



Universiteit
Leiden
The Netherlands

Statistical methods for mass spectrometry-based clinical proteomics

Kakourou, A.A.

Citation

Kakourou, A. A. (2018, March 8). *Statistical methods for mass spectrometry-based clinical proteomics*. Retrieved from <https://hdl.handle.net/1887/61138>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/61138>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/61138> holds various files of this Leiden University dissertation

Author: Kakourou, Alexia

Title: Statistical methods for mass spectrometry-based clinical proteomics

Date: 2018-03-08

Statistical Methods for Mass Spectrometry-based Clinical Proteomics

Alexia Artemis Kakourou

Cover design: Dimitris Sevastakis & BIBAIOTEXNIA, Athens

Printing: Neoanalyssis, Athens

Copyright © by 2018 Alexia Kakourou. All rights reserved. No part of this publication may be reproduced without prior permission of the author.

ISBN: 978-960-93-9733-9

Research leading to this thesis was funded by the EU under the FP7 Marie Curie Action MEDIASRES (Novel Statistical Methodology for Diagnostic Prognostic and Therapeutic Studies and Systematic Reviews), (Grant/Award Number: FP7/2011/290025) and MI-MOmics (Methods for Integrated Analysis of Multiple Omics Datasets), (Grant/Award Number: FP7/Health/F5/2012/305280).

Statistical Methods for Mass Spectrometry-based Clinical Proteomics

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 8 maart 2018
klokke 15:00 uur

door

Alexia Artemis Kakourou
geboren te Athene in 1986

Promotor:

Prof. dr. J. J. Houwing-Duistermaat

Co-Promotor:

Dr. B. J. A. Mertens

Leden promotiecommissie:

Prof. dr. W. Vach, Institute of Medical Biometry and Statistics, University of Freiburg, Freiburg, Germany

Dr. R. Pheiffer, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, United States

Prof. dr. J. J. Goeman

Dedicated to my family for their unconditional love and support.

Πιάσε την αστραπή στον δρόμο σου,
άνθρωπε, δώσε της διάρκεια, μπορείς!
*Catch the lightning as you go,
man, make it last ,you can!*

– *Odysseus Elytis, The lifelong moment, 1978*

Acknowledgements

Hereby I would like to acknowledge all the important people who contributed, each one in their own way, to the successful completion of my research thesis and PhD journey.

I am deeply grateful first of all to my supervisor Dr. Bart Mertens for offering me the opportunity to “seize the day”, introducing me to the field of Clinical Biostatistics and supporting me all the way to the end with continuous motivation and inspiration. Bart, I am honored to be your first PhD student. I would equally like to express my gratitude to my co-supervisor Prof. Werner Vach for his invaluable guidance and fruitful feedback throughout these years of challenge and accomplishments. I am also thankful to my promotor Prof. Jeanine Houwing-Duistermaat for her support and interest in my work, as well as for offering me the model of a successful woman in the field of Biostatistics.

In addition, I would like to thank my reading committee members Dr. Ruth Pfeiffer and Prof. Jelle Goeman for their time, their interest in this work and their positive feedback.

Thanks are owed to my colleagues from the Department of Medical Statistics and Bioinformatics who have largely contributed to my growth as a researcher and scientist by accepting me among them and creating a fruitful milieu for the continuous exchange of ideas and expertise. Special thanks go to all my colleagues with whom I shared office. Roula, Rosa, Mia, Giorgos, Carlo, Markus, Irene, Ning-Ning and Jesse, thank you for your friendship, support and tolerance throughout these years of stress. I would also like to thank a number of individuals, within and outside the Department of Statistics who played a distinct role in my PhD pursuit and in my life in Leiden in general: Thank you Bruna, Renaud, Anna, Theodor, Eleni, Katerina, Mar, Lora, Zhenia, Roberta and Benjamin.

My gratitude also goes to Wilma Mesker for believing in me and offering me the opportunity to work with her as an applied statistician in such an inspiring and challenging, multi-disciplinary clinical research project in the Department of Surgery. Many special thanks to my Mass Spectrometry experts from the Centre of Proteomics and Metabolomics Yuri van der Burg and Simone Nicolardi for deepening my understanding of the proteomic data world.

