an ongoing basis by the Chief Investigator, and adjustments made accordingly (in conjunction with the Sponsor).

The Chief Investigator will be responsible for the day to day monitoring and management of the study. The Chief Investigator will ensure there are adequate quality and number of monitoring activities conducted by the study team. This will include adherence to the protocol, procedures for consenting and ensure adequate data quality.

The Chief Investigator will inform the Sponsor should he/she have concerns which have arisen from monitoring activities, and/or if there are problems with oversight/monitoring procedures.

The UCLH/UCL Joint Research Office, on behalf of UCL as Sponsor, will conduct random audits on a selection of studies in its clinical research portfolio. Monitoring and auditing will be conducted in accordance with the UK Policy Framework for Health and Social Care Research, and in accordance with the Sponsor's monitoring and audit policies and procedures.

16. TRAINING

The Chief Investigator will review and provide assurances of the training and experience of all staff working on this study. Appropriate training records will be maintained in the study files.

17. INSURANCE AND INDEMNITY

University College London holds insurance against claims from participants for harm caused by their participation in this clinical study. Participants may be able to claim compensation if they can prove that UCL has been negligent. However, as this clinical study is being carried out in a hospital, the hospital continues to have a duty of care to the participant of the clinical study. University College London does not accept liability for any breach in the hospital's duty of care, or any negligence on the part of hospital employees. This applies whether the hospital is an NHS Trust or otherwise.

Participants may also be able to claim compensation for injury caused by participation in this clinical study without the need to prove negligence on the part of University College London or another party. Participants who sustain injury and wish to make a claim for compensation should be advised to do so in writing in the first instance to the Chief Investigator, who will pass the claim to the Sponsor's Insurers, via the Sponsor's office.

Hospitals selected to participate in this clinical study shall provide clinical negligence insurance cover for harm caused by their employees and a copy of the relevant insurance policy or summary shall be provided to University College London upon request.

18.RECORD KEEPING AND ARCHIVING

UCL and each participating site recognise that there is an obligation to archive study-related documents at the end of the study (as such end is defined within this protocol). The Chief Investigator confirms that he/she will archive the Trial Master File at UCLH for the period stipulated in the protocol and in line with all relevant legal and statutory requirements. The Principal Investigator at each participating site agrees to archive his/her respective site's study documents in line with all relevant legal and statutory requirements. Study documents will be archived for a minimum of 5 years from the study end, and no longer than 20 years from the study end.

The Trial Master File will be archived at UCL, in accordance with the UCL Retentions Schedule. It will be archived for a minimum of 5 years from the study end, and no longer than 20 years from study end.

19. INTELLECTUAL PROPERTY

All background intellectual property rights (including licences) and know-how used in connection with the study shall remain the property of the party introducing the same and the exercise of such rights for purposes of the study shall not infringe any third party's rights.

All intellectual property rights and know-how in the protocol, the study data and in the results arising directly from the study, but excluding all improvements thereto or clinical procedures developed or used independently of the study by each participating site, shall belong to UCL. All intellectual property rights deriving or arising from the material or any derivations of the material provided to UCL by the participating site shall belong to UCL. Each participating site agrees that by giving approval to conduct the study at its respective site, it agrees hereby to effectively assign all such intellectual property rights ("IPR") to UCL and to disclose all such know-how to UCL.

Each participating site agrees to, at the request and expense of UCL, execute all such documents and do all acts necessary to fully vest the IPR in UCL.

Nothing in this section shall be construed so as to prevent or hinder the participating site from using know-how gained during the performance of the study in the furtherance of its normal activities of providing or commissioning clinical services, teaching and research to the extent that such use does not result in the disclosure or misuse of confidential information or the infringement of an intellectual property right of UCL or its funder. This does not permit the disclosure of any of the results of the study, all of which remain confidential.

20. PUBLICATION AND DISSEMINATION

UCLH will own the data arising from the study. On completion of the study, data will be analysed and a final study report prepared. The findings will be published in peer-reviewed scientific journals and may also be presented at conferences. The funder (NIHR) will be acknowledged in all publications arising from this study. Study participants will be given the option to be contacted about the results of the study and to participate in future involvement activities. The study protocol and statistical analysis scripts will be made available on request to the principal investigator. Individual patient data will not be made available publicly. Anonymised survey findings from clinicians will be made publicly available.

All proposed publications will be discussed with and reviewed by the Sponsor prior to publishing other than those presented at scientific forums/meetings. Resulting publications and/or abstracts will be emailed to the JRO.

