Data Management Plan

General information

Name and contact details	Name: Lizbeth Burgos Ochoa Email: l.burgosochoa@students.uu.nl Project Affiliation: VU medical center
Name of project and group	Project: COMPARISON OF METHODS TO PERFORM MEDIATION ANALYSIS WITH TIME-TO-EVENT OUTCOMES Supervisors: Judith Rijnhart, Martijn Heymans, Jos Twisk
Description of your research	"Comparison of methods to perform mediation analysis with time-to-event outcomes" is a research project that aims to compare the statistical performance of four methods to perform mediation analysis with time-to-event outcomes. This is done by means of Monte Carlo simulations and an illustration with an empirical dataset. The compared methods are I) the classical mediation approach with Cox PH model (ab and c-c' methods); II) the classical mediation approach with the AFT model (ab and c-c' methods); III-IV) Potential outcomes approach for both, Cox and AFT models, respectively.
Project duration	Start: 01-09-2017 End: 09-05-2018
Names of people and their responsibilities for data management	Lizbeth Burgos Ochoa: data generation, storage, archiving Supervisors: storage, archiving External party (Netherlands Study of Depression and Anxiety): collection, storage, documentation, archiving
Funding body	Not applicable
Grant number	Not applicable
Partner organizations	Netherlands Study of Depression and Anxiety (NESDA)

About this Data Management Plan

Date written	29-04-2018
Date last update	No updates have been made
Version	First version

Data Management

<table>
<tr><th colspan="2"><b>1. Data collection</b>
Describing the data you will be creating/collecting</th></tr>
<tr><td colspan="2"><b>General description</b>
The main objective of the current research project requires the use of two types of data: 1) Data generated by the Monte Carlo simulations, and 2) Empirical dataset to illustrate the compared methods.</td></tr>
<tr><td>1.1</td><td><b>Will the project use existing or third party data?</b>
☐ No
☐ Own / group previous research
☒ Academic collaborators
☐ Commercial collaborators
☐ Publicly available database/archive
☐ Specialist commercial data provider
☐ Other (please specify)

The illustration of the compared methods is done through an empirical dataset. Our empirical dataset corresponds to the study by Gerrits et al., 2014, based on data from the Netherlands Study of Depression and Anxiety (NESDA) (Penninx et al., 2008), a longitudinal cohort study, designed to investigate the long-term course and consequences of depressive and anxiety disorders. The dataset was accessed and downloaded through the NESDA website. The default provided format is a .sav file (for SPSS). The NESDA data can only be accessed upon request by sending (and getting the approval) of a Data Analysis Plan (DAP)* to the NESDA Research Committee. Therefore, the provided data cannot be used or redistributed to other parties not specified in the DAP.
*DAP at the end of the document.</td></tr>
<tr><td>1.2</td><td><b>What type()s of data will you collect or create, in what file format(s)?</b>
R code (.rmd): For the generation of the simulations datasets, estimation, performance measures, results tables, and plots.
Raw results (.txt files): Results tables created with the R code.
Additional File 1 (PDF file): Contains the processed results tables (additional file for paper).
Additional Images (PDF files)
All these files are necessary to generate and analyze the data in order to write a thesis/publish a paper.</td></tr>
<tr><td>1.3</td><td><b>How will you collect and/or create your data?</b>
R code is written as an R Notebook file. Each file contains the necessary code to perform the Monte Carlo simulations and produce relevant output, such as tables and plots. Raw results files are a direct product of running the R code. The Additional file will be created from the raw results tables. Data from the NESDA study was collected by the NESDA research group, complying with the necessary ethical and privacy requirements. More information in (Penninx et al., 2008). The requested data files were accessed (after approval) through NESDA's researcher's website and stored by the principal researcher (Lizbeth Burgos Ochoa).</td></tr>
<tr><td>1.4</td><td><b>What tools, instruments, equipment, hardware or software will you use to capture, produce, collect or create the data?</b>
Process: R, Microsoft Word, Microsoft Excel, Microsoft Power Point (All Available)
View: R, Microsoft Word, Microsoft Excel (All available)</td></tr>
</table>

Analyse: R (Available)

