AI Clinician XP2

AI Clinician XP2 - A pilot study of the AI Clinician running in real time in the ICU

Version 1.1 9th November 2023

MAIN SPONSOR: Imperial College London

FUNDERS: NIHR/NHS-X

IRAS Project ID: 321582

REC reference: N/A

**Protocol authorised by:**

|  |  |  |
| --- | --- | --- |
| **Name & Role** | **Date** | **Signature** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**Study Management Group**

Chief Investigator: Matthieu Komorowski

Co-investigators: Aldo Faisal, Anthony Gordon

Study Management: Elizabeth Fagbodun

**Clinical Queries**

Clinical queries should be directed to Matthieu Komorowski who will direct the query to the appropriate person

**Sponsor**

Imperial College London is the main research Sponsor for this study. For further information regarding the sponsorship conditions, please contact the Head of Regulatory Compliance at:

Research Governance and Integrity Team

Imperial College London and Imperial College Healthcare NHS Trust

Room 215, Level 2, Medical School Building

Norfolk Place

London, W2 1PG

**Tel**: **0207 594 1862**

[Imperial College - Research Governance and Integrity Team (RGIT) Website](https://www.imperial.ac.uk/research-and-innovation/research-office/research-governance-and-integrity/)

**Funder**

NIHR/NHS-X

This protocol describes the AI Clinician study XP2 – A pilot study of the AI Clinician running in real time in the ICU and provides information about procedures for entering participants. Every care was taken in its drafting, but corrections or amendments may be necessary. These will be circulated to investigators in the study. Problems relating to this study should be referred, in the first instance, to the Chief Investigator.

This study will adhere to the principles outlined in the UK Policy Frame Work for Health and Social Care Research. It will be conducted in compliance with the protocol, the Data Protection Act and other regulatory requirements as appropriate.

**Table of Contents Page No**

[1. INTRODUCTION 7](#_Toc33098103)

[1.1. BACKGROUND 7](#_Toc33098104)

[1.2. RATIONALE FOR CURRENT STUDY 7](#_Toc33098105)

[2. STUDY OBJECTIVES 7](#_Toc33098106)

[3. STUDY DESIGN 7](#_Toc33098107)

[3.1. STUDY OUTCOME MEASURES 7](#_Toc33098108)

[4. PARTICIPANT ENTRY 7](#_Toc33098109)

[4.1. PRE-REGISTRATION EVALUATIONS 7](#_Toc33098110)

[4.2. INCLUSION CRITERIA 7](#_Toc33098111)

[4.3. EXCLUSION CRITERIA 8](#_Toc33098112)

[4.4. WITHDRAWAL CRITERIA 8](#_Toc33098113)

[5. ADVERSE EVENTS 8](#_Toc33098114)

[5.1. DEFINITIONS 8](#_Toc33098115)

[5.2. REPORTING PROCEDURES 8](#_Toc33098116)

[6. ASSESSMENT AND FOLLOW-UP 9](#_Toc33098117)

[7. STATISTICS AND DATA ANALYSIS 9](#_Toc33098118)

[8. REGULATORY ISSUES 9](#_Toc33098119)

[8.1. ETHICS APPROVAL 9](#_Toc33098120)

[8.2. CONSENT 9](#_Toc33098121)

[8.3. CONFIDENTIALITY 10](#_Toc33098122)

[8.4. INDEMNITY 10](#_Toc33098123)

[8.5. SPONSOR 10](#_Toc33098124)

[8.6. FUNDING 10](#_Toc33098125)

[8.7. AUDITS 10](#_Toc33098126)

[9. STUDY MANAGEMENT 11](#_Toc33098127)

[10. PUBLICATION POLICY 11](#_Toc33098128)

[11. REFERENCES 11](#_Toc33098129)

