BAYESIAN METHODS TO ADDRESS MULTIPLE COMPARISONS AND MISCLASSIFICATION BIAS IN STUDIES OF OCCUPATIONAL AND ENVIRONMENTAL RISKS OF CANCER

A thesis by publications presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Public Health

Massey University, Wellington, New Zealand

Marine Corbin

2013

Abstract

In this thesis I explore the application of several Bayesian approaches, implemented with standard statistical software, in environmental and occupational epidemiology. These methods are applied to case-control studies of occupational risks for lung and upper aerodigestive tract cancers conducted in New Zealand and Europe. The findings are of interest in themselves, but the focus of the thesis is on the application of Bayesian methods to produce these findings. It is not intended to represent a comprehensive overview of all Bayesian methods, but rather to explore Bayesian methods which are most appropriate for the studies which are presented here.

In the first section, I review the underlying theory involved in such analyses.

In the second section, I use Bayesian methods to address the problem of multiple comparisons. In occupational case-control studies, we may collect information on hundreds of occupations/exposures for which there is little or no prior evidence. For those occupations/exposures, we get a false positive finding by chance about 5% of the time. This means that if we repeat the study in a new population, these chance associations are likely to exhibit 'regression to the mean' and will not show such extreme risks again. Bayesian methods can be used to 'shrink' effect estimates based on how strong the regression to the mean is likely to be.

In the third section, I use Bayesian methods for assessing and correcting systematic error. Although the methods I use can be applied to several situations (selection bias, misclassification, residual confounding), I apply them to the specific situation of

misclassification of the main exposure. In particular, I apply four different methods for such sensitivity analyses: multiple imputation for measurement error (MIME); imputation based on specifying the sensitivity and specificity (SS), Direct Imputation (DI) of the ‘true’ exposure using a regression model for the predictive values and imputation based on a fully Bayesian analysis.

I conclude by summarising the strengths, limitations, and areas of future development for the use of these methods. It is anticipated that, in 5-10 years time, such analyses may become standard supplements to ‘traditional’ forms of analysis, i.e. that Bayesian methods may be routinely used, and may form part of the ‘epidemiological toolkit’ for assessing and correcting for both random and systematic error.

Author's declaration

This thesis was produced according to Massey University's "Thesis-by-Paper" requirements. That is, it is based on research that is published, in-press, submitted for publication, or is in final preparation for submission. Each individual chapter is set out in the style of the journal to which it has been submitted. Consequently, some of the submitted chapters are relatively succinct, there is some repetition (particularly in the Methods sections) and there are small stylistic differences between chapters. To supplement the relative brevity of some of the chapters, the appropriate sections of the background and methods chapter have been extended.

I have stated my contribution to each chapter in Appendix IV.

Acknowledgements

First of all I would like to thank both of my supervisors Neil Pearce and Milena Maule for their constant guidance, support and faith in me during this long adventure and through the distance.

Neil, thank you for welcoming me in New Zealand and at Centre for Public Health Research (CPHR) and for giving me the chance to embark on this PhD. Thanks for your advice and encouragement over the years and thanks for all the opportunities you gave me to extend my knowledge and experience. Thanks also for always finding the time and the ways to meet regularly and answer my questions, even though the different locations, internet connection and time differences did not always make it very easy.

Milena, grazie della tua amicizia e di essere sempre stata qua per me durante tutti questi anni, anche durante i primi mesi di vita di Matteo. Grazie di avermi dato la motivazione e di avermi incoraggiata a iniziare questo dottorato. Grazie del tuo immenso aiuto sia sul piano lavorativo che sul piano morale e di aver condiviso con me tutti i momenti alti e bassi nella realizzazione di questa tesi. Sei stata bravissima a saper ridarmi energia e fiducia ogni volta che ne avevo bisogno e ricorderò sempre sia le insalate di formule che tutte le risate insieme.