1.5 **What is the estimated size of the data?**

Data stage	Specification of the type of research data	Software choice and file format	Total number of data files	Total Data size
R code	R code to run simulations, analyze empirical data and produce output	Original format: .RMD Processed in: R software	9	213 KB
NESDA dataset	Dataset from NESDA study	Original format: .sav Processed in: R software	1	14 KB *Not provided in the archive
Raw results	Results tables in form of text files	Original format: .txt Processed in: Microsoft Excel/ Microsoft Word	16	96 KB
Additional File 1	Results tables in a single PDF file	Original format: .txt Processed in: Microsoft Word	1	598 KB
Additional images	Images for the final article not created within R code.	Original format: .jpeg Processed in: Microsoft Power Point	2	96 KB

2. Data storage and security

Ensuring that all research data are stored securely and backed up or copied regularly during your research

2.1 **Where will you store your data?**

Please describe how safe storage is guaranteed. Specify your method if your data is collected and/or transported in different locations.

☒ On university departmental network storage (Utrecht University)
☐ On university personal network storage
☐ In a Virtual Research Environment
☒ Physical storage (internal hard drive)
☒ Cloud service (Dropbox)
☒ Online repository (Github)

All the data is stored on an internal hard drive and a cloud service (Dropbox). After the end of the study, all the data (except the NESDA dataset, see 4.1) will be stored in the university network storage and a Github repository. All the files are in different locations, thus minimizing the chance for loss of all data at any given time since the backup is possible from a second/third location.

2.2	**When will your data be backed up?** There will be regular back-upping. Internal Hard-disk backup > Weekly Dropbox > Weekly University Network Storage > Once after project completion Github > Once after project completion
2.3	**Are there any commercialization, ethical or confidentiality restrictions about handling your data?** Please specify briefly.
	☒ Contractual obligations ☐ Requirements by law: protection of personal data (e.g. privacy law): specify in 4.1 ☒ Requirements by law: copyright, intellectual property: specify in 4.1 ☒ Ethical restrictions (e.g. ethical review): specify in 4.1 ☐ Commercial considerations (e.g. patentability) ☐ Formal security standards ☐ No requirements ☐ Other, namely: The obligations with NESDA require that the data will not be used for other purposes than the ones specified in the approved Data Analysis Plan. Other restrictions will be addressed in section 4.1.
2.4	**How will access to the data be managed during the project?** During the project, only the researchers involved have access to the data derived from the project. After the project is finished, all the data, except the NESDA dataset, as mentioned above, will be open access.
2.5	**What are the main risks to data security?** Accidental deletion, malfunction of internal hard disks, theft. If data became unusable or got lost, then this would prove highly detrimental to future analysis, and the publication of the generated data. If the NESDA dataset is filtered, the information could be used by third parties in a detrimental way. However, the individual privacy of the participants would not be compromised given that we do not have access to variables that link any participant with the identification number. Only the principal investigator from NESDA and their data manager have access to such information, more information can be found in (Penninx et al., 2008).

2.6	**What measures do you take to comply with the security requirements and to mitigate the risks?** Describe how you can restore your data in the event of data loss and who is responsible. If applicable, please describe procedures to ensure personal data are handled confidentially and who is responsible. ☒ Access restrictions ☐ Encryptions ☐ Data processing ☐ De-identification / Anonymisation ☒ Regular back-ups ☒ Master copy stored on university network storage ☐ Master copy stored elsewhere ☐ Other, namely: … The locations where the NESDA dataset is stored are only accessed by the researchers. In the case of the R code and other generated files, to comply the risk of information loss continuous backups are made. Furthermore, the generated data (except the NESDA dataset) will be archived in the university network storage.
2.7	**How do you differentiate between raw and processed data?** Please explain briefly why you (do not) differentiate.
	☐ I will not differentiate ☒ I will create a new file for processed data ☐ I will create a new file for processed data and I will lock raw data ☐ Other, namely: … The only raw data are the raw results tables. They will be processed for the main text of the research paper and the additional files linked to the article.
2.8	**Are there any non-digital data or outputs that the project will generate? Where will these outputs be stored?** All the data and output derived from the project are digital files.
2.9	**Do you expect to have any supplementary costs for storage not covered by the project budget?** The storage costs are either free or covered by the project budget.

