Plan Overview

A Data Management Plan created using DMPonline

Title: PROBAST + AI Delphi Survey

Creator: Constanza Andaur

Principal Investigator: Maarten van Smeden, Karel Moons

Data Manager: Constanza Andaur, Tabea Kaul

Project Administrator: Constanza Andaur, Tabea Kaul

Contributor: Anneke Damen

Affiliation: UMC Utrecht

Template: UMC Utrecht DMP

ORCID iD: 0000-0002-5529-1541

ORCID iD: 0000-0003-2118-004X

Project abstract:

PROBAST was designed to assesses the risk of bias and applicability of prediction models. It supports the interpretation and conclusions about potential risks of bias or distortion (as compared to their development setting) in the predictive performance that may be encountered when the prediction models are to be applied. Recent advances in artificial intelligence require the development of targeted assessment tools. We aim to develop the extension of PROBAST (Prediction model Risk Of Bias ASsessment Tool) for prediction models developed and/or validated using artificial intelligence/machine learning techniques, namely PROBAST-AI. For this, we will set two Delphi rounds to collect experts' opinion and consensus.

Amendments (July 14, 2022):

After reviewing the comments from round 1 of the PROBAST-AI Delphi survey and considering calls for improvement which we have received from key stakeholders since the first launch of PROBAST in 2019, we have decided to extend our research to update and expand the current tool PROBAST. We envision that the new PROBAST will include reworded quality and risk of bias assessment items that apply to studies developing, validating, or updating multivariable diagnostic and prognostic prediction models for individualized risk prediction, as well as an extension to prediction models using AI/ML techniques and specific questions regarding algorithmic fairness. The new tool will make a clearer distinction between the quality assessment of development of prediction models and their evaluation/validation. The working title for this newly developed tool is **PROBAST + AI** . For this, we will set two Delphi rounds to collect experts' opinion and consensus.

ID: 78317

Start date: 01-10-2022

End date: 30-04-2023

Last modified: 22-04-2024

Copyright information:

PROBAST + AI Delphi Survey

1. General features

1.1. Please fill in the table below. When not applicable (yet), please fill in N/A.

DMP template version	
ABR number *(only for human-related research)*	N/A
METC number *(only for human-related research)*	N/A
DEC number *(only for animal-related research)*	N/A
Acronym/short study title	PROBAST + AI Delphi
Name Research Folder	PROBAST_2.0
Name Division	Julius Center
Name Department	Epidemiology
Partner Organization	UMC Utrecht
Start date study	01.10.2022
Planned end date study	30.04.2023
Name of datamanager consulted*	Evelien Kruisselbrink
Check date by datamanager	

1.2 Select the specifics that are applicable for your research.

- Non-WMO

We aim to collect survey data. Specifically, we aim to capture experts' opinion about the suitability of several statements to assess the risk of bias in prediction models developed with machine learning/ artificial intelligence techniques throughout two to three rounds of survey.

Amendments (July 14, 2022):

After receiving comments from round 1 of the PROBAST-AI Delphi survey, we decided to expand our aim to developing a tool, namely PROBAST 2.0, which aims to assess the risk of bias and applicability/ appropriateness of prediction models developed with and without machine learning/artificial intelligence techniques. Specifically, we aim to capture experts' opinion about several revised statements of the 2019 published tool PROBAST, which assess the risk of bias and applicability/appropriateness of prediction models and added several statements which are applicable to prediction models developed with machine learning/artificial intelligence techniques. We aim to collect survey data for this purpose throughout two to three rounds of survey.

Amendments (December 9th, 2022):

The working title was changed from PROBAST 2.0 to PROBAST +AI.

2. Data Collection

2.1 Give a short description of the research data.

We will collect survey data. Our survey aims to capture experts' opinion about the suitability of several statements to assess the risk of bias and applicability in prediction models developed with machine learning/artificial intelligence techniques. The final product is the development of a risk of bias assessment tool, namely PROBAST-AI based on experts' consensus. Unfortunately, there is no existing data available that can be reuse.