I would also like to acknowledge all the people from MEDIASRES with whom I had the pleasure to work alongside and shared great moments of training and achievements. Dear MEDIASRES fellows: Anna B., Mia, Corine, Leyla, Soheila, Susanne, Hong, Sung-Won, Anna W., Markus, Matteo, Federico and Ketil, thank you for the trip

we took together and good luck to all of you.

Another important person I would like to thank is Dimitris for his essential contribution to the cover design of this thesis and for enhancing the last phase of my PhD endeavor.

Last but not least, I want to express my gratitude to my family and an important person for embracing me with unconditional love, continuous encouragement and endless motivation. To Giorgos for his unlimited support, understanding and much more. To my mum Kitty for her psychological and practical support throughout this journey, my dad George and my sisters Natalie and Lydia for always being there for me, my grandparents for believing in me and for setting the foundation of my education. Σας χρωστάω ότι είμαι.

Table of Contents

Acknowledgements	vii
1 Introduction	1
1.1 Introduction	1
1.2 Data acquisition	2
1.3 Data pre-processing	4
1.4 Limit of detection (LOD)	5
1.5 Proteomic prediction	6
1.5.1 Types of models	7
1.5.2 Construction of diagnostic rules	8
1.5.3 Assessment of diagnostic rules	9
1.6 Outline of the thesis	10
2 Combination of prediction rules for proteomic diagnosis	15
2.1 Introduction	16
2.2 Combination Method	18
2.2.1 Convex Combination via Linear Mixtures	19
2.2.2 Model-based Combination	19
2.3 Application and Analysis	21
2.3.1 Model Choice	21
2.3.2 Results	23
2.3.3 Post-hoc Analysis	24
2.4 Simulation Study	29
2.5 Discussion	35
3 Combination of omics data for prediction of binary outcomes	39
3.1 Introduction	40
3.2 Methods	41
3.2.1 Double cross-validation prediction	41
3.2.2 Parallel combination of predictions	42
3.2.3 Sequential combination of prediction	44
3.3 Performance evaluation	44
3.3.1 Calibration measures	45

3.3.2	Discrimination measures	45
3.4	Application	46
3.4.1	Data presentation	46
3.4.2	Model choice: Logistic regularized regression	47
3.4.3	Results	48
3.5	Summary and discussion	53
4	Accounting for isotopic clustering in Fourier transform MS data analysis	55
4.1	Introduction	56
4.2	Materials and Methods	57
4.2.1	Data description	57
4.2.2	Identification Algorithm - Data preprocessing	59
4.2.3	Summary measures	61
4.3	Application and analysis	65
4.3.1	Identification algorithm implementation	65
4.3.2	Model fitting and results	66
4.4	Discussion	73
5	Bayesian variable dimension logistic regression with paired proteomic measurements	77
5.1	Introduction	78
5.2	Data	80
5.3	Bayesian variable-selection model on intensity-shape pairs	81
5.3.1	The logistic regression model	81
5.3.2	Variable-dimension logistic regression model	81
5.3.3	Prior specification	82
5.3.4	MCMC model fitting	83
5.4	Results	86
5.4.1	Application	86
5.4.2	Convergence	86
5.4.3	Post-hoc analysis	86
5.4.4	Assessment of predictive performance	90
5.4.5	A simulation example	92
5.5	Discussion	95
6	Adapting censored regression methods to adjust for the LOD in proteomic diagnosis	105
6.1	Introduction	106
6.2	Data	107
6.2.1	Data description	107
6.2.2	Data structure and limit of detection (LOD)	108
6.3	Methods	110
6.3.1	Censored regression	110

6.3.2	Random effect censored regression	111
6.3.3	Random effect censored regression applications	113
6.4	Application and analysis	114
6.4.1	Model choice	114
6.4.2	Model fitting and performance measures	115
6.4.3	Results	116
6.4.4	Variable selection	118
6.5	Discussion	121
6.6	Conclusion	123
Bibliography		131
Summary		137
Samenvatting		143
List of Publications		149
Curriculum Vitae		151