	3. Data documentation Documenting your data to help future users to understand and reuse it
3.1	**How will files be named?** Raw files will be named as follows: “results.sim.approach.model.exposuremediator.txt”. For example, “results.sim.cl.AFT.BC.txt” is the file name for the results table of the classical mediation approach (cl), with an AFT model, in an scenario with Binary exposure and Continuous mediator (BC). Similarly, R code files will be named as “Simulations_approach _ exposuremediator.rmd”.

3.2 **How will folders be named and structured?**

1.R code

This folder contains the code to perform the Monte Carlo simulations and the illustration with the empirical dataset. It is divided into three subfolders: Classical Mediation Approach, Potential Outcomes, Empirical Illustration. All the documents in this folder are R Notebook files, which are opened with R Studio. Each file has all the code required to generate (or load) the datasets, do the estimation procedure and produce the final results (tables and plots). Specific instructions for running the code can be found inside each file.

-The Classical Mediation Approach folder contains four R Notebook files to perform the simulations for the two methods based on the classical mediation approach. To avoid mixing results, each file corresponds to an exposure-mediator-type combination (i.e. normal exposure-normal mediator, normal exposure-binary mediator…etc.).

-The Potential Outcomes folder contains four R Notebook files to perform the simulations for the two methods based on the Potential Outcomes mediation approach. As before, each file corresponds to an exposure-mediator-type combination (i.e. normal exposure-normal mediator, normal exposure-binary mediator…etc.).

-The empirical illustration folder contains instructions to analyze the empirical dataset with the four methods compared in the simulation study. Unfortunately, as the variables used are part of the Netherlands Anxiety and Depression Study (NESDA) dataset we cannot provide access to the used dataset. Further explanation on this matter can be found in the Data Analysis Plan. However, given that the provided code was designed to be generic, we encourage researchers to use and adapt the code to fulfill their own means.

2.Raw results

This folder contains two subfolders, Classical Mediation Approach, and Potential Outcomes. Each subfolder contains eight text files (.txt) with the raw results derived from the simulations performed in the R Notebooks (four from the classical approach, four from the potential outcomes approach). The structure of the Raw results folder is similar to the R code folder.

3. Additional File 1

This folder contains a pdf file with the full results tables derived from the Raw results. In the paper, this file is referred as Additional File 1.

4. Data Management Plan

This folder contains a single PDF document corresponding to the Data Management Plan (DMP). The DMP is a formal document that outlines how the data were handled during the research project, and how it will be handled after the project is completed.

5. Additional Figures

Two PDF files containing additional figures: Single mediator model and Compared methods.

3.3 **How do you handle version control to maintain all changes that are made to the data?**

☐ No version control (e.g. original files are overwritten)

☐ Version control software, namely: …

☒ Data/version number in filename/folder

☒ 'Track changes' feature in the software

☒ By saving the script with which I process my data

☐ Other, namely:

For R documents version control is done by numbering the versions. For data processed in Excel, a new spreadsheet is created within the same file to preserve the original data and allow different methods of analysis/versions to be tracked.

3.4 **What metadata standard will be used, if any?**[i]

☐ No metadata standard is used

☐ Generic metadata standard (e.g. Dublin Core)

☐ Standard automatic Windows metadata (e.g. from Word, Excel)

☐ Specialised metadata standard, namely: …

☒ Another metadata standard

Metadata will be added to the R code files with information of the simulation that can be run with each file or the analysis that can be performed.

3.5 **What supporting information/documentation will you create to enhance understanding of the data?**

A readme.txt file will be added to the main folder containing the data for open access.

4. Data access, sharing, and reuse

Managing access and security, sharing your data

4.1 **Are there any restrictions placed on sharing/reuse of some / all of your data?**

The dataset from NESDA cannot be redistributed to other parties due to the intellectual property of the organization and ethical issues. The dataset contains information regarding health outcomes which should not be published without approval. Publications will be revised and approved by the NESDA research committee. More information regarding the procedure to use data from NESDA can be found in https://www.nesda.nl/pro-index/nesda-analysis-plans/

4.2 **With whom will you share your data at which stage in your research? You can use the table below.**

	Would not share with anyone	Would share with my immediate collaborators	Would share with others in my research center or	Would share with scientists in my field	Would share with scientists outside of my field	Would share with anyone

			at my institution			
Immediately after the data has been generated		X				
After the data has been analyzed		X				
Immediately before publication		X				
Immediately after completion of the project			X (Not applicable to NESDA dataset, see 4.1)	X (Not applicable to NESDA dataset, see 4.1)	X (Not applicable to NESDA dataset, see 4.1)	X (Not applicable to NESDA dataset, see 4.1)

4.3 **If intending to share any part of the data, do your participant consent forms include information about intentions for sharing, retention of data and steps taken to protect participants privacy and confidentiality?**

☒ Not applicable.
☐ Yes. Please specify the relevant formula in the consent form.

