Appendix B

Analysis Plan/Application data/materials

Analysis plan #: Date received: Date approval:

To be filled in by The Maastricht Study.

Applications submitted on Tuesday in even weeks, are in general discussed on Thursday in odd weeks. (With the exception of July and August).

1. **Title**

2. **First author**
   Name:

   Position:

   Institute:

   Address:

   Email:

   Phone number:

3. **Co-authors**
   Informed co-owner(s): ☐ Yes ☐ No
   Name(s) co-owner(s) that were informed:

   Provide list of co-owner(s) who agreed to be co-author:

   Are you a student and will this work be part of your bachelor/master thesis? ☐ Yes* ☐ No
   If yes, please provide details about your program:

* Please send your final thesis to the MT of the Maastricht Study via research.dms@mumc.nl

4. Research questions and hypotheses

5. Background

Background and rationale for addressing the research questions and hypotheses.

6. Design and sample

Study design and main in- and exclusion criteria of the study sample, e.g. cross-sectional study in participants with type 2 diabetes.

7. Variables

All requested variables should be identified. Please list the variable names from the online data dictionary of The Maastricht Study, available via: https://disqover.dh.maastrichtuniversity.nl

Variable Name	General Description	"Co-owner(s)"
Main independent variable(s)		
Outcome variable(s)		
Confounders		

8. Statistical analyses

Concisely describe the statistical analyses. This should include: 1. Statistical testing; 2. Model structure; 3. Sensitivity analyses; and 4. Interaction and stratified analyses.
A clear distinction should be made between confounders, interactions and mediators.
(For statistical advice See Attachment 1).

9. Mock Tables

Include mock-up of key tables.

10. Timeline

A timeline for completion and submission of the paper.

11. Compensation

What compensation is proposed by the applicant
(for information refer to 'Procedure data materials' see https://www.demaastrichtstudie.nl/data-guidelines)

12. Agreement for the of data and/or materials of the Maastricht Study

This agreement is for the analysis plan entitled:

The participating researchers are:

I certify that I am aware of the rules described in 'Procedure Data/Materials - The Maastricht Study' which include:

- The data/materials should be treated confidentially
- The data/materials may not be shared with others who are not included in this project
- I agree with the "Maastricht Study Data License Agreement" as stated in Appendix D (see below)
- The approval is valid for 1 year: After a year a written progress report should be submitted.
- For publications the rules as described in the 'Procedure Publicatie' are applicable.

Date

Name first author and signature

Appendix D

Maastricht Study Data License Agreement

This end-user License Agreement is a legal agreement between (fill in institution name and address) . legally represented by (fill in name).

The "Licensee"

and

Maastricht University/University Hospital Maastricht, The Maastricht Study, legally represented by The Maastricht Study Management Team, the “Licensor”