**Glossary of Abbreviations**

|  |  |
| --- | --- |
| Human evaluators | ICU doctors (any level) assessing the AI in the background, without informing clinical practice |
| ICU | Intensive Care Unit |
| Sepsis-3 | The current international definition of sepsis. Requires the presence of a suspected infection (administration of antibiotics and sampling of bodily fluids for bacteriological culture) and some degree of organ failure (an increase in SOFA score of 2 points or more) |
| SOFA | Sequential Organ Failure Assessment – an organ failure score |
| EHR | Electronic Health Record |
| SQL | Structured Query Language |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**Keywords**

Sepsis, decision support system, artificial intelligence, machine learning, electronic health record

**Study Summary**

|  |  |
| --- | --- |
| **TITLE** | AI Clinician XP2 - A pilot study of the AI Clinician running in real time in the ICU |
| **DESIGN** | Prospective multi-centre observational pilot study in 4 ICUs from 2 different NHS Trusts. |
| **AIMS** | Confirm ability to identify patient participants, test the operational robustness of the AI Clinician software when operating in real-time in a live ICU environment, evaluate clinician’s agreement with AI suggestions, evaluate integration in the routine workflow |
| **OUTCOME MEASURES** | Number of patients identified per site and per week; system / usability data; patient time series data; doses of treatment chosen by clinicians; assessment of AI decisions by clinician evaluators; clinicians participant data. |
| **POPULATION** | Patients: adult patients in intensive care; human clinician evaluators: senior ICU doctors (senior registrars, fellows or consultants) |
| **ELIGIBILITY** | Adult patients with sepsis in intensive care , senior ICU doctors (senior registrars, fellows or consultants) |
| **duration** | 6 months |

# INTRODUCTION

* 1. BACKGROUND

Sepsis is life-threatening organ dysfunction due to severe infection and affects 250,000 patients annually in the UK (pre-COVID-19), of whom 48,000 die. In addition, virtually all COVID-19 intensive care unit (ICU) deaths had sepsis. It is a leading cause of death and the most expensive condition treated in hospitals. It was recognised as a top research priority by the James Lind Alliance, a partnership of patients and clinicians to prioritise the most pressing unanswered questions facing the NHS.

The cornerstone of sepsis resuscitation is the administration of intravenous fluids (IVF) and/or vasopressors (drugs that squeeze the blood vessels to increase blood pressure) to maintain blood flow to prevent organ failure. However, there is huge uncertainty around the individual dosing of these drugs in an individual patient, partially due to high sepsis heterogeneity. The current guidelines provide recommendations at a population-level but fail to individualise the decisions. Wrong decisions lead to poorer outcomes and increased ICU-resource use. A tool to personalise these medications could improve patient survival.

We have developed a new method to automatically and continuously review and recommend the correct dose of these medications to doctors, which was created using artificial intelligence (AI) techniques applied to large medical databases. The method we used is called reinforcement learning, and we call the technology the “AI Clinician”.

In the AI Clinician XP1, we tested the safety of the AI Clinician when running in “shadow mode”, i.e. in pseudonymised batches of patient data presented to off-duty ICU clinicians. This enabled us to 1) develop methods and software to connect to real-time electronic health records (EHR); 2) check the safety of the algorithm when used in a contemporary UK ICU patient cohort.

In XP2, the AI Clinician will be running in real-time on dedicated computers at the bedside of actual patients in 4 ICUs across 2 NHS Trusts (Three ICUs at ICHT and one ICU at UCLH).

This present experiment will **test the feasibility of running the AI Clinician in real-time in operational ICUs**, in preparation for a future large scale multicentre randomised trial that will test for an improvement in clinically relevant outcomes. At this stage and in the interest of focusing on prescribers first, we will only be testing the use of the system by ICU doctors. Studies with nurses will be conducted in the future.

* 1. RATIONALE FOR CURRENT STUDY

Hypotheses

* We can include at least one subject per centre per week, and the system can be used at least twice for each patient included.
* The system is reliable: when triggered, it runs appropriately and produces an output within 2 minutes more than 90% of the time (at least 86 successful uses in 96 planned uses). This duration is automatically timed within the interface.
* The quality of AI suggestions, assessed by on-duty clinicians, is appropriate (we define “quality of AI suggestions” below).
* Patients’ outcomes are better when actual doctors’ decisions match the suggested AI decisions.