There is no long-term value on the data collected, thus it is unlikely the data will be shared and/or preserved. However, there are no restriction for its reuse or sharing with third-parties for verification of study results.

The volume of data will be less than 1 GB. The proportions of raw data is closely to 100 MB, processed data 50 MB, reports 50 MB (including figures, word and power point documents). We consider the scale of the data small, which does not involves challenges for accessing, sharing or transferring data. Applications that will be use from raw data to outputs are provided in Figure 1.

Subjects	Volume	Data Source	Data Capture Tool	File Type	Format	Storage space
Human	200	eCRF	REDCap	Quantitative	.csv	0 - 10 GB

Amendments (July 14,2022):
After receiving comments of the PROBAST-AI Delphi round, we have extended the aim of our survey to now also capture experts' opinions on several improved statements to assess the risk of bias and applicability/appropriateness of prediction models (PROBAST). The updated PROBAST tool will be extended to include statements which assess the risk of bias and applicability/appropriateness of prediction models developed with machine learning/artificial intelligence techniques (AI-extension). The final product is the development of a risk of bias and applicability/appropriateness assessment tool, namely PROBAST 2.0 based on experts' consensus. We will reuse the data (comments) gathered through the first Delphi round of PROBAST-AI.

Amendments (December 9th, 2022):

The working title was changed to PROBAST + AI.

Subjects	Volume	Data Source	Data Capture Tool	File Type	Format	Storage space
Human	200	eCRF	Castor	Quantitative	.csv	0 - 10 GB

2.2 Do you reuse existing data?

- No, please specify
- Yes, please specify

We aim to collect experts' opinion about the suitability of several statements to assess the risk of bias in prediction models developed with machine learning/artificial intelligence techniques. No such tool has been developed previously following the EQUATOR guidelines, thus there is no existing data available that can be reused.
Amendments (July 14,2022):
Yes/No:
After receiving comments of the PROBAST-AI Delphi round 1, we have extended the aim of our survey to capture experts' opinions on several improved statements to assess the risk of bias and applicability/appropriateness of prediction models (PROBAST). The updated PROBAST tool will be extended to include statements to assess the risk of bias and applicability/appropriateness of prediction models developed with machine learning/artificial intelligence techniques (AI-extension).

2.3 Describe who will have access to which data during your study.

During data collection:
The principal investigator will have access to identifying direct personal data because we will send personalised links to answer the survey (linking table: linking responses to participants). The principal investigator will be in charge of denying access to identify survey responses within REDCap (build-in features in REDCap). Therefore, other members of the research team will only have access to pseudoanonymized data to run analysis. Neither the principal investigator nor other members of the research team will have access to delete records of survey responses.

Type of data	Who has access
Direct identifying personal data	Principal investigator, Datamanager
Table linking survey responses to participants	Principal investigator, Datamanager
Pseudonymized data	Research Team

Amendments (July 14, 2022):
As recommended by the Data managers of the Julius Center, UMCU, we will switch the survey from REDCap to Castor for the PROBAST 2.0 Delphi survey.

Amendments (December 9th, 2022):

The working title of the project was changed from PROBAST 2.0 to PROBAST +AI.

2.4 Describe how you will take care of good data quality.

Survey data from experts will be collected in an electronic case report form (eCRF) in the data capture tool: REDCap. In the eCRF, validation checks are built in. For example, the principal investigator can select which variables contain personal data and anonymized them using a build-in feature in REDCap. We contacted the current REDCap administrator in the Julius Center to obtain access to REDCap (Daniel Boateng, d.boateng-2@umcutrecht.nl) and the system administrator to obtain technical support (Johan Smits; j.smits-9@umcutrecht.nl).

Data collection will be frozen before analysis. The data will be checked by another member of the research team authorized by the principal investigator.

