Master Statistics and Data Science

Proposal for the Thesis Project

MSc Statistics and Data Science
Leiden University

Period of the project
Carrying out the project will account for 30 ECTS, which is equivalent to 21 weeks of 40 hours.

- Intended period of the project: Feb 4th - July 15th (Defense: before August 15th)
- Intended date for the Midterm Progress Meeting: April 10th

Student
Name: Yixiao Tang
Student number: s3855422
e-mail (U-mail): y.tang.9@umail.leidenuniv.nl

Medical Supervisor
Name: Jan H.N. Lindeman
e-mail: j.h.n.lindeman@lumc.nl
Institute: Leiden University Medical Center (Depts. Vascular and Transplantation Surgery)
Function: Provide guidance on research progress.

Statistics Supervisor
A second supervisor from the master of Statistics and Data Science is needed when the daily supervisor is no staff member of the departments/Institutes of the master.
Name: Hein Putter
e-mail: h.putter@math.leidenuniv.nl
Institute: Leiden University Medical Center (Depts. Medical Statistics and Bioinformatics) & Mathematics Institution
Function: Offer expertise on statistical methods.

EMOS (European Master in Official Statistics)
project (no)

Intended independent reader (to be filled in by Statistics and Data Science supervisor)
Each thesis is judged independently by a second member of the Statistics and Data Science organizations. The independent reader is not involved in the research of the project and has no direct hierarchical relation with the supervisor(s).

Name:
e-mail:
Institute:
Function:

1 Summary

1.1 Title and abstract (max 150 words)

A quantitative analysis strategy for time-to-event analysis with composite endpoints

Traditional time-to-event analyses with composite endpoints based on the first event, face challenges such as the non-proportional hazards assumption, difficulties in relative risk comparisons, competing risks, and equal weighting of events with differing clinical impacts.

This study proposes a novel approach using Restricted Mean Survival Time (RMST) as the estimator to overcome non-proportional hazards assumption and provide a robust quantitative measurement for evaluating the treatment effect. For clinical benefit analysis, Quality-Adjusted Life Years (QALYs) will be incorporated to weigh different endpoints based on their impact on quality of life and address competing risks by assessing cumulative quality-adjusted survival rather than focusing solely on the first event. For cost-effectiveness analysis, treatment costs will also be converted to evaluate the economic value of the intervention.

The method will be applied to data from the FOURIER trial, to provide a more precise and clinically meaningful analysis of cardiovascular outcomes.

1.2 ECTS Justification for the preparation of the Thesis Proposal (max 500 words)

The investment for the study and writing of this thesis proposal should be 4 ECTS. Please clarify how many hours were spent on the activities needed for the writing of this thesis proposal.

Write down here your justification of your hours for this thesis proposal, know that you can also include hours spent on further improvement of academic skills (For example presentation workshop / writing workshop or module etc.).

No need for justification.

2 DESCRIPTION OF THE PROPOSED RESEARCH PART OF THE THESIS PROJECT

Write a concise proposal of a maximum of 1200 words. This should contain:

2.1 The Research Problem

Traditional time-to-event analyses with composite endpoints have several limitations. These include:

1. **Non-proportional hazards assumption**: Many clinical trials face issues with hazard ratios that vary over time, reducing the reliability of proportional hazard models.
2. **Relative risk**: The hazard ratio is a relative measurement of an outcome that is not comparable and can’t be intuitively understood by patients. And it is not comparable among different studies.
3. **Equal weighting of events**: Composite endpoints treat all events (e.g., nonfatal myocardial infarction, stroke, cardiovascular death) equally, despite their differing impacts on patients' quality of life.
4. **Competing risks**: Different endpoints are competing, once one of the composite endpoints occurs, there is no chance for the second event to be recorded.

2.2 Research aims

This study aims to address the limitations of traditional composite endpoint analyses by:

1. **Applying the Principles of Restricted Mean Survival Time (RMST)** as a quantitative and clinically relevant outcome measure for composite endpoints in cardiovascular trials.
2. **Developing a QALY-based weighting way** for evaluating composite endpoints, incorporating the impact of each event on patients' quality of life.
3. **Testing this novel methodology** using data from the FOURIER trial to provide a more comprehensive evaluation of evolocumab's treatment effect.

2.3 Research plan

Reconstruct Dataset
Create a reconstructed dataset for the Fourier trial with longitudinal follow-up time and simulate the sequential events table.

Event Weighting Using QALYs
Assign QALY weights to each type of clinical event (e.g., nonfatal myocardial infarction, stroke, cardiovascular death) based on their relative impact on quality of life. The weights will be determined by consulting existing literature and expert opinion.

Outcome Analysis with Restrict Mean Survival Time (RMST):
RMST will be used to estimate the average survival time between the evolocumab and placebo groups over a predefined time horizon.