# STUDY OBJECTIVES

**Primary objective**

We want to test the **feasibility of running the AI Clinician** in real-time in the ICU of 4 sites across a period of 6 months **and identify patients suitable for the system**.

**Secondary objectives**

1) We will assess the **robustness of the model predictions** by collecting the actual doses of IVF and vasopressors given to patients and compare to the doses recommended by the AI. We will also collect pseudonymised patient data time series and a number of outcomes: daily organ function, ICU and hospital length of stay, and ICU and in-hospital mortality. This will allow us to assess ability of the AI model to predict state progression.

2) Each time the tool is used, **feedback on the output of the AI in real-time** will be collected from participating clinicians in real-time through a unique interface: [before revealing the AI suggestions] 1) what is your prescribed dose of IVF and vasopressors; 2) how confident are you in your judgement (on a scale of 1 to 10); [at this point, the AI suggestions are revealed] 3) agreement/disagreement with the AI-suggested dose of IV fluids (5-point scale: “likely too low”, “possibly too low”, “likely safe”, “possibly too high”, “likely too high”); 4) agreement/disagreement with the AI-suggested dose of vasopressor (on a 5-point scale); 6) will you modify your prescription based on the AI suggestion? (yes/no); 7) how confident are you in your judgement after seeing the AI suggestion (scale of 1 to 10)?; 8) would you intervene if the AI dose was to be administered automatically? (yes/no); 9) text box for free text comments.

3) We will also gather **qualitative feedback from users** on the tool. At the end of the study, we will **interview** 2 participating clinicians from each centre (8 people in total), to gather their thoughts on the user interface, workflow integration (when is it likely that they will use the system?), trust in the system, and suggestions for improvement.

4) Confirm **operational robustness and technical reliability of the software** when operating in real-time in operational ICU environments. We will measure indicators of system reliability and availability (e.g. how many times does it produce an output when triggered? What percentage of time is the system available, at different times during the day and the lag of the system (e.g. how long it takes to produce an AI output, when the system is triggered. This can be measured automatically with a timer built in the interface, so it doesn’t need manual timing by the user.). We will also automatically collect usage data (date and time of system use), to get information about when users are likely to use the system.

# STUDY DESIGN

This is a **prospective observational pilot study in 4 ICUs across 2 NHS Trusts** (ICHT and UCLH).

The study will be pseudonymised, following processes we developed during XP1. The output of the AI will be presented at the bedside on a dedicated research laptop to on-duty clinicians who will be asked to review and rate the AI suggestions. As such, we will not be seeking patient consent (see section “consent”). No patient personal or identifying data will be collected as such no identifiers will leave the NHS.

There will be **no randomisation** (between standard of care and care supported by the AI) and **no power calculation** in the study.

Over a 6-month period, across 4 sites, we have a target of 48 patients. Including one subject per centre per week would be 12 subjects in 3 months, which equals to 48 subjects across all 4 sites. The study period will be 6 months to allow some space for drop out and capacity issues.

At the beginning of the study period (during setup), 3 to 4 doctors per site will be recruited into the study: 1 consultant and 2 to 3 registrars, for a total of up to 16 clinician participants. Fully informed written consent will be taken from these clinicians. We will collect the following information from each participant: gender, job title, years of ICU experience.

**Patient identification**

Once the site is activated, daily, on each site, a local clinical research nurse screens all patients present on ICU for sepsis, based on sepsis-3 criteria, and all other inclusion/exclusion criteria. They will confirm the suitability for inclusion with the local PI or study lead. Once included in the study, the timing of estimated sepsis onset is recorded.

**Actual system use**

The system use is optional and can be **triggered manually** by 3 events: during the routine ward round (morning/afternoon, the doctor at the bedside notifies the local clinical research nurse), in case of haemodynamic instability (for example in case of hypotension: MAP < 65 mmHg, the bedside nurse would in this case notify the research nurse) or on demand by clinician participants (e.g. in case of uncertainty with regards to optimal course of action, they would notify the research nurse).

