Introduction & instructions

Protocol template: development and evaluation of AI/ML-based tools for healthcare and medicine

About this document

This is a protocol template for studies developing new or evaluating existing AI/ML-based tools and algorithms in healthcare and medicine. Protocols are living documents that describe the intended conduct of the study. Having a pre-specified protocol is scientific best-practice, and also commonly expected by regulatory authorities, data stewards, ethics boards, or other stakeholders.

This template was created based on past experience designing and conducting evaluations of AI/ML-based tools. It is a work-in-progress – it has not been approved or endorsed by any regulatory bodies and it may contain errors or omissions. With these caveats, I'm making it available "as-is" for anyone interested in using it. I'm hoping it's useful to students and researchers working in the domain.

Feedback or ideas for improvement are very much appreciated *and can be sent to info@kathrynrough.net.*

Instructions for using this document

A good protocol is concise and free of jargon. Protocols provide enough information for the audience to understand the research question and proposed methods. It allows readers to assess the strengths, limitations, and appropriateness of the study design and analyses.

Instructions and guidance to the protocol authors are provided in blue italic text, and should be removed from the finalized protocol. Headings and subheadings that are not relevant to a given study can also be removed.

Template protocol

Title page

Title	*Informative title that communicates the purpose of the algorithm or tool being investigated and the study design*
Protocol version	*Numerical identifier (e.g., version 1.3)*
Date of last protocol update	*DD Month YYYY*
Protocol author/Primary Investigator	*Name* Contact: *physical address and email address*
Additional contributor(s) *(optional)*	*Name(s)*
Study sponsor *(optional)*	Organization: *Name* Contact: *physical address and/or email address*
Study registration *(optional)*	Site: *e.g., clinicaltrials.gov* Identifier: *site-assigned identifier*
Conflict of interest declaration	*Helpful guidance on conflict of interest reporting can be found at: https://www.icmje.org/recommendations/browse/roles-and-responsibilities/author-responsibilities--conflicts-of-interest.html*

Table of Contents

Abstract

The abstract should be a stand-alone summary of the protocol that captures the key components of the study (without referencing the full protocol).

Title: *Informative title that communicates the purpose of the algorithm or tool being investigated and the study design*

Background and rationale: *Context and motivation for conducting study*

Study objectives: *Recommend formatting as a bulleted list and indicating which objectives are primary, secondary, and exploratory. This list should be identical to the objectives listed in the full protocol.*

Data sources: *Description of the data sources. Clearly indicate whether each data source will be used for model training, model evaluation or both.*

Study design and population for model training: *Briefly describe the study population (including all selection criteria) and the type of study design (e.g., randomized trial, cohort study, cross-sectional study) used for model training.*

Outcome label definition: *Define the labels used to train the model*

Model training: *Clearly describe the training task set-up or the approach to fine-tuning an existing model. Briefly characterize the types of data used as features.*

Study design and population for evaluation: *Briefly describe the study population and the type of study design used for evaluation. Mention all comparators that will be evaluated.*

Evaluation metrics and analyses: *Briefly describe the metrics and analyses that will be used to address the study objectives.*

1. Amendments and updates

Protocol version	Date	Sections updated	Description and rationale
e.g., v1.1	*DD Month YYYY*	*List of all sections updated*	*Description of, and brief justification for, updates. Can be helpful to specify if the update occurred before finalization/ registration of protocol.*
…	…	…	…

2. List of abbreviations

Providing a complete catalogue of abbreviations used throughout the protocol is a particularly useful resource when the document will be read and reviewed by multidisciplinary audiences.

Abbreviation	Meaning
e.g., AI	*Artificial intelligence*
…	…

3. Milestones

Planned dates for milestones or deliverables can be indicated in the table below. This section may be required by certain stakeholders; otherwise, it may be omitted.