1. **Clinical Benefit Analysis**
   The RMST will be adjusted for quality of life by applying QALY weights, providing a measure of quality-adjusted survival. The difference in quality-adjusted RMST will quantify the clinical benefit of evolocumab compared to placebo.

2. **Cost-Effectiveness Analysis**
   Combines QALY weights with treatment costs to evaluate whether the intervention provides good value for money.

Comparison with Traditional Hazard Ratio Analysis
Compare RMST-based findings with traditional hazard ratio analyses to highlight differences and advantages.

2.4 Expected Results/end product:

QALY-weighted Restrict Mean Survival Time Estimation of the time-to-event trails.

2.5 Reference list

1. Sabatine, M. S., Giugliano, R. P., Keech, A. C., Honarpour, N., Wiviott, S. D., Murphy, S. A., Kuder, J. F., Wang, H., Liu, T., Wasserman, S. M., Sever, P. S., & Pedersen, T. R. (2017). Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. *The New England Journal of Medicine*, *376*(18), 1713–1722. https://doi.org/10.1056/NEJMoa1615664

2. Erviti, J., Wright, J., Bassett, K., Ben-Eltriki, M., Jauca, C., Saiz, L. C., Leache, L., Gutiérrez-Valencia, M., & Perry, T. L. (2022). Restoring mortality data in the FOURIER cardiovascular outcomes trial of evolocumab in patients with cardiovascular disease: a reanalysis based on regulatory data. BMJ Open, 12(12), e060172–e060172. https://doi.org/10.1136/bmjopen-2021-060172

3. D'Agostino, J. (2000). Debate: The slippery slope of surrogate outcomes. *Current Controlled Trials in Cardiovascular Medicine*, *1*(2), 76–78. https://doi.org/10.1186/CVM-1-2-076

4. Wouters OJ, Naci H, Samani NJ.QALYs in cost-effectiveness analysis: an overview for cardiologists.Heart 2015;101:1868-1873.https://doi.org/10.1136/heartjnl-2015-308255

5. McCaw, Z. R., Yin, G., & Wei, L.-J. (2019). Using the Restricted Mean Survival Time Difference as an Alternative to the Hazard Ratio for Analyzing Clinical Cardiovascular Studies. *Circulation (New York, N.Y.)*, *140*(17), 1366–1368. https://doi.org/10.1161/CIRCULATIONAHA.119.040680

6. DOLAN, P., METCALFE, R., MUNRO, V., & CHRISTENSEN, M. C. (2008). Valuing lives and life years: anomalies, implications, and an alternative. *Health Economics, Policy and Law*, *3*(3), 277–300. https://doi.org/10.1017/S1744133108004507

7. Wei, Y., & Royston, P. (2017). Reconstructing Time-to-event Data from Published Kaplan–Meier Curves. The Stata Journal, 17(4), 786-802. https://doi.org/10.1177/1536867X1801700402

8. Putter, H., Fiocco, M., & Geskus, R. B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*, *26*(11), 2389–2430. https://doi.org/10.1002/sim.2712

3 ACTIVITIES TO GET EXPERIENCE AS A WORKING STATISTICIAN/DATA SCIENTIST IN A POSSIBLE FUTURE WORKING ENVIRONMENT

Formulate, together with your supervisor(s), three concrete working activities in which you will participate during your thesis project. One of these activities should involve developing communication skills. A list of possible activities can be found on Brightspace and in the information for supervisors, but it is allowed to formulate other activities. Give for each of the activity a short description and a short motivation for choosing the activity.

3.1 Activity 1: Survival Lunch Speech

- Short description of the activity (max 150 words)

The Survival Lunch is a bi-weekly seminar organized by the Survival Team of the Medical Statistics Department at LUMC. During each session, a speaker presents their research, followed by an open discussion among participants, which includes professors, medical statisticians, PhD students, and a few master's students. As part of my thesis project, I will prepare and deliver a 20-30 minute presentation on my research topic, explaining key statistical methods, findings, challenges, and application scenarios. This will be followed by a Q&A session where participants can provide feedback and engage in discussion.

- Motivation for choosing the activity (max 150 words)

This activity will help me develop my presentation skills, particularly in presenting complex statistical concepts to an audience with diverse expertise. I will improve my ability to structure a talk, clearly explain methodologies, and respond to critical questions. Additionally, receiving feedback from experienced statisticians and researchers will enhance my understanding of the strengths and limitations of my work.

3.2 Activity 2: Vascular Research Meeting Speech

- Short description of the activity (max 150 words)

The Vascular Research Meeting is a regular seminar where clinicians and medical researchers discuss ongoing vascular research projects. As part of my thesis project, I will prepare and deliver a presentation about the clinical application (FOURIAL Trial) of the proposed method, focusing on the statistical methodology used to address it. The presentation will include a discussion on the strengths and weaknesses of existing approaches and potential improvements. Feedback from clinicians will help understand how the real application field needs.