When triggered, the research laptop is brought at the bedside by the research nurse, and the system is launched. The local clinical research nurse will connect the system up and notify the bedside doctor. Participating clinicians only see pseudonymised patient data on the laptop. We record each time the reason for using the system (one of 4 options: “routine ward round”, “haemodynamic instability”, “on clinician request”, “other” – with the option to enter free-text).

Our custom made webapp connects to a pseudonymised SQL copy of the patient EHR database. Patient data from admission until the current time point is extracted (using an existing SQL query). Data is automatically pre-processed (using processes developed during XP1): the total fluid volume, total urine output and cumulated fluid balance since ICU admission are computed; a carry-forward for the previous 36 hours is conducted to impute missing values (important for vitals, labs and ABGs); units are converted to the ones expected by the algorithm (matching MIMIC-III). Time series corresponding to the previous 12h (if available, otherwise we use the longest time period of data available) is fed into the algorithm. The time series data is also presented to the clinician in graphical format (plots) for sanity checks (for example, the absence of important parameters such as urine output may cause high uncertainty in the model output). The output of the AI clinician (a recommended dose of IVF in 25 mL increments, and a recommended dose of vasopressors in mcg/kg/min) is presented to the participating clinician.

Feedback on the output of the AI is collected from participating clinicians:  [before revealing the AI suggestions] 1) what is your prescribed dose of IVF and vasopressors; 2) how confident are you in your decision (on a scale of 1 to 10); [at this point, the AI suggestions are revealed] 3) agreement/disagreement with the AI-suggested dose of IVF (5-point scale: “likely too low”, “possibly too low”, “likely safe”, “possibly too high”, “likely too high”); 4) agreement/disagreement with the AI-suggested dose of vasopressor (on a 5-point scale); 6) will you modify your prescription based on the AI suggestion? (yes/no); 7) how confident are you in your decision after seeing the AI suggestion (scale of 1 to 10)?; 8) would you intervene if the AI dose was to be administered automatically? (yes/no); 9) text box for free text comments. This should take around 2 minutes.

Patients will stay in the study for up to 48h after estimated sepsis onset, to match the period used during training of the AI Clinician model. Their indicators of organ function are recorded hourly for 48 hours after each system use.

Each time the system is used, we keep track of user and pseudonymised patient participant, date and time, as well as the reason for using the system. We also record patient’s data as described in the section “data collected/study outcome measures”: demographics and time series of data from ICU admission to current time of evaluation.

For each included patient, the research nurse will also collect ICU and in-hospital mortality within 30 days of inclusion. If a patient dies, we will keep and use their pseudonymised data.

Each clinician will be required to use the system a minimum of 4 times across the study period.

* 1. STUDY OUTCOME MEASURES

|  |  |
| --- | --- |
| **Data of interest** | **Study outcome measures** |
| Patient identification | Number of subjects identified and presented to a bedside doctor each week in each centre  Number of times the system is used for each patient |
| System/ usability data | Date and time of each system usage (each time it is used by clinicians).  Data availability: what percentage of essential and optional data fields are available 24/7.  System availability: success/failure of generating a response. Delay in generating response when the system is triggered.  Number and nature of technical issues (drop-outs, freezes). An independent online form (survey-type) will be created to log all technical issues that the users may encounter (e.g., system unavailable, login issues etc). This survey will be kept on the same research laptop, but separate from the AI Clinician application, so it can’t be affected by server outage for example.  Server status, down-time events, planned and unplanned outages. These events can be monitored remotely and logged by the ICT team. |
| Anonymised patients’ data | Patient demographics (age in years, gender, primary diagnosis)  From ICU admission to current time of evaluation:   * Vital signs and lab values including arterial blood gases * Doses received of intravenous fluids and vasopressors * Urine output and fluid balance   Presence of sedation, mechanical ventilation, dialysis (binary)  Outcomes: organ function (hourly SOFA for up to 48 hours after the decision time), ICU and hospital mortality |
| Clinician participants data | At the start of the study:   * Gender, grade and seniority (years of ICU experience)   At each evaluation of the AI, the clinician will be asked:  In your assessment:   * Is the patient in sepsis? In septic shock? * Is the patient likely hypovolaemic or likely to be fluid-responsive? * Is the patient vasoplegic? * Does the patient need IVF? How much? * Does the patient need vasopressors? How much?   Consider the AI recommendations:   * Do you agree with the AI recommendation for IVF / vaso? (likert scale 1-5) * Is the AI recommendation too low, just right, too high (both drugs)? * Will you change your prescription based on the AI suggestion, and why? (no because they’re already on a correct dose or no because the AI appears unsafe ; yes meaning I believe the AI is correct and should be followed) * If the AI was to administer the treatment automatically, would you intervene to stop it? * Is the AI suggestion useful? * Is the tool easy to use? * Do you need an explanation for the AI suggestion? Would you prefer if the AI provided an explanation such as “give fluids to reduce the risk of AKI at H48 by 10%”… * Do you have any other comment or notes about the clinical context (please do not include personal, identifiable information)?   At the end of the study, for 2 participants per centre:   * Qualitative interviews (with audio recording, for transcription +/- thematic analysis) |