1. By signing this License agreement, or by accessing, storing, copying, processing or otherwise using the data from The Maastricht Study “the Data”, the Licensee agrees to be bound by the terms of this License Agreement.
2. The Data will be provided to Licensor by means of delivery of a data carrier. The data carrier and the Data contained therein remains the property of the Licensor. Licensee acknowledges and agrees that Licensor holds all right, title and interest in and to the Data and the intellectual property rights or related rights therein. Licensee agrees that the intellectual property rights or related rights of Licensor are not transferred, assigned or affected in any way as a result of this License Agreement.
3. This License grants the Licensee a nonexclusive, nontransferable, no-cost, royalty free right for limited duration to use the Data solely for internal, non-commercial, non-clinical, academic research purposes only.
   The Licensee is authorized to store the Data on a single laptop, personal computer, tablet or workstation (“Computers”) The Data may not be stored so that it is accessible to multiple users over an intranet.
4. The Licensor makes clear that no condition is made or to be implied, nor is any warranty given or to be implied, as to the accuracy of the Data, or that it will be suitable for any particular purpose or for use under any specific conditions. Furthermore, the Licensor disclaims all responsibility for the use which is made of the Data.
5. The Licensee agrees to indemnify the Licensor and hold the Licensor harmless from and against any and all claims, damages and liabilities asserted by third parties (including claims for negligence) which arise directly or indirectly from the use of the Data by Licensee.
6. Unless expressly stipulated in this License Agreement, no part of the Data may be reproduced, published, disseminated, modified, transmitted or transferred in any form or by any means, electronic or in print, without the express permission of the Licensor. The Licensee may be held legally responsible for any copyright infringement that is caused or encouraged by the failure to abide by these terms and conditions.
7. Licensee is not permitted under this License to use the Data commercially. Use for which any financial return or other consideration is received shall be defined as commercial use.
8. This License Agreement becomes effective on the final date of signature of this License agreement, or at the moment the Licensee first accesses, stores, copies, processes or otherwise uses the Data, whichever comes first, and is entered for a period of 12 months.
   Licensor may terminate this License Agreement for any reason with thirty (30) days prior written notice. Without prejudice to any other rights, Licensor may terminate this License Agreement with immediate effect if Licensee fails to comply with the terms and conditions of this License Agreement.
9. In case of termination of this License Agreement, the Licensee must return the original data carrier to Licensor and destroy all copies of the Data immediately.
10. This License Agreement is governed by Dutch law. Any dispute shall be brought exclusively before the competent courts of Maastricht, the Netherlands.

Attachment 1

Statistical and methodological advice – The Maastricht Study

Version 1.0 (25-11-2019)

The aim of this document is to provide the researcher some guidance in regression modeling, confounder selection and statistics as usually performed in The Maastricht Study. Generally speaking, this document is intended as a guideline, not as a mandatory protocol. Deviations are possible, as every research analysis plan differs.

1. Model structure

Advice: structure the statistical models as shown below

It is important to explicitly describe the model structure in the statistical analysis plan. Beforehand, define on what model the conclusion will be based (advice: the fully adjusted model). This prevents biased reporting of results. The advised standard model structure is as follows:

- Model 1: crude;
- Model 2: as model 1 with additional adjustment for age, sex, type 2 diabetes status*;
- Model 3: as model 2 with additional adjustment for other confounders (e.g. lifestyle).

*If no other glycaemia-related covariates are included in the models (e.g. glucose metabolism status, HbA_{1c}), adjustment for type 2 diabetes status is required due to the study design which is characterized by an alternative recruitment approach for individuals with type 2 diabetes.

Researchers may choose to perform a 'complete case analysis' or 'available case analysis'. A 'complete case analysis' (i.e. the participants with any missing data are excluded in all models) has the advantage of making the comparison of consecutive regression models clearer. However, this analysis excludes more participants in earlier models, which may introduce selection bias in earlier models (i.e. the values may not be missing completely at random). When an 'available case analysis' is performed, the consecutive models will often have different sample sizes, as the number of missing values is usually higher in models with more covariates. As a consequence, differences in regression coefficients between models may be attributable to the inclusion of additional covariates (i.e. less confounding), the sample size differences (i.e greater selection bias), or both. In any case, the number of participants in the final model should be the same for both methods. Accordingly, it is important to always investigate selection bias by comparing the in- and excluded individuals (see 5. Exclusion of participants).

2. Confounders

a. The selection of confounders

Advice: select confounders based on literature and logical reasoning

The selection of confounders is an important process that should be well thought-out. It should be based on literature and logical reasoning. We strongly discourage univariate testing as a method to select confounders (except for exploratory analyses). Researchers should preferably not adjust for mediators, ascending/descending proxies and descendants of the outcome, since this could introduce overadjustment bias. We advise to explore the effect of such variables as part of an additional analysis (see 3b. Sensitivity analyses). For more on overadjustment bias see: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2744485/

b. Strongly advised confounders

Advice: always adjust for age, sex, and type 2 diabetes status; strongly consider diet, physical activity, smoking, alcohol use, and educational level

