

# DATA COLLECTION BIAS ASSESSMENT FORM

## Introduction Page

Name of project:	
------------------	--

Stated goal of project:	
-------------------------	--

Definition of the project:	
----------------------------	--

Motivation of the project:	
----------------------------	--

Team (including departments/ specialties):	
---	--

# ALGORITHM PACKAGE LEAFLET

This package leaflet is provided with machine-learning algorithms based on human behavioural data that aim for diagnostic, profiling, and/or predictive outcomes. The aim of this package leaflet is to gather information about the data collection phase as a part of algorithm development and/or machine learning.

## DESCRIPTION OF CHOICES IN DATA COLLECTION

### Algorithm goal

---

Goal of the resulting prediction model/classification model:	
What is the added value of reaching this goal?	
What is your concept definition of the outcome measure?	
Describe how the current most common outcome measure differs from the outcome measure that you expect to find?	
How were concepts translated into quantifiable data?	

### Study design

---

How would you describe the outcome measure?	
What feature/concept needs to be measured?	
How did you determine the measurements for quantification of these concepts?	
Which of these measurements are necessary for this study?	

## Materials and methods

Measurement methods/instruments used:	
Indicate per instrument its strengths and weaknesses	
How were these instruments validated?	
Or how do you plan to validate these instruments?	

## Sampling parameters

How is the target population delineated?		
What are the official statistical characteristics of your target population?		
Describe the difference between your target characteristics and the official characteristics of your target population?		
How can your sample be described?	<b>What is your ideal target sample?</b>	<b>What is your actual obtained sample?</b>
What other considerations are affecting your actual obtained sample?	<b>E.g. Think of financial constraints and practical issues such as language or technology</b>	
How large should the sample minimally be to statistically representative of the population?		
Describe your sampling strategy:	<b>List of the different criteria you used to describe your sample.</b>	<b>What is your sampling method?</b>

What is your recruitment strategy?	What is your ideal strategy?	What is your actual strategy?
How did you distribute your instruments for data collection?		
Describe the data collection process in steps		
In which phase did you encounter drop-outs/ non-response?		
Does the sample accurately match the proportions in the population?		
Was there a difference between the sampling plan and implementation?		
What measures have been used to address imbalances?		
Who was excluded from the sample?		
How do these people differ from the measures criterium of the people of the sample?		
Which statistical checks or other measures were used to eliminate inaccurate results?		
Team description in terms of ideal sample distributions:		



Were the data checked on:		Was there an influence of:		Could any of these reasons have caused refusal to participate?	
<input type="checkbox"/>	response rates between subgroups	<input type="checkbox"/>	the timing of the measurements	<input type="checkbox"/>	the (sensitive) nature of the study
<input type="checkbox"/>	response rates compared to other sources	<input type="checkbox"/>	the geographical demarcation of the study	<input type="checkbox"/>	personality traits of potential test subjects
<input type="checkbox"/>	variations within the study	<input type="checkbox"/>	the area of sampling	<input type="checkbox"/>	the social context of the study
<input type="checkbox"/>	distributions based on different hypotheses about non-response?	<input type="checkbox"/>	socio-economic status or health of test subjects	<input type="checkbox"/>	technical aspects of the study
		<input type="checkbox"/>	the method of approaching test subjects	<input type="checkbox"/>	language or cultural background
				<input type="checkbox"/>	response burden

## Reflection on bias risk

To what extent do relevant characteristics of test subject categories differ?			
What threshold have you determined to quantify good / moderate / low performance of input data?			
What checks have been made to ensure there were no errors arising from data entry?			
Have the right questions been asked? (check the appropriate box)	<table border="1"> <tbody> <tr> <td> <p><b>Face validity check:</b> Does it look like the collection method will collect what is required?</p> <p>Yes</p> <p>No</p> </td> <td> <p><b>Construct validity check:</b> Did you use literature, standard questionnaires or other means to check if your questions measure the construct you are interested in?</p> <p>Yes</p> <p>No</p> </td> </tr> </tbody> </table>	<p><b>Face validity check:</b> Does it look like the collection method will collect what is required?</p> <p>Yes</p> <p>No</p>	<p><b>Construct validity check:</b> Did you use literature, standard questionnaires or other means to check if your questions measure the construct you are interested in?</p> <p>Yes</p> <p>No</p>
<p><b>Face validity check:</b> Does it look like the collection method will collect what is required?</p> <p>Yes</p> <p>No</p>	<p><b>Construct validity check:</b> Did you use literature, standard questionnaires or other means to check if your questions measure the construct you are interested in?</p> <p>Yes</p> <p>No</p>		
Do the chosen methods produce data in a consistent manner?			
How well do results fit with other (research) results or theories?			

Have the data collection methods been re-tested?	
To what extent can the findings be generalized?	
Was there a reactive effect; behavioural change resulting from the methods used?	



## Appendix 2 Description of sample

<b>Total N</b>	Ideal	
	Actual	
	Difference	
<b>Age range</b>	Ideal	
	Actual	
	Difference	
<b>Gender distribution</b>	Ideal male	
	Ideal female	
	Ideal other	
	Actual male	
	Actual female	
	Actual other	
	Difference	
<b>Ethnicity</b>	Ideal	
	Actual	
	Transfer index	
<b>BMI</b>	Ideal	
	Actual	
	Difference	
<b>Social economic/Career</b>	Ideal	
	Actual	
	Difference	
<b>Geographical Distribution</b>	Ideal	
	Actual	
	Difference	



**Appendix 3 Data collection and response**

<b>Data Collection step</b>	<b>Drop-out (e.g. number of drop-outs)</b>	<b>Why</b>	<b>Non-response (e.g. number of non-response)</b>	<b>Why</b>