21. REFERENCES

1. NHS. NHS Long Term Plan. 2019 Jul. Available: https://www.longtermplan.nhs.uk

2. Ashley EA. Towards precision medicine. Nat Rev Genet. 2016;17: 507–522. doi:10.1038/nrg.2016.86

3. Department of Health and Social Care. Chief Medical Officer annual report 2016: generation genome. GOV.UK; 2017 Jul. Available: https://www.gov.uk/government/publications/chief-medical-officer-annual-report-2016-generation-genome

4. Weed LL. Medical records that guide and teach. N Engl J Med. 1968;278: 593–600. doi:10.1056/NEJM196803142781105

5. McConnell H. International efforts in implementing national health information infrastructure and electronic health records. World Hosp Health Serv. 2004;40: 33–40. Available: https://www.ncbi.nlm.nih.gov/pubmed/15114922

6. Kalra D, Fernando B. Approaches to enhancing the validity of coded data in electronic medical records. Prim Care Respir J. 2011;20: 4–5. doi:10.4104/pcrj.2010.00078

7. Shah AD, Quinn NJ, Chaudhry A, Sullivan R, Costello J, O'Riordan D, et al. Recording problems and diagnoses in clinical care: developing guidance for healthcare professionals and system designers. BMJ Health Care Inform. 2019;26. doi:10.1136/bmjhci-2019-100106

8. Theadom A, de Lusignan S, Wilson E, Chan T. Using three-channel video to evaluate the impact of the use of the computer on the patient-centredness of the general practice consultation. Inform Prim Care. 2003;11: 149–156. doi:10.14236/jhi.v11i3.563

9. openEHR: An open domain-driven platform for developing flexible e-health systems. Available: http://www.openehr.org/

10. NHS Digital. Introduction to Care Connect API. [cited 29 Aug 2019]. Available: https://nhsconnect.github.io/CareConnectAPI/

11. INTEROpen. Introduction to CareConnect API. Available: http://www.interopen.org/

12. Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on natural language processing. J Am Med Inform Assoc. 2004;11: 392–402. doi:10.1197/jamia.M1552

12. Meystre S, Savova G, Kipper-Schuler K, Hurdle J. Extracting Information from Textual Documents in the Electronic Health Record: a Review of Recent Research. Yearb Med Inform. 2008; 128–144.

13. Wu S, Miller T, Masanz J, Coarr M, Halgrim S, Carrell D, et al. Negation's not solved: generalizability versus optimizability in clinical natural language processing. PLoS One. 2014;9: e112774. doi:10.1371/journal.pone.0112774

14. NHS Connecting for Health. Design Guide Entry - Terminology - Matching. Health and Social Care Information Centre; 2007 Mar. Available: https://webarchive.nationalarchives.gov.uk/20160921150545/http://systems.digital.nhs.uk/data/cui/uig

15. Poulos J, Zhu L, Shah AD. Data gaps in electronic health record (EHR) systems: An audit of problem list completeness during the COVID-19 pandemic. Int J Med Inform. 2021 Jun;150:104452. doi: 10.1016/j.ijmedinf.2021.104452. Epub 2021 Apr 1. PMID: 33864979; PMCID: PMC9759969.

16. Kraljevic Z, Searle T, Shek A, Roguski L, Noor K, Bean D, et al. Multi-domain clinical natural language processing with MedCAT: The Medical Concept Annotation Toolkit. Artif Intell Med. 2021 Jul;117:102083. doi: 10.1016/j.artmed.2021.102083. Epub 2021 May 1. PMID: 34127232

17. Fugard and Potts, Supporting thinking on sample sizes for thematic analyses: a quantitative tool. International Journal of Social Research Methodology, 18:6, 669-684, DOI: 10.1080/13645579.2015.1005453

21.1 APPENDICES

21.2 APPENDIX 1: Schedule of Assessments

Interventions for clinicians

Visit No:	1	2	3	4
Window of flexibility for timing of visits:			At least 4 weeks after consent	At least 2 weeks after MiADE switch on
Informed Consent	X			
Training	X			
Pre-MiADE survey		X		
Pre-MiADE observed consultation (for clinicians in qualitative substudy)		X		
MiADE switch on			X	
Post-MiADE consultation and interview (for clinicians in qualitative substudy)				X
Post-MiADE survey				X

Interventions for patients

Visit No:	1	2	3
Window of flexibility for timing of visits:		Ideally 24 hours after information giving, but can be same day	Ideally on same day or next day
Information giving	X		
Informed consent		X	
Observed consultation		X	
Semi-structured interview			X

21.3 APPENDIX 2: Associated Documents

Document Name	Version	Document Date
Clinician information leaflet	2	24/07/2023
Patient information leaflet	2	24/07/2023
Clinician consent form	2	24/07/2023
Patient consent form	2	24/07/2023