# PARTICIPANT ENTRY

* 1. INCLUSION CRITERIA

**For patients:**

* Adult > 18yr
* Admitted to an ICU in a participating centre
* With early (within 24 of onset) sepsis (as defined by the sepsis-3 definition)
* For full escalation (no ceiling of care, e.g. patient “not for vasopressors”)
* Expected to survive more than 24hr
* Has not opted-out for use of their data for research (NHS and NHS-X website)

**For clinician participants:**

* ICU doctors at the senior registrar, ICU fellow or consultant level
  1. EXCLUSION CRITERIA

**For patients:**

Not for full active care, e.g. not for vasopressors

Not expected to survive more than 24hr

Elective surgical admission (these patients are regularly on antibiotics but given as a prophylaxis, with no sepsis)

Opted-out for use of their data for research (NHS and NHS-X website)

**For clinician participants:**

Declined participation

* 1. WITHDRAWAL CRITERIA

**For clinician participants:**

Requested withdrawal from study. Clinicians may request withdrawal by contacting the study coordinators. In the event that a clinician requests withdrawal from the study, no further data will be collected from them, however all data collected will be used.

# ADVERSE EVENTS

This is an observational study, so there is no direct risk for patients or the human evaluators.

# 

# ASSESSMENT AND FOLLOW-UP

For each included patient, the research nurse will also collect ICU and in-hospital mortality within 30 days of inclusion.

The study ends 30 days after inclusion on the system.

# STATISTICS AND DATA ANALYSIS

Data and all appropriate documentation will be stored for a minimum of 10 years after the completion of the study, including the follow-up period.

# REGULATORY ISSUES

* 1. ETHICS APPROVAL

The Study Coordination Centre has obtained approval from the xxx Research Ethics Committee (REC) and Health Research Authority (HRA). The study must also receive confirmation of capacity and capability from each participating NHS Trust before accepting participants into the study or any research activity is carried out. The study will be conducted in accordance with the recommendations for physicians involved in research on human subjects adopted by the 18th World Medical Assembly, Helsinki 1964 and later revisions.

* 1. CONSENT

Consent will be sought from participating clinicians. A full explanation will be provided in the form of a written explanation covering the details of the study, their role and any potential risks of participating. Any remaining questions that participants have after reviewing the written documentation will be answered by the investigator team.

All clinician participants are free to withdraw from the study at any stage upon written request to the investigator team without giving reasons. Their data will remain in the study unless they request for it to be removed.

Consent to enter the study must be sought from each participant only after a full explanation has been given, an information leaflet offered, and time allowed for consideration.  Signed participant consent should be obtained.  The right of the participant to refuse to participate without giving reasons must be respected. All participants are free to withdraw at any time from the protocol treatment without giving reasons and without prejudicing further treatment.