The data that will be shared was not collected from participants.

4.4 **Who has authority to grant (additional) access to your data?**
Please describe briefly.

☐ Only you
☐ A colleague from the project, namely: …
☐ Supervisor
☐ Funder
☒ Collaborator/research partner organization
☐ Other, namely:

For the data that is not open access, the NESDA dataset, as mentioned before, can only be accessed upon request by NESDA.

4.5 **How will you manage copyright and Intellectual Property Rights issues?**
Who owns the data? How will the data be licensed for reuse? Please describe briefly your choices and their consequences.

All the data, except NESDA dataset, will be open access, so anyone can reuse it and reproduce the results. The VU medical center owns the data derived from the project.

4.6 **What is the audience for reuse?**
Researchers interested in replicating the results from the study, other researchers interested in the field of mediation and /or survival analysis. Empirical researchers interested in performing mediation analysis with time-to-event outcomes.

	5. Data preservation and archiving Preserving your data
5.1	**Which criteria will you use to decide which data has to be archived?** Please briefly describe your choices. ☒ Type of data (raw, processed) and how easy it is to reproduce it ☒ The relevance of content for others ☐ Usability of format for others ☒ Data underlying publications ☐ Verification of research ☐ Available time ☐ Available money ☐ Other, namely: … All the data generated by the researcher's data (raw and processed) will be archived. The purpose of this is for possible research verification and verification of data underlying publications.
5.2	**How long should your data be preserved? Are there any requirements regarding the disposal of data?** The data generated by the researchers (R code, raw results, images, etc.) will be archived in Github and the university network storage for a minimum period of 10 years. The data from NESDA is stored and managed by the organization, but, storage for a period of 15 years can be guaranteed (in accordance with Article 454, paragraph 3 of the Medical Treatment Agreement Act (Wet op de geneeskundige behandelingsovereenkomst, WGBO). After completion of the project, the researchers of this project must delete the NESDA dataset from their own storage units.
5.3	**Which data repository is appropriate for archiving your data?** Github was selected as the online repository for the open access data derived from the project. The research archive can be accessed through the following link: https://github.com/LizBurgosOchoa/Mediation-and-survival
5.4	**Does the archive have specific requirements concerning file formats, metadata etc?** Github accepts all the types of files that will be uploaded for the archive. A readme file will be added to the research archive to guide the reader through the file structure.

5.5	**What costs (if any) will your selected repository charge? Who pays?** No costs are derived from the storage of the research archive.
5.6	**Who is responsible for the data after the project ends?** Principal researcher and supervisors.

Description of the NESDA dataset

As the NESDA dataset cannot be shared with others, but the researchers registered in the Data Analysis Plan, we provide a general description of the data.

The Netherlands Study of Depression and Anxiety (NESDA), is a longitudinal cohort study, designed to investigate the long-term course and consequences of depressive and anxiety disorders. The used dataset consists of 1122 individuals with remitted depressive or anxiety disorder followed up for a period of four years. The dataset contained the following variables:

Sex: binary, male or female
Age of respondent: continuous, in years
Education from respondent: continuous, in years
Recency of last episode: binary, ≤1 year, over a year.
QIDS score: continuous, Quick Inventory of Depression Symptomatology scores, values from 0 to 20
CPG score: ordinal, Chronic Pain Grade score, values from 0 to 4
Time-to-recurrence: continuous, time until the recurrence of depression (in months)
Event: binary, event or non-event, did the respondent had recurrence of depression?

