Application procedure for Data and/or Materials (version Sept 2026)

Application Procedure for Data and/or Materials – The Maastricht Study

Data and materials from The Maastricht Study are available for individual researchers. Obtaining these data and materials requires an application to be submitted to the Management Team[a] (MT) of The Maastricht Study.

There are 2 types of applications:

1. A proposal for a PhD project fully based on The Maastricht Study data can be submitted The Maastricht Study via email at um-researchdms@maastrichtuniversity.nl using the form "A.The_Maastricht_Study_PhD_plan.DOCX"). This proposal briefly describes the project outline, the aims and research questions, indicates a timeline and presents the supervisors of the project. The application for a PhD project proposal may not contain more than 3 research questions. The PhD supervisory team must include at least one member of the MT (as first and/or co-supervisor).

2. A proposal for an individual research article or an internship can be submitted via the online data request portal. This proposal includes a description of the background of the topic, the research question(s), a list of the requested data, and the researchers involved.

Proposals are evaluated based on the following:

- Scientific approach and content, possible overlap with other (on-going) projects, and feasibility (also in terms of funding and personnel). The applicant will receive an email with the MT's decision within 30 days after submission.
- In case major changes are requested, the revised application must be accompanied by a response to the comments of the reviewer(s). A final decision of the MT will be received within 30 days.
- In case the applicant does not agree with the decision of the MT, they may submit a substantiated appeal to the MT.
- Requests for changes to a previously-approved research proposal (e.g. alterations to the research question, new dependent or independent variables) must be submitted via the online data request portal.
- It is the responsibility of the applicant to consult with the domain experts[b] of all requested data (including covariates) and/or materials before submitting a request via the online data request portal.
- Statistical analysis plan: the research proposal should include a sufficiently detailed statistical analysis plan, including the proposed statistical methods, model structure, covariate selection, and, where relevant, interaction and sensitivity analyses. Applicants are advised to consult *Appendix 1: Statistical and methodological guidance – The Maastricht Study* when preparing their analysis plan.
- Domain experts of data and/or materials have the right to be co-authors of a paper, depending on the content and/or research question, as well as the required effort and contributions to the research project (Vancouver rules).
- In case the MT is listed as domain expert for certain variables, it is not necessary to send the research proposal to each individual MT member.
- Members of the MT have the right to be co-authors on publications, reflecting their contribution to making The Maastricht Study possible. The names of the MT members who will be co-authors are determined during the evaluation of the proposal. The applicant will be informed of these names via the online application form at the time of notification of the approval or rejection of the research proposal.

The applicant of a data request must comply with the following:

- One year after the approval of a research proposal, the researcher must inform the MT of the progress of the project (use form: Yearly_Progress_Update_Research_Proposal.DOCX).
- One applicant can have a maximum of 3 ongoing data requests, in order to prevent too much data being claimed by one researcher. Once a final manuscript is submitted to the MT (via form: C.Application_for_review_manuscript_abstract.DOCX), or the applicant has withdrawn his/her application, it is no longer considered ongoing.
- Data and materials provided by The Maastricht Study remain property of The Maastricht Study.
- The applicant is not allowed to use the data/materials for scientific purposes other than described in the approved research proposal, and is not allowed to share the data with other researchers without written approval from the MT.
- The applicant will be responsible for any additional costs for the project, including but not restricted, the cost of biomarker measurements in biobanked materials of Maastricht Study participants.
- When an application for a PhD project or data is approved, the applicant agrees to comply with the 'Publication Procedure– The Maastricht Study'.
- Handling fee per data request (equivalent to 1 scientific publication) is:
    - € 3300 for internal data requests (i.e. for papers that have Maastricht University and/or MUMC employees as first and last authors)
    - € 5500 (excl. any applicable V.A.T.) for external data requests (i.e. all other situations)

The publication fee applies to all researchers requesting Maastricht Study data, including domain experts and members of the Maastricht Study management team. The data application fee is needed to keep data management, including data distribution to researchers, up and running.

A PhD student of The Maastricht Study is additionally expected to comply with the following:

- A PhD project proposal must be submitted to 'The Maastricht Study' within 6 months after starting the PhD project.
- In addition to project proposal, the PhD student must submit a data request with research plan for each individual article using the online data request portal
- During the first 2 years the PhD student is expected to spend at least 0.4 fte on activities involving data collection[d] for The Maastricht Study (assuming a 4-year PhD project (at 1.0 fte)).
- If needed, the PhD student is expected to be available for data collection activities for The Maastricht Study during out-of-off hours (e.g. evening hours and Saturdays).

[a] Members of the Management Team: Prof. M. Brouwers, Dr. C. van der Kallen, Dr. A. Koster, Dr. M. Greevenbroek, Dr. A. Wesselius, Prof. S. Köhler, Prof. B. de Galan.

[b] Domain expert: An MUMC and/or UM employee with a senior position in a research discipline who made a substantial contribution to the collection of data and/or materials, whether financially or content based. In case a domain expert no longer holds a position at the UM or MUMC, the position as domain expert is terminated. The head of the department will appoint a new domain expert. A list of domain experts is available from the MT. If a variable and/or materials have more than one domain expert, a contact person will represent the interests of the other domain experts.

[c] i.e. not an employee of Maastricht University or Maastricht University Medical Centre

[d] Data collection: actual collection of data during onsite visits of participants to the research centra, as well as logistical activities necessary for data collection and data cleaning or data processing for The Maastricht Study.

Appendix 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, HbA1c), 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).

- NIT_alcoholtot (obtained via FFQ) or N_ALCOHOL_CAT (aantal glazen per week) (former is more precise, latter has less missing values)

In addition: when the researcher is interested in the separate confounding effects of alcohol consumption and other dietary habits he/she is advised to include 'NIT_alcoholtot' (obtained via FFQ) and the Dutch Healthy Diet calculated without alcohol ('DHD_zonder alcohol component') in the analyses.

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