Milestone	Planned date
E.g. Date of data transfer	*DD Month YYYY*
…	…

4. Background and rationale

Four main elements should be addressed in this section:

- *What problem will the AI/ML-based tool be applied to? Describe a specific health condition (e.g., atrial fibrillation), a broader issue in health care or medical practice (e.g., medication errors), or a specific task (e.g., administrative coding). Briefly provide the reader with adequate information about the problem so they are able to engage with later sections of the protocol. Describe the size/severity/burden of the problem.*
- *What is already known about tools and approaches (AI/ML-based or otherwise) applied to this problem? Provide a brief summary of existing tools and approaches, including any relevant limitations.*
- *What are the current gaps or limitations of existing tools or approaches?*
- *How will this study address the articulated gaps?*

5. Study Objectives

Objectives can be formatted as a bulleted list or table (example below). Indicate which objectives are primary, secondary, and exploratory. The primary objective is the main aim of the study. If sample size or power calculations are performed, they should address the primary objective.

Also clearly communicate the metric(s) or measures of effect used to assess each objective and the population in which the analysis will be performed (e.g., specific subgroups or evaluation datasets).

Objective	**Population(s)**	**Metric(s)**
Primary objective		
Secondary objective(s)		

Exploratory objective(s)		

6. Research methods

6.1. Existing data sources

Describe any data sources that will be used to train, fine-tune, or evaluate the model. If multiple datasets will be used, separate subheadings for each are recommended.

For existing data sources (e.g., an electronic health record database), briefly characterize what information is available in the data source, with references to external documents with more detailed information, when available. Describe which populations are represented in the data source, how long they are followed in the data, the ability of the dataset to accurately capture the intended selection criteria and labels, and any other information on the strengths/weaknesses of the data source relevant to assessing its suitability for the intended research. Include a brief rationale for choosing the data source.

When available, provide information on the provenance of the source data (data provider, version number, date of last ETL, etc.).

6.2. Original data collection

If new data will be collected as part of this study (either from study participants or human raters), describe the selection criteria, what type of data will be collected, the method of collection, the duration of data collection, and the rationale for these choices.

When data will be collected from study participants, provide information on the informed consent process and ability to withdraw from the study.

When data will be collected from human raters, provide details on the qualifications of the raters, the training they received for the study, instructions to the raters, and other relevant information on how their feedback is collected or structured. Adequate detail should be provided so others could replicate the process in future studies.

As appropriate, refer to further study design details in other sections of the protocol.

6.3. Data cleaning and transformation

Describe data preparation and cleaning procedures, including the use of common data models (e.g., OMOP CDM v5.4; FHIR v5.0.0). Summarize any mechanisms or procedures to ensure data quality and integrity. Plans to transform or recategorize the original data, correct inconsistencies or errors, or impute values should be briefly described. Any linkage methods between data sources can be briefly summarized here (with details provided in an Appendix).

6.4. Model development: study design and setting

Describe the overall study design used for training a model (or fine-tuning an existing model) and the rationale for major study design decisions. Consider addressing:

- *Study design: If the study fits one of the classical epidemiologic designs (case control, cohort, randomized trial, cross-sectional, etc.), label it as such.*
- *Setting and population: Who is the target population for the AI/ML-based tool? Who is the population represented in the training data? How well do the training data sources represent that target population? Address topics including geographic location, therapeutic setting, demographics, and duration of follow-up.*
- *Training set selection criteria: Describe the selection criteria (i.e., inclusion and exclusion criteria) for the training data. Provide sufficient detail to allow replication of the selection criteria in future studies; detailed code lists or algorithms can be included as an Appendix.) When the same data will be*

used for training and evaluation, describe the approach to partitioning the data (or cross-validation).

- *Study time periods: For studies with a longitudinal component, clearly define:*
    - *When eligibility criteria will be assessed (i.e., entry to study)*
    - *When the prediction(s) will be made*
    - *Time window for assessment of the features*
    - *Time window for assessing the label/outcome*
    - *When an individual is no longer in the study/under follow-up (it can be helpful to consider events such as death, having the outcome, and other forms of censoring)*

Particularly for studies with a longitudinal component, a study design diagram can be helpful to clearly define study time periods. For an example of study design diagrams, see Scheeweiss et al (2019).