Descriptives statistics are shown below:

Characteristics	Population (%) N= 1122
Sex, female	765 (68.2%)
Age (years) Mean (SD)	43.4 (12.8)
Education (years) Mean (SD)	12.5 (3.2)
Recency of last episode, (≤1 year)%	160 (14.3%)
QIDS score, Mean (SD)	5.4 (3.7)
Chronic Pain Grade (CPG)	
CPG 0-1	720 (64.2%)
CPG 2	246 (21.9%)

CPG 3	101 (9.0%)
CPG 4	54 (4.8%)

References

Gerrits, M. M., van Oppen, P., Leone, S. S., van Marwijk, H. W., van der Horst, H. E., & Penninx, B. W. (2014). Pain, not chronic disease, is associated with the recurrence of depressive and anxiety disorders. *BMC Psychiatry*, *14*(1). https://doi.org/10.1186/1471-244X-14-187

Penninx, B. W. J. H., Beekman, A. T. F., Smit, J. H., Zitman, F. G., Nolen, W. A., Spinhoven, P., … For the NESDA Research Consortium. (2008). The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *International Journal of Methods in Psychiatric Research*, *17*(3), 121–140. https://doi.org/10.1002/mpr.256

NESDA ANALYSIS PLAN

Please e-mail completed form to Marissa Kok, NESDA study, A.J. Ernststraat 1187, 1081 HL Amsterdam. E-mail: ma.kok@ggzingeest.nl

1. First author information:

Name of first author: Lizbeth Burgos Ochoa

E-mail address:l.burgosochoa@vumc.nl

Telephone: 0626271674

Site: □ UL/LUMC □ UMCG/RUG ■ VU/VUMC

□ Other:

Which NESDA senior investigator will supervise?:

Brenda W. Penninx

If there is specific funding for the conduct of data analyses, please mention:

There is no specific funding for conducting the data analyses.

2. Working title of plan:

Comparison of methods for statistical mediation analysis with survival data.

3. Give a brief summary of your analysis plan that includes the following:

a. Research question

What is the performance of different methods for mediation analysis with survival data under data conditions commonly encountered in epidemiological research practice?

b. Brief background and rationale for addressing the research question in NESDA

Statistical mediation analysis, which has been widely discussed in the field of psychology after Baron and Kenny's influential paper [1], has recently raised interest from epidemiologists. Mediation analysis can be used to identify underlying mechanisms that lead of exposure-outcome relationships [2]. In epidemiologic research these outcomes of interest often are a time-to-event variable (survival time). This type of outcome variable provides information about the time until an event occurs (treatment response or adverse event) and is common in fields like oncology and cardiology. For example, to investigate the effect of an intervention on the survival of patients with heart disease or to explore the clinical effectiveness of a cancer treatment given the endpoint of overall patient survival.

Nonetheless, guidance for researchers interested in performing mediation analysis with survival time variables is limited [2].

The traditional approach to estimate indirect effects in mediation analysis is based on three regression models. In the first model the total effect of the exposure on the outcome is estimated, i.e. the *c* coefficient. In the second model the effect of the exposure on the mediator is estimated, i.e. the *a* coefficient. In the third model the direct effect of the exposure on the outcome, i.e. the *c'* coefficient, and the effect of the mediator on the outcome, i.e. the *b* coefficient, are estimated. Subsequently, the indirect effect can be computed as the product of the *a* and *b* coefficients, also known as the product-of-coefficients method $(\widehat{ab})$ or as the difference between the total effect $\hat{c}$ and the direct effect $\widehat{c'}$, also known as the difference-between-coefficients method $(\hat{c} - \widehat{c'})$.

For modeling survival data, Cox proportional hazards (PH) regression is the most commonly used method. However, its use incurs into a problem while performing mediation analysis. When the mediator and outcome are both continuous with normally distributed error terms, the product of coefficients $\left(\widehat{ab}\right)$ and the difference between coefficients $\left(\hat{c} - \widehat{c'}\right)$ approaches for calculating indirect effects, will yield identical estimates [5]. Yet, this is not the case when the outcome is time-to-event. For example, in the study of Gerrits et al. [6], based on NESDA data, the product of coefficients approach was used to calculate the indirect effect based on Cox PH regression. However, the indirect effect and direct effect do not add up to the total effect, showing that the results from the product-of-coefficients approach and the difference-between-coefficients approach are mathematically inconsistent. Because of this mathematical inconsistency, it is not evident which of the values resulting from these two methods can be interpreted as the indirect effect [2].