We strongly advise that certain variables are added as confounders, either due to study design (see 1. Model structure) or their relevance. These include:

- Age and sex (relevance; mandatory).
- Type 2 diabetes status (by study design; mandatory if no other glycaemia-related confounder is included).
- Diet, physical activity, smoking status, alcohol use (relevance; encouraged).
- Educational level as proxy of socioeconomic status (relevance; encouraged).

c. Mandatory confounder pairs

Advice: when blood pressure and/or lipid profile are covariates, researchers should also adjust for relevant medication use

When the association is adjusted for certain confounders, we strongly advise that variables that strongly influence the aforementioned confounder are included in the model. This is the case for:

- Blood pressure variables (e.g. office systolic blood pressure, mean arterial pressure); adjust for blood pressure-lowering medication (yes/no).
- Lipid variables (e.g. LDL cholesterol, TC/HDL cholesterol ratio); adjust for lipid-modifying medication (yes/no).

d. Confounder preference

Advice: when choosing between similar confounders, the choice is preferably based on literature

It is generally up to the individual researcher to choose between similar covariates (e.g. body mass index versus waist circumference). In their decision, researchers should consider the clinimetric properties of the measurements and the number of missing participants. In case of physical activity, for example, the activity monitor (ActivPAL), although arguably preferable, has more missing values than the CHAMPS questionnaire. Some suggestions:

- Educational level (as proxy of socioeconomic status).
- Lipids: TC/HDL cholesterol ratio (additional adjustment for triglycerides unnecessary).
- CHAMPS questionnaire or activity monitor (latter has more missing data).
- Smoking status (never, former, current) and not pack years.
- MINI depression and not PHQ-9 (former is gold standard, latter has more missing data).
- Dutch Healthy Eating Index (not Mediterranean Diet Score).

3. Additional analyses

a. Stratified analyses: testing for interaction

Advice: test for interaction with sex and diabetes status (or glucose metabolism status)

We strongly advise that certain interactions are investigated, either due to their relevance or for study design-related reasons. These include:

- Sex: strongly encouraged (based on international consensus).
- Type 2 diabetes status: strongly encouraged (by design); OR glucose metabolism status: encouraged (as part of an biologically plausible exploratory analysis).
- Other: not unless there is a specific hypothesis (or a plausible mechanism). This should be described in the research proposal.

To study this, add an interaction term (e.g. LDL cholesterol*sex) and examine the P-value in the fully adjusted model (which is least contaminated by confounding). Interaction terms for ordinal/nominal variables should be added as dummy variables (e.g. as LDL*prediabetes and LDL*type 2 diabetes; not LDL*glucose metabolism status).

Generally, the P-value cut-off for statistical significance of an interaction is 0.05. In case of small samples sizes or more exploratory analyses, an exception of $p < 0.10$ can be made. When a statistically significant interaction is found, stratified results must be presented in addition to the overall results.

b. Sensitivity analyses

Advice: perform certain sensitivity analyses to study result robustness

Sensitivity analyses are important to investigate the robustness of the results (i.e. whether substitution of a determinant or an important confounder with a similar variable, change in in/exclusion criteria [e.g. exclusion of all individuals with diabetes], etc. strongly influences the results). It is important to describe the planned sensitivity analyses in your research analysis plan. Still, additional analyses can be added based on the findings of the study. Generally speaking, the following sensitivity analysis (i.e. exchanging the variables mentioned below) is strongly advised:

- Glucose metabolism status (if central determinant, outcome or confounder): HbA1c, fasting plasma glucose, and post-load glucose.

The following sensitivity analyses can be considered:

- Office blood pressure: 24-h ambulatory blood pressure.
- Educational level (as proxy of socioeconomic status): income and occupational status.
- CHAMPS questionnaire: activity monitor (ActivPAL).