#	Question	Yes	No	N/A
1.	Do you use a certified Data Capture Tool or Electronic Lab Notebook?	Amendment (July 14, 2022): yes	x	
2.	Have you built in skips and validation checks?	x		
3.	Do you perform repeated measurements?		x	
4.	Are your devices calibrated?			x
5.	Are your data (partially) checked by others (4 eyes principle)?	x		
6.	Are your data fully up to date?	x		
7.	Do you lock your raw data (frozen dataset)	x		
8.	Do you keep a logging (audit trail) of all changes?	x		
9.	Do you have a policy for handling missing data?	x		
10.	Do you have a policy for handling outliers?	x		

Amendments (July 14,2022):
For the expanded PROBAST 2.0 Delphi, we will change the data capture tool to Castor. Castor is a self-service tool. We contacted the current Castor administrator of the Julius Center to obtain access to Castor (dm_julius@umcutrecht.nl). More specifically, we were in contact with Evelien Kruisselbrink.

Amendments (December 9th, 2022):
The working title was changed from PROBAST 2.0 to PROBAST + AI.

2.5 Specify data management costs and how you plan to cover these costs.

Explanation.
REDCap is available via UMC Utrecht, thus no license fee is required. Storage will be on UMC Utrecht server, thus no extra cost is necessary. Our data has no long-term value, thus we do not plan to archive it in repositories.

#	Type of costs	Division ("overhead")	Funder	Other (specify)
1.	Time of datamanager	x		
2.	Design of eCRF			Principal Investigator
3.	Storage	x		

Amendments (July 14, 2022):
For the expanded PROBAST 2.0 Delphi survey, we will change from REDCap to Castor. Castor is available via UMC Utrecht, thus no license fee is required. Storage will remain on the UMC Utrecht server, thus no extra cost is necessary. Our data has no long-term value, thus we do not plan to archive it in repositories.

Amendments (December 9th, 2022):
The working title was changed from PROBAST 2.0 to PROBAST +AI.

2.6 State how ownership of the data and intellectual property rights (IPR) to the data will be managed, and which agreements will be or are made.

UMC Utrecht is and remains the owner of all collected data for this study. Our data consist on survey responses of experts. IPR are not applicable, thus it does not need additional protection.

3. Personal data (Data Protection Impact Assessment (DPIA) light)

Will you be using personal data (direct or indirect identifying) from the Electronic Patient Dossier (EPD), DNA, body material, images or any other form of personal data?

- Yes, go to next question

I will process direct identifying personal data. I have checked the full DPIA checklist and I do not have to complete a full DPIA. I therefore fill out this DPIA light and proceed to 3.1

3.1 Describe which personal data you are collecting and why you need them.

Which personal data?	Why?
Name and email address of participants	To be able to invite participants for taking part in the Delphi process and to send them the survey and reminders. The invitee to the survey are experts on the topic and therefore, their email are available through open resources (i.e. articles, conferences websites).
Gender, Expertise, Academic background	To describe our study population

3.2 What legal right do you have to process personal data?

- Other, please explain

We extended an invitation to contact us if people were interested in participate in the development of PROBAST-AI by using academic publications and social media (Twitter). We will send a link to the survey to whom have previously contacted us.

At the beginning of the survey, we will ask if they consent to participate in the study, after we have provided information regarding the purpose of the study and how their data will be handle. In addition, we ask them if they consent to be acknowledged by name in any potential academic publication. Also, we will ask them to invite other known experts to approach us, so we can send them an invitation, as well.

In case any participants refuse to participate in the survey, they will not have access to the survey and their personal information (name, e-mail) will be delete.

Amendment:

We will notify participants which took part in the PROBAST-AI Delphi round 1 that we have expanded our aim to create a risk of bias tool that is now called PROBAST 2.0. Participants will have to give their consent to participate in this second round of the Delphi survey.

Amendment (December 9th, 2022):

The working title was changed from PROBAST 2.0 to PROBAST +AI.

3.3 Describe how you manage your data to comply to the rights of study participants.

Right of Access

A linking table to personal data will be only accessible to the principal investigator who can re-identify participants when necessary and deliver, correct, or delete the data. This file will be saved in a secure research folder within UMC server (C_PersonalData), only accessible to the principal investigator.