- Motivation for choosing the activity (max 150 words)

Presenting in front of an interdisciplinary audience will improve my ability to translate statistical findings into clinically meaningful insights. This experience will be essential for working as a statistician in a medical research setting, where clear communication and methodological advancements play a critical role.

3.3 Activity 3: Collaborate with Clinicians

- Short description of the activity (max 150 words)

This activity involves direct collaboration with a clinical department that works with patient data. My role will be to introduce a new statistical method that can be applied to their research, helping them analyze clinical data in an alternative way. This process includes understanding their research needs, explaining the benefits of the proposed method, and assisting with its implementation.

- Motivation for choosing the activity (max 150 words)

Working with clinicians will provide valuable experience in translating statistical methods into practical applications. I will gain insight into the challenges of working with real-world medical data. This collaboration will also enhance my problem-solving and interdisciplinary communication skills, ensuring that statistical findings are accessible and useful to non-statisticians. By bridging the gap between statistical theory and clinical practice, I will develop key skills for a career in medical data science and biostatistics.

4 WORK PLAN AND SUPERVISION

4.1 Supervision

Describe the arrangements regarding the type and frequency of meetings between student and daily supervisor(s) and on roles and responsibilities. If there is a second supervisor from Statistics & Data Science, also describe type and frequency of communications between the second supervisor, the daily supervisor and the student.

Medical Supervisor (Jan): Provide guidance on research progress, review drafts, and ensure the research addresses key clinical questions and contributes meaningfully to the field.

Statistics Supervisor (Hein): Offer expertise on statistical methods, review drafts of the technical part, and ensure methodological rigor.

Student (Yixiao): Conduct literature reviews, data analyses, and manuscript preparation, seeking insights from supervisors.

We plan to have weekly meetings every Tuesday for 1 hour to update progress, problems encountered, and action plans for next week.

4.2 Time Schedule

Carrying out the thesis project should take 30 EC (exclusive 4 EC to write this proposal). Present a feasible time schedule of your activities. Note that 30 EC corresponds to 21 weeks full time work. Make a detailed plan (week by week) so that at each supervisor-student meeting, it can be discussed if things are still going as planned, and if not, how to tackle that. Be aware that writing takes time. Indicate what elements can be cut / reduced if necessary.

February
Week 1-4: Thesis proposal. Conduct literature reviews. Data reconstruction and sequential events simulation.

March
Week 5-8: Implement QALY-based weighting for composite endpoints and Conduct RMST analyses.

April
Week 9-12: Conduct Clinical benefit and cost-effectiveness analyses and compare results with traditional methods. (Midterm Progress Meeting in this month)

May
Week 13-16: Interpret results and draft initial sections of the manuscript. Share drafts with supervisors for feedback. Corroborate with clinicians to introduce and apply the method on their research.

June
Week 17-20: Refine the manuscript based on feedback. Prepare visualizations and presentations for dissemination. Give a speech in the Cardiovascular department.

July
Week 21: Finalize the manuscript and submit it for review. Give a speech in the Medical Statistics & Bioinformatics department.

4.3 Infrastructure

Describe the arrangements offered to the student to facilitate the students' work progress (For example, guest employment, a desk, shared office, computer, access to a computing server)

A shared office and monitor offered by the Medical statistics department.

4.4 Other Courses / Activities:

- What courses (how many ECTS) still need to be obtained during the thesis project before graduation. Please adapt your time schedule to incorporate this.
- Are there other reasons that may make it impossible to spend ±21 consecutive weeks on the thesis project

No other courses are needed.

5 AGREEMENT PAGE

The supervisors and student hereby declare that they agree to the arrangements in this proposal.

- The supervisors and student hereby declare that they have applied, and will apply good scientific practices, that follow the University Academic Integrity Regulations and the Ethical Guidelines from Statistical Practice . When in conflict with each other, the University Academic Integrity Regulations should be followed.

- The supervisors hereby declare that they are aware that all supervisors, the independent reader, and the board of examiners and panel member of independent audits should be able to assess the complete thesis. Furthermore all supervisors and independent reader should be able to view the corresponding programming code.

- The student hereby declares that he has provided the supervisor with the Documentation for Supervisors

- The student hereby declares that both this proposal and its resulting thesis, will be free of plagiarism (cf. Rules and Regulations of the Board of Examiners).

	Name	**Signature**	**Date**
Student	**Yixiao Tang**		
Supervisor	**Jan Lindeman**		
	Hein Putter		

This signed proposal should be submitted to the Thesis Committee by uploading it in Brightspace. In case of problems please contact thesis@stat.leidenuniv.nl.