The mathematical inconsistency of the product-of-coefficients and difference-between-coefficients approaches happens because in a Cox model the residual variance is fixed. Consequently, the scale of the coefficients in nested Cox regression models will differ resulting in the mathematical inconsistency of $\widehat{ab}$ and $\hat{c} - \widehat{c'}$.

Throughout the years several methods for mediation analysis with survival data have been developed, such as the method based on the counterfactual framework, the semiparametric robust estimation method and the dynamic path analysis approach. Though some of the previously mentioned methods are starting to be used in research, e.g. counterfactual approach [13, 14], it is not clear yet, whether any of them would be more suitable for situations encountered in practice. So far, they have been illustrated by their authors through a single dataset and, to the best of our knowledge, they have not been compared between each other in a formal manner using simulation studies.

The aim of this study is to evaluate and compare the performance of different methods for mediation analysis and survival analysis under conditions that reflect situations encountered in practice. For that purpose the study from Gerrits et al. [6], which is based on data from the Netherlands Study of Depression and Anxiety (NESDA), will serve as basis for the statistical analyses. We intend to work with this data, because it is an example of what can be encountered in common clinical research, its topic and results are relevant for the clinical field, and contains the information necessary to conduct the statistical analyses.

c. Describe what has been done in NESDA in this area, and how this current plan extends earlier findings and does not have overlap with earlier analyses/papers

The purpose of the study from Gerrits et al [6], was to examine to what extent chronic disease in general and specific diseases and pain symptoms are associated with recurrence of depressive and anxiety disorders in patients with a prior history of a depressive and/or anxiety disorder in a major longitudinal study. Furthermore, by applying the product of coefficients approach based on Cox PH regression, the authors aimed to determine whether subthreshold depressive and anxiety symptoms partly mediate associations between chronic diseases and pain, on the one hand, and recurrence of depressive and anxiety disorders, on the other. The obtained results showed that subthreshold depressive symptoms mediated the associations between pain and depression recurrence. The indirect effects were significant for various types of pain, a higher number of pain locations and higher Chronic Pain Grade, suggesting that there is an overall effect of the pain variables on depression recurrence through aggravated subthreshold depressive symptoms.

As mentioned in the background information, various methods for mediation analysis with survival data have been developed the last years. The aim of our study is to methodologically investigate which method for mediation analysis with survival data is the most accurate for estimating the mediated effect. We would like to use the mediation analysis as presented in Gerrits et al as an empirical data example in our paper. Furthermore, the coefficients from the mediation analysis performed by Gerrits et al. will be used as input for the Monte Carlo simulation studies that will be done to assess the performance of the available methods for mediation analysis.

d. Variables and Wave of data collection to be used in main analysis (the main predictor and outcome variables must be identified)

The wave of data collection that will be used for the analysis corresponds to the same used in the study from Gerrits et al. As described by the authors, the baseline data collection took place between 2004 and 2007, with follow-up assessments, two and four years later. Also, the variables used will be same as in the previously mentioned study,

a) two exposure variables: *chronic disease* and *pain*,
b) continuous-time survival outcome: *recurrence of depressive and/or anxiety disorder*,

c) mediator variable: *subthreshold symptoms of depression and anxiety,*

d) covariates: *age, gender, and number of years in education, recency of the last episode of depressive or anxiety disorder and medication.*

e. Outline of analyses

First, a series of Monte Carlo simulations will be conducted to evaluate and compare the performance of different methods developed to integrate mediation and survival analysis under conditions that reflect situations encountered in practice.

A single continuous mediator model will be investigated with a discrete or continuous exposure variable and a continuous-time survival variable as outcome. Different mediation scenarios will be modeled. The two most common types of censoring in clinical research will be taken into account (censoring due to dropout or study duration). The coefficients from the model based on the NESDA empirical dataset (Gerrits et al.) will be used as reference for the input values of the simulated datasets.

The assessed methods will be the approach based on the counterfactual framework [10], the semiparametric robust estimation method [11] and the dynamic path analysis approach [12]. The performance of these methods will be assessed based on their parameter estimates (bias), standard errors, confidence intervals (along with coverage probabilities), power and Type 1 error.