Right of Rectification

If a participant wants to rectified responses, the principal investigator can deliver a code to the participant to access her/his responses and rectified them.

Right of Objection

If an invitee to the survey refuses to participate, they can close the survey and her/his name and e-mail will be deleted from our database.

Right to be forgotten

We inform invitee that their participation is voluntary and that they can stop taking part in the research at any moment. Removal of collected data can be done by the principal investigator.

3.4 Describe the tools and procedures that you use to ensure that only authorized persons have access to personal data.

Name and e-mail address of participants will be stored in a secure folder (C_PersonalData). The principal investigator will be the only person with access to this information.

3.5 Describe how you ensure secure transport of personal data and what contracts are in place for doing that.

We will not transport any personal data outside the UMCU network drives.

4. Data Storage and Backup

4.1 Describe where you will store your data and documentation during the research.

The digital files will be stored in the secured Research Folder Structure of the UMC Utrecht. All original raw data will be stored here and analyses will be done on working copies. We will need <1 GB storage space, so the capacity of the network drive will be sufficient. Documentation will be stored in the Research Folder Structure.

While the survey is open to be answer, raw data will be store in REDCap. After the survey is closed, we will delete all raw data store in REDCap and exported to the UMC Research Folder Structure.

We will make available within the main folder (PROBAST-AI) a 'readme' text file with the details about contributors, conditions for re-use, format, file types, and description of sub-folders structure.

Amendments (July 14, 2022):

With the expansion of the research project to develop the final product PROBAST 2.0, we are going to switch from REDCap to Castor. After the survey is closed, we will delete all raw data store in Castor and export it to the UMC Research Folder Structure.

The main folder (previously PROBAST-AI) was renamed to PROBAST 2.0.

4.2 Describe your backup strategy or the automated backup strategy of your storage locations.

During data collection, automatic backups will be made in the Electronic Data Capture Tool REDCap. Upon completion of data collection, all data will be exported and saved in the Research Folder Structure where they are automatically backed up by the UMC Utrecht backup system. All data in REDCap will be deleted afterwards.

Amendments (July 14, 2022):

For PROBAST 2.0, the Electronic Data Capture Tool changed from REDCap to Castor. Upon completion of data collection, all data will be exported and saved in the Research Folder Structure where they are automatically backed up by the UMC Utrecht backup system. All data in Castor will be deleted afterwards.

Amendment (December 9th, 2022):

The working title was changed from PROBAST 2.0 to PROBAST +AI.

5. Metadata and Documentation

5.1 Describe the metadata that you will collect and which standards you use.

We will make available a codebook with the data dictionary (survey questions, units of measurement, assumptions) for the raw data. This information is generated in REDCap while the survey is being created, so the file will be upload in the main folder (PROBAST-AI Delphi) after data collection. Alongside, another R script will be available with analyses and detailed methodology.

Amendments (July 14, 2022):

For PROBAST 2.0, we will switch from REDCap to Castor, as advised by the data management of the UMCU. The main folder was renamed to PROBAST 2.0. I will make a codebook available with the data dictionary (survey questions, units of measurement, assumptions) for the raw data. The information is generated in Castor while the survey is being created so the file will be uploaded in the main folder (PROBAST 2.0) after data collection. Alongside, another R script will be available with analyses and detailed methodology.

Amendment (December 9th, 2022):
The working title was changed from PROBAST 2.0 to PROBAST +AI.

5.2 Describe your version control and file naming standards.

We will distinguish versions by indicating the version in the filename of the master copy by adding a code after each edit, for example V1.1 (first number for major versions, last for minor versions). The most recent copy at the master location is always used as the source, and before any editing, this file is saved with the new version code in the filename. The file with the highest code number is the most recent version.

6. Data Analysis

6 Describe how you will make the data analysis procedure insightful for peers.

I have written an analysis plan in which I state why I will use which data and which statistical analysis we plan to do in which software. The analysis plan is stored in the project folder, so it is findable for my peers.
The data will be processed using the statistical software R (R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/). The scripts will contain comments, such that every step in the analysis is documented and peers can read why I made certain decisions during the analysis phase.