Thanks to all my workmates at CPHR for making me feel so quickly part of the 'family'. In particular, thanks to Jeroen Douwes for welcoming me back at CPHR for the last part of my PhD and for helping me through the examination process. Thanks to my mock examiners Jeroen, Steve Haslett, Laura Howe, Andrea 't Mannetje, Amanda Eng and Collin Brookes for their constructive comments. Thanks to Steve for his availability and for his very helpful guidance. Thanks to Amanda and Collin, my "mentor PhD students", for all their valuable advice and support. Thanks to Dave McLean, Andrea 't Mannetje, Soo Cheng and Fiona McKenzie for their help with the lung cancer study. Thanks to Mathu and Helene for their support and coaching and for our weekly quiz nights and a particular thank you Mathu for hosting me in your lovely apartment every time I came back to Wellington. Thanks to Katharine for being such a supportive roommate during the ultimate phase of this PhD. Thanks to Soo and Grace for keeping me going with the magic tiger balm and essential oils. Thanks to Kerry and Soo for the many rides home when I stayed late at work. Thanks also to Hilary for being always so helpful and thanks to Nathalie and Vicki for their help in the last minute rush.

Grazie ai miei colleghi dell'Unità di Epidemiologia dei Tumori per la loro accoglienza e per avermi viziata dal mio primo giorno a Torino. Innanzitutto grazie mille a Franco Merletti di avermi accolta prima come stagista e poi come dottoranda, di avermi spinta e indirizzata nella scelta di questo dottorato e di avermi dato tutte le opportunità possibili per condurre questo progetto. Grazie a Lorenzo Richiardi per il suo importante contributo a questa tesi e per i suoi consigli che mi hanno aiutata tante volte. Grazie a tutti i "stanzonesi" (Milena, Lorenzo, Daniela, Costanza, Emanuele e Enrica) per tutti i buoni momenti passati insieme e i tradizionali pranzi dagli "Oscar" che mi mancano. Grazie anche a Daniela Aimar di avermi ospitata durante alcune settimane nella sua casa.

I want to thank all my coauthors and in particular thanks to Sander Greenland and Kyle Steenland for their guidance and advice. Thanks also to Jonathan Bartlett for his help and input on Chapter VI.

My stay in New Zealand would not have been such a nice experience if I had not had a nice home to go to every night. Thanks to Carl Lin, Jacob, KC, Steve Mainwaring, Mousumi, Matilda and Swann for being such amazing flatmates.

I am grateful to all my friends for always staying in contact even through the distance. Un spécial gros merci à Claire, Morgane, Elena et Manue pour leurs visites à Wellington et/ou à Turin qui m'ont fait énormément plaisir.

I also wish to thank my wonderful family. Merci à Maman, Martin, Clémentine, Marjolaine, Corentin et Capucine de m'avoir soutenue et encouragée pendant toutes mes études. Merci d'avoir supporté mes crises de nerfs à chaque départ, quand je décidais de déballer ma valise cinq minutes avant de partir parce que j'avais oublié quelque chose. Merci aussi d'avoir fait le voyage tous les six à Turin et à Welllington. Merci à Papy et Mamie de m'avoir également soutenue et accompagnée dans tous mes projets. Merci de m'avoir emménagée et déménagée lors de tous mes déplacements en Europe (même sous la neige) et de m'avoir continuellement gâtée. Merci aussi de vos 2 consécutives visites en Nouvelle-Zélande ! Merci aussi à ma cousinette Hélène de m'avoir hébergée lors de mes visites à Londres.

Finally, Sebastián, without this PhD I would probably have never met you but without you I would probably have never managed to finish this thesis in time. Thanks for your support for the last few years and for always believing in me and thanks also for your very special care for the last months. ¡Mil gracias por todo!

Table of Contents

List of tables

List of figures

List of Abbreviations

CI	Confidence interval
CL	Confidence/Credibility limits
DI	Direct imputation of the ‘true’ exposure using a regression model for the predictive values
EB	Empirical Bayes
FBA	Fixed-parameter bias analysis
HR	Hierarchical regression
ISCO	International Standard Classification of Occupations
ISIC	International Standard Industrial Classification
logOR (or ln(OR))	log Odds Ratio
MCMC	Markov Chain Monte Carlo
MIME	Multiple imputation for measurement error
ML	Maximum likelihood
NACE	National Industrial Classification of All Economic Activities
NZSCO	New Zealand Standard Classification of Occupations
NZSEI	New Zealand Socio-Economic Index
OR	Odds Ratio
PBA	Probabilistic bias analysis
SB	Semi-Bayes
SI	Simulation Intervals
SL	Simulation Limits
SS	Imputation based on specifying the sensitivity and specificity
UADT	Upper aerodigestive tract