Journals occasionally request additional adjustment for variables that are not (truly) confounders (see 2a. The selection of confounders). In this case, it is advised to do so as part of additional (sensitivity) analyses, not the main analysis.

4. Reviewing assumptions: normal distribution and linearity

Advice: inspect normality visually

To determine whether a variable is normally distributed, visual inspection is preferred (i.e. via histogram, Q-Q plots or P-P plots). The use of a statistical test (e.g. Kolmogorov-Smirnov test, Shapiro-Wilk test) is less informative, as it will almost always indicate non-normality in large study populations.

Please note that for linear regression, the assumption holds that the *residuals* should be normally distributed, and not the variables *itself*. Even if both the determinant and outcome variables are non-normally distributed, the residuals can be normally distributed (and vice versa). Normal distribution of residuals can be examined via residual histograms, Q-Q plots or P-P plots.

Advice: inspect linearity of a continuous determinant-outcome association with the use of quintiles

In linear regression, the determinant should be linearly associated with the outcome. If the determinant is continuous, the recommendation is to divide the determinant into quintiles and visually inspect linearity. If quintiles are not feasible due to small sample sizes, quartiles or tertiles should be used. Other graphs, e.g. scatterplots, are difficult to interpret visually in large study populations.

5. Exclusion of participants

Advice: exclusion of individuals with diabetes other than type 2 depends on your analysis

Many previous studies from The Maastricht Study have *always* excluded individuals with diabetes *other than type 2*. It is not clear why, and it should not be considered standard practice for two reasons. First, individuals with diabetes *other than type 2* are part of the general population. Second, automatic exclusion could be considered unethical, as these participants have given The Maastricht Study their time and effort in order to provide us their data.

If glucose metabolism status (i.e. normal glucose metabolism – prediabetes – type 2 diabetes) is the main determinant or outcome, it is valid to exclude individuals with other types of diabetes.

Advice: compare characteristics of included and excluded individuals from the analysis

We strongly advise comparing the characteristics of individuals included in versus excluded from the analyses to investigate whether the study sample is representative of the total study population. This gives insight into whether individuals excluded from the analysis were (un)healthier, which could affect generalizability of the results. The decision to *statistically* test differences (included versus excluded) is up to the individual researcher.

6. Composite scores

Researchers regularly create composite scores of similar variables to be used as determinant, outcome, confounder, or mediator (e.g. inflammation composite score, which consists of CRP, SAA, E-Selectin, etc.). Such a composite score needs to be calculated based on the final study population in order to have a mean of 0 and a SD of 1 (syntaxes are available from The Maastricht Study data management team). Importantly, the construction of the composite score should be biologically sensible. For example, only if you expect blood pressure to have a similar pathophysiological effect on the three cognition domains may you construct a (global) cognition composite score. Of note, the individual measures or domains of the composite score are preferably investigated individually (as part of the sensitivity analyses), especially if they are used as the main outcome or determinant.

7. Cross-sectional research semantics

Advice: avoid terms that imply longitudinal data / interventional data / causality when describing cross-sectional associations

In case of describing cross-sectional The Maastricht Study data, researchers should avoid terms that imply longitudinal data / interventional data / causality. A few examples are:

- Increase / Decrease: imply change over time. Use: greater, higher, lower etc.
- Change / Effect: imply change or effect due to intervention. Use: difference, association.
- 'Baseline' characteristics: this implies multiple time points. Use: participant characteristics

8. STROBE checklists

STROBE stands is an international, collaborative initiative of epidemiologists, methodologists, statisticians, researchers and journal editors involved in the conduct and dissemination of scientific studies, with the common aim of providing guidance on how to correctly report (observational) research. The STROBE checklist was developed in order to improve the quality of research and to standardize research publications. An increasing number of journals requests authors to add a STROBE checklist as a supplementary file when submitting their manuscript. It is therefore advised to review this checklist prior to submitting an Appendix B. For more information, please visit: https://www.strobe-statement.org/index.php?id=available-checklists .