6.5. Model development: Training task and outcome label

Clearly define the training task set-up and labels, as well as the rationale for choosing the label. Sufficient detail should be provided for the label to be replicated. When data will be labeled by raters, be sure to describe the credentials of the raters, any training they received, and the process for data annotation (either in this section or referencing details already provided in Section 6.2). Describe any adjudication procedures. For longitudinal studies, clearly define the window for assessing the label.

If the label has been validated or used in other settings, reference them.

6.6. Model development: Features

Provide a high-level description of the features that may be used in the model (e.g., structured fields of the electronic health records). An exhaustive list of features does not need to be pre-specified. If feature selection, representation, or engineering will be part of the planned iterative model development process, that can be explicitly stated in this section.

Open-ended, non-commital language tends to be helpful in this section. Over-specifying feature-related details in the protocol will not improve the study's scientific rigor but may create a need for future protocol amendments if the modeling approach changes.

6.7. Model development: training and fine-tuning

Briefly describe the iterative model development process (or plans for fine-tuning an existing model). Plans to experiment with different model architectures or ensembling may be mentioned. If a validation set (or cross-validation) will be used to support modeling decision making, briefly describe the validation set. Specify the software, programming languages, or packages that will be used for model training or fine-tuning.

Open-ended, non-commital language also tends to be helpful in this section. Over-specifying details of modeling approaches will not improve the study's scientific rigor but may create a need for future protocol amendments if the modeling approach changes.

6.8. Evaluation: study design

Describe the overall study design used for evaluating the model and rationale for major study design decisions. Consider addressing the points outlined in the instructions in Section 6.4, as well as:

- *Comparators: What are the comparators? Will the model be compared to human performance, an existing tool, and/or previously published algorithms? Provide a brief justification for the choice of comparators.*
- *Evaluation populations and subgroups: Describe all populations and subgroups where evaluations will be performed (which should match the populations listed in the Study Objectives). Clearly define the selection criteria for each subgroup; it may be helpful to summarize this information in a table. If datasets not used in training will be used for evaluation, briefly comment on similarities and differences from the training data and any implications for external validity.*

6.9. Evaluation: metrics

Describe the metrics that will be used to evaluate the performance of the model (these should match the metrics listed in the Study Objectives section). The rationale for the metrics should be briefly described. For complex or non-standard metrics, briefly explain the metric and cite references that provide more detailed methodological descriptions.

6.10. Evaluation: analyses

Briefly describe all planned evaluation-related analyses, including any descriptive analyses that will be performed to characterize the evaluation data or models themselves. If any formal statistical testing will be performed, clearly state the statistical tests that will be used and their corresponding hypotheses. If confidence intervals will be generated, state which methods will be used. Also describe any planned sensitivity or subgroup analyses.

Mention the software, programming languages, or packages that will be used for the evaluation analyses.

6.11. Sample size calculations

Describe any sample size or power calculations here. Clearly state all assumed parameters used in the calculations, with justifications.

7. Data management

Describe the data management procedures used in the study, including infrastructure for data transfer, storage, and backup. Provide information on processes and systems that will be used to maintain appropriate security for any sensitive data.

8. Limitations

Describe limitations of the data sources, study design, and analytic methods. This may include commenting on the representativeness of the target population in the training and evaluation data, potential for errors or mis-measurement in the features, potential for errors or mis-measurement in labels, or gaps between the evaluation tasks and the intended real-world application space. Any steps taken to mitigate a given limitation should also be mentioned in this section.

9. Protection of human subjects

Describe the safeguards taken to comply with legal requirements for ensuring the rights and well-being of participants and relevant data privacy laws. Discuss any ethics committee/human subjects review board processes.

If the study is considered "exempt" by the relevant ethics committee, this should be stated with the reason for exemption.

10. Plans for communicating results

Any plans for submission, communication, or publication of interim or final results can be described in this section.

11. References

Citations for any references or documents referred to on the protocol. Using consistent reference formatting is often more important than the particular reference style.

12. Appendices

Documents listed in the table below can be maintained separately from the protocol.

Number	Document title	Date of last update
e.g., Appendix 1	*Title*	*DD Month YYYY*
...	*...*	*...*