7. Data Preservation and Archiving

7.1 Describe which data and documents are needed to reproduce your findings.

The data package will contain: the study protocol describing the methods and materials, the raw data, the script to process the data, the scripts leading to tables and figures in the publication, a codebook with explanations on the variable names, and a ‘read_me.txt’ file with an overview of files included and their content and use.

7.2 Describe for how long the data and documents needed for reproducibility will be available.

Data and documentation needed to reproduce findings from this non-WMO study will be stored for at least 5 years.

7.3 Describe which archive or repository (include the link!) you will use for long-term archiving of your data and whether the repository is certified.

After finishing the project, the data package will be stored at the UMC Utrecht Research Folder Structure and is under the responsibility of the Principal Investigator of the research group. When the UMC Utrecht repository is available, the data package will be published here.

7.4 Give the Persistent Identifier (PID) that you will use as a permanent link to your published dataset.

I will not make the dataset available in an external repository. Therefore, a PID is not necessary.

8. Data Sharing Statement

8.1 Describe what reuse of your research data you intend or foresee, and what audience will be interested in your data.

Members of the research team might be reusing all research data in the final dataset to generate new research questions. The raw data can be of interest for other researchers for verification of study results.

8.2 Are there any reasons to make part of the data NOT publicly available or to restrict access to the data once made publicly available?

- Yes (please specify)

Our data will be shared with third parties after approval of the Principle Investigator. In the event that researchers outside the research team would like to reuse our data, this can only be granted if the research question is in line with the original consent given by the study participants or if it is for verification purposes. Every application therefore will be screened upon this requirement.
The data will be available for share and reuse immediately after the publication of the two main papers (PROBAST-AI statement and PROBAST-AI Explanation & Elaboration) until 1 year after.
Amendments (July 14, 2022):
After receiving comments from the PROBAST-AI Delphi round 1, we decided to expand our research to develop an expanded tool called PROBAST 2.0. This tool will include updated statements to assess the risk of bias and applicability/appropriateness of prediction models (PROBAST) and will include an extension for prediction models that use machine learning/ artificial intelligence techniques. Therefore, we will approach participants of PROBAST-AI Delphi round 1 again for their participation in this expanded survey.
Our data will be shared with third parties after approval of the Principle Investigator. In the event that researchers outside the research team would like to reuse our data, this can only be granted if the research question is in line with the original consent given by the study participants or if it is for verification purposes. Every application therefore will be screened upon this requirement.
The data will be available for share and reuse immediately after the publication of the two main papers ('PROBAST 2.0 statement' and 'PROBAST 2.0 Explanation & Elaboration') until 1 year after.

Amendment (December 9th, 2022):
The working title was changed from PROBAST 2.0 to PROBAST +AI. Therefore, the publications of the main papers will be renamed to 'PROBAST + AI statement' and 'PROBAST +AI Explanation & Elaboration').

8.3 Describe which metadata will be available with the data and what methods or software tools are needed to reuse the data.

All data and documents in the data package mentioned in 7.1 will be shared under restrictions. This Data Management Plan, codebook of the data and scripts of analysis in R will also be available.

8.4 Describe when and for how long the (meta)data will be available for reuse

- (Meta)data will be available as soon as article is published

(Meta)data will be available as soon as related articles are published until 1 year after.

8.5 Describe where you will make your data findable and available to others.

We will inform potential users that our data is available within the publication of 'PROBAST-AI Statement' and 'PROBAST-AI Elaboration and Explanation' papers.
Amendments (July 14, 2022):
The project will be expanded and was renamed to PROBAST 2.0. Hence, the title of the publications will change. We will inform potential users that our data is available within the publication of 'PROBAST 2.0 Statement' and 'PROBAST 2.0 Elaboration and Explanation' papers.
Amendment (December 9th, 2022):

The working title was changed from PROBAST 2.0 to PROBAST +AI. Therefore, the publications of the main papers will be renamed to 'PROBAST + AI statement' and 'PROBAST +AI Explanation & Elaboration').