Potential patients will only be identified by local NHS staff at the recruiting site and they will maintain the same duty of confidentiality owed to all patients by medical and nursing staff. All personal identifiable data will be kept within the NHS hospital and routine NHS databases. Pseudonymisation of patient data will be completed by the trusts research informatics team who form a part of the extended care team. Because of all of the above we will not seek specific patient consent for this phase of the study.

* 1. CONFIDENTIALITY

The Chief Investigator will preserve the confidentiality of participants taking part in the study as per the Data Protection Act.

Data will be pseudonymised

The General Data Protection Regulation (2016/679) broadly defines personal data breaches as a security incident that leads to the confidentiality, integrity or availability of personal data being affected. There will be a personal data breach whenever any personal data is lost, destroyed, corrupted or disclosed; if someone accesses the data or passes it on without proper authorisation; or if the data is made unavailable, for example, when it has been encrypted by ransomware, or accidentally lost or destroyed. Personal data breaches will be immediately reported to the Sponsor, the Data Protection Officer, and RGIT. Any breaches will be documented in the TMF/ISFs, and will follow Sponsor/Data Controller reporting processes.

We have put measures in place to limit the risk of data confidentiality breach by following Caldicott principles where applicable. Identifiable data will not leave trust servers. The study will only use pseudo-anonymised data so confidentiality of patients and clinician participants will be maintained as linkage codes will be kept within the trust, however the Principal Investigator will ensure to preserve the confidentiality of participants and fulfil transparency requirements under the General Data Protection Regulation (GDPR).

Data and all appropriate documentation will be stored for a minimum of 10 years after the completion of the study. Private health information such as patient identity, date of birth and date of death are not required. The lookup table mapping new patient identifiers and their initial identity will only be kept within trust systems.

Data will be kept encrypted in a secured lab accessible only (via swipe and pin code access) to members of the Imperial College Critical Care Research lab (under Prof. Anthony Gordon). The computer itself where the data will be kept is password protected, and its hard drives are encrypted.

* 1. INDEMNITY

Imperial College London holds negligent harm and non-negligent harm insurance policies which apply to this study

* 1. SPONSOR

Imperial College London will act as the main Sponsor for this study. Delegated responsibilities will be assigned to the NHS trusts taking part in this study.

* 1. FUNDING

NIHR/NHS-X are funding this study. Clinician participants will be remunerated up to £100 for their time spent on the study.

* 1. AUDITS

The study may be subject to audit by Imperial College London under their remit as sponsor and other regulatory bodies to ensure adherence to GCP and the UK Policy Frame Work for Health and Social Care Research.

# STUDY MANAGEMENT

The day-to-day management of the study will be co-ordinated through Elizabeth Fagbodun.

# PUBLICATION POLICY

The results of the project will be published in a peer-reviewed, high impact, and ideally fully open access journal.

# REFERENCES

<https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-10-1>

" ...***test of the methods and procedures to be used on a larger scale*** if the pilot study demonstrates that the methods and procedures can work" [[2](https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-10-1#ref-CR2)];

"...investigation designed to ***test the feasibility of methods and procedures*** for later use on a large scale or ***to search for possible effects and associations*** that may be worth following up in a subsequent larger study" [[3](https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-10-1#ref-CR3)].