Once these methods have been tested in a formal manner they will be applied to the NESDA empirical dataset to evaluate their performance in what could be usual research practice. After this, the next goal will be issuing recommendations for applied researchers in the epidemiological and medical field about the circumstances under which these methods could be more efficiently applied.

All the statistical simulations and analyses will be conducted using the software R [15].

4. Proposed authors:

Lizbeth Burgos Ochoa

Judith J.M. Rijnhart

Jos W.R. Twisk

Martijn W. Heymans

Brenda W. Penninx

5. **Timeline for completion and submission of the manuscript:**

   October-December 2017: Literature review, simulation protocol.

   January-March 2018: Run simulations and tests on the empirical dataset.

   March-April 2018: Writing process of results and discussion.

   May 2018: Submission of manuscript

I hereby state that I will use the data only for addressing the research question described in point 3, and not for other purposes unless I submit a new analysis plan.

Signed Date

Lizbeth Burgos Ochoa 25-10-2017

References

1. Baron RM, Kenny DA. The Moderator-Mediator variable distinction in Social Psychological research: Conceptual, strategic, and statistical considerations. J Pers Soc Psychol. 1986;51:1173–82.

2. Gelfand LA, Mackinnon DP, Derubeis RJ, Baraldi AN. Mediation Analysis with Survival Outcomes: Accelerated Failure Time vs . Proportional Hazards Models. Front Psychol. 2016;7 March:1–10.

3. Sedlis SP, Pamela M. Hartigan, Koon K. Teo, David J. Maron, John A. Spertus, John Mancini, et al. Effect of PCI on Long-Term Survival in Patients with Stable Ischemic Heart Disease. N Engl J Med. 2015;20373:1937–46. doi:10.1056/NEJMoa1505532.

4. Gustave Roussy I, Molina A, Kheoh T, Haqq CM, Research J, Fizazi K, et al. Abiraterone acetate for treatment of metastatic castration-resistant prostate cancer: fi nal overall survival analysis of the COU-AA-301 randomised, double-blind, placebo-controlled phase 3 study. Lancet Oncol. 2012;13:983–92. doi:10.1016/S1470-2045(12)70379-0.

5. MacKinnon DP, Warsi G, Dwyer JH. A simulation study of mediated effect measures. Multivariate Behav Res. 1995;30:41–62.

6. Gerrits MMJG, Oppen P Van, Leone SS, Marwijk HWJ Van, Horst HE Van Der, Penninx BW. Pain , not chronic disease , is associated with the recurrence of depressive and anxiety disorders. BMC Psychiatry. 2014;14:1–9.

7. Ngwa JS. Comparing methods for modeling longitudinal and survival data , with consideration of mediation analysis. Boston University; 2013. https://open.bu.edu/handle/2144/15210.

8. Tein J, Mackinnon DP. Estimating mediated effects with survival ata. In: Yanai H, AO R, K S, Y K, JJ M, editors. New Developments in Psychometrics: Psychometrics Society Proceedings. Tokyo: Springer-Verlag; 2003. p. 405–12.

9. Lange T, Hansen J V. Direct and Indirect Effects in a Survival Context. Epidemiology. 2011;22.

10. Lange T, Vansteelandt S, Bekaert M. A Simple Unified Approach for Estimating Natural Direct and Indirect Effects. Am J Epidemiol. 2012;176:190–5.

11. Tchetgen Tchetgen EJ. On Causal Mediation Analysis with a Survival Outcome. Int J Biostat. 2011;7:1–38. doi:10.2202/1557-4679.1351.

12. Strohmaier S, Røysland K, Hoff R, Borgan O, Pedersen TR, Aalen OO. Dynamic path analysis - a useful tool to investigate mediation processes in clinical survival trials. Stat Med. 2015;34:3866–87.

13. Nordahl H, Rod NH, Frederiksen BL, Andersen I, Lange T, Diderichsen F, et al. Education and risk of coronary heart disease: Assessment of mediation by behavioral risk factors using the additive hazards model. Eur J Epidemiol. 2013;28:149–57.

14. Rochon J, Bois A du, Lange T. Mediation analysis of the relationship between institutional research activity and patient survival. BMC Med Res Methodol 2014 141. 2014;14:9.

15. R-Core-Team. R: A language and environment for statistical computing. 2017.