<https://www.nature.com/articles/s41598-019-56927-5#Sec10>

<https://www.frontiersin.org/articles/10.3389/fmed.2021.589197/full>

<https://www.sciencedirect.com/science/article/pii/S0923753419596430>

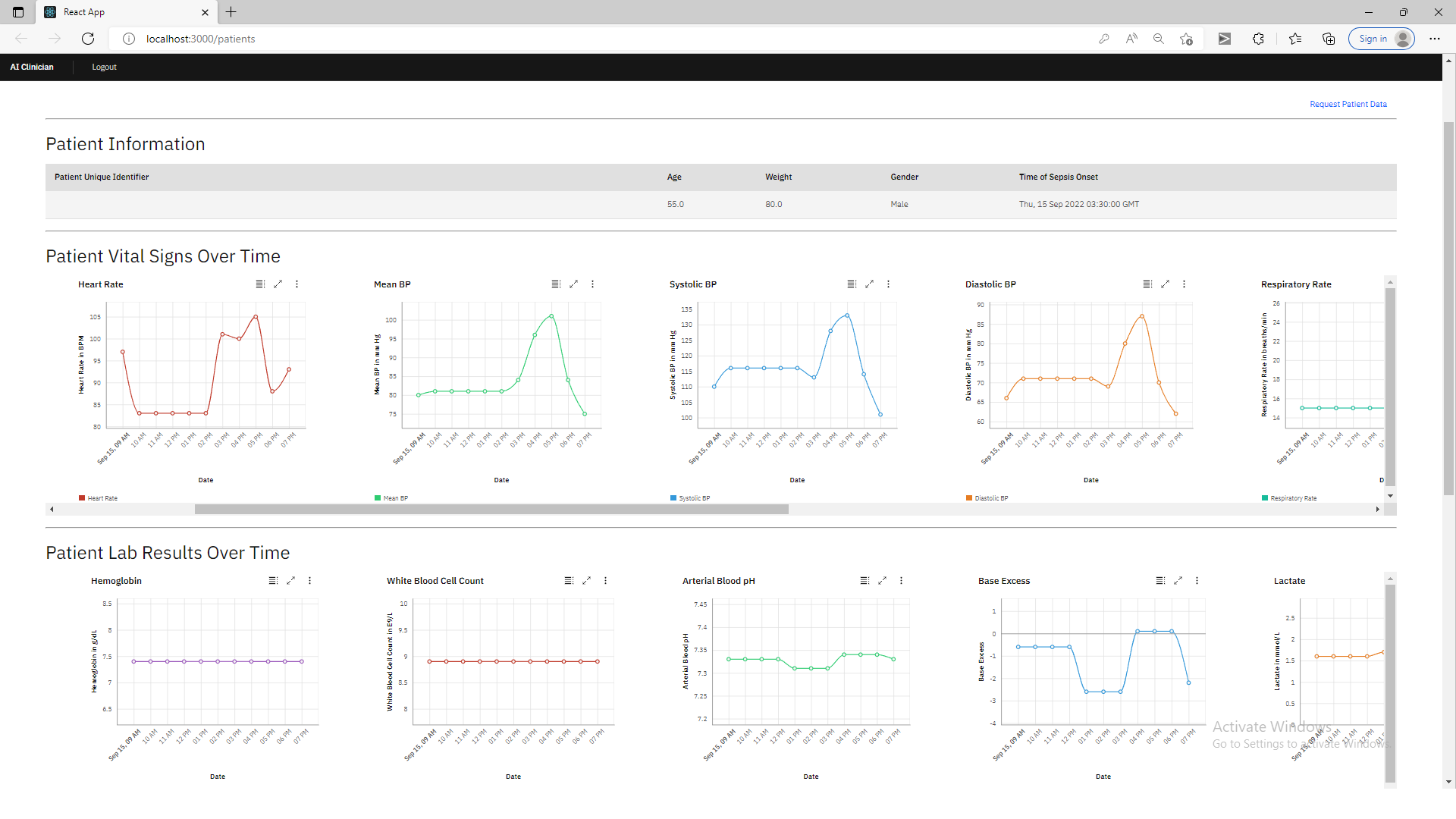
**EXAMPLE APPENDICES**

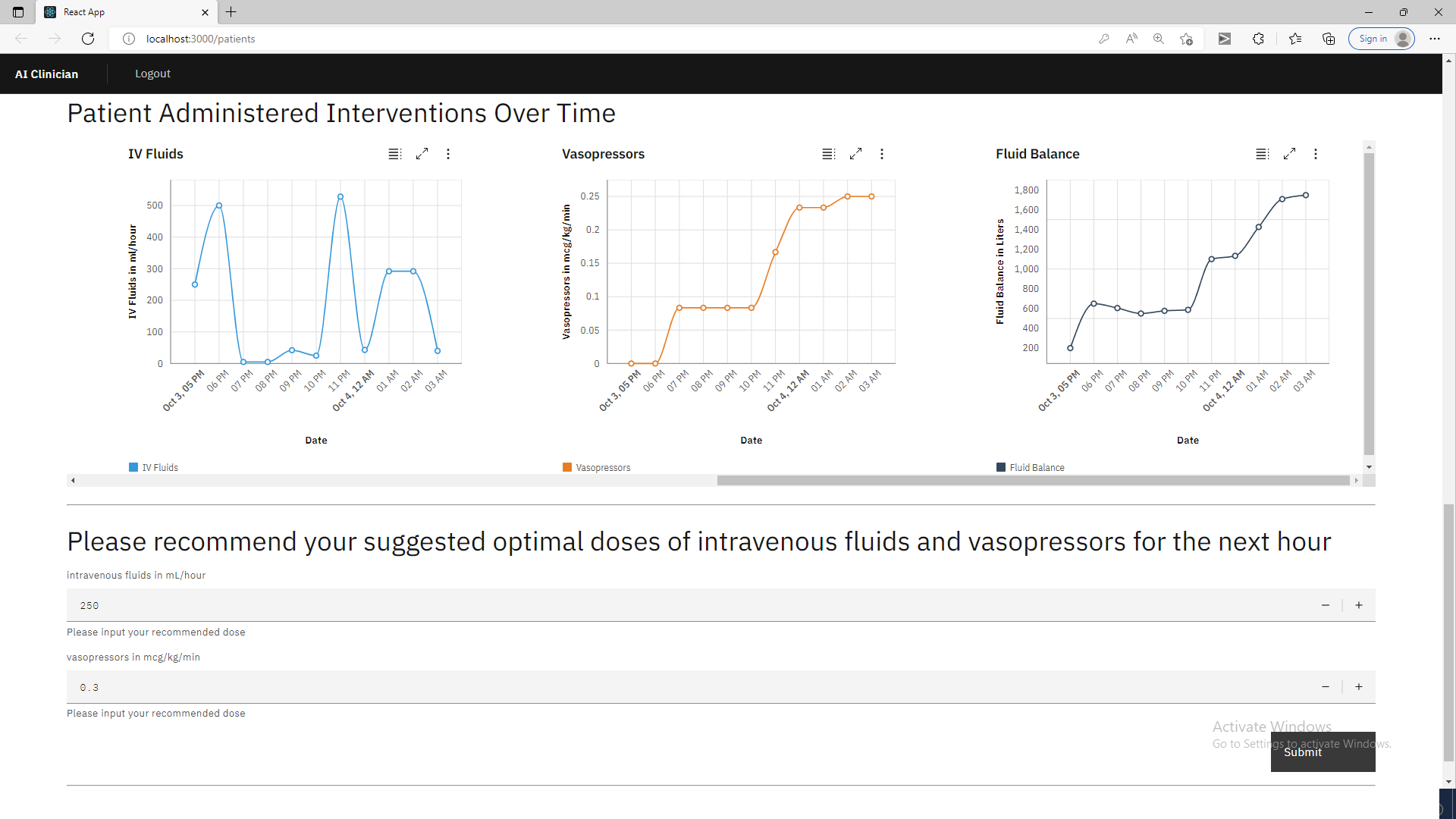
Appendices should be additional information to the protocol and can consist of:

**Appendix 1. Summary of investigations, treatment and assessments**

|  |  |  |
| --- | --- | --- |
| **Investigations / Assessments** | |  |
|  | Screening | At each assessment session |
| System Set-Up |  |  |
| Medical History | X |  |
| Clinical Evaluation | X | X |
| Operational Reliability checks |  | X |
| Informed consent (Clinician) | X |  |

**Appendix 2. Screenshots of the AI Clinician Interface**

Screenshot 1: Patient demographics, and sections showing the time series of vital signs and lab results.

Screenshot 2: Patient time series of interventions, and data entry form for the doses prescribed by the human clinician. 

Screenshot 3: Pop-up window showing the assessment of the AI Clinician recommended dosage.