

***Statistical Issues Associated with the Analysis of
Binary Outcomes in Randomised Controlled Trials
when the Effect Measure of Interest is the Relative
Risk***

Lisa Yelland

BMa&CompSc (Hons)

*Discipline of Public Health
School of Population Health and Clinical Practice
Faculty of Health Sciences
The University of Adelaide
Australia*

*Thesis submitted in fulfillment of the requirements for the degree of Doctor of
Philosophy, September 2011*

Contents

Abstract.....	vi
Declaration	viii
Manuscripts Contributing to this Thesis.....	x
Presentations Arising out of this Thesis	xi
Awards Arising out of this Thesis.....	xii
Acknowledgements.....	xiii
Abbreviations	xv
1 Introduction	1
1.1 Problems with Odds Ratios	1
1.2 Advantages of Relative Risks.....	4
1.3 Problems with Relative Risks.....	5
1.4 Thesis Aim.....	6
1.5 Thesis Outline.....	6
2 Literature Review and Aims.....	8
2.1 Independent Data	8
2.1.1 Methods for Estimating Relative Risks	8
2.1.2 Comparison of Methods.....	11
2.1.3 Specific Aim 1	13
2.2 Clustered Data	13
2.2.1 Analysing Clustered Data	14
2.2.2 Simulating Clustered Binary Outcomes.....	18
2.2.3 Specific Aim 2	21

2.2.4	Methods for Estimating Relative Risks	21
2.2.5	Specific Aim 3	22
2.3	Perinatal Trials	22
2.3.1	Multiple Births.....	23
2.3.2	Specific Aim 4.....	24
3	Methods	25
3.1	Summarising Simulation Results	25
3.1.1	Common Approaches.....	25
3.1.2	Approaches Used in this Thesis	26
3.2	Computational Resources	28
3.3	Simulating Clustered Data.....	29
3.4	Avoiding Invalid Success Probabilities	31
4	Relative Risk Estimation in Randomised Controlled Trials: A Comparison of Methods for Independent Observations	34
4.1	Preface	34
4.2	Statement of Authorship.....	35
4.3	Article	36
4.3.1	Abstract.....	36
4.3.2	Introduction.....	36
4.3.3	Methods.....	37
4.3.4	Simulation Results.....	45
4.3.5	Example	61
4.3.6	Discussion.....	64
4.3.7	Recommendations	67
4.4	Additional Discussion.....	69

5	Adjusted Intraclass Correlation Coefficients for Binary Data: Methods and Estimates from a Cluster Randomised Trial in Primary Care	71
5.1	Preface	71
5.2	Statement of Authorship	72
5.3	Article	73
5.3.1	Abstract	73
5.3.2	Introduction	74
5.3.3	Methods	76
5.3.4	Results	81
5.3.5	Discussion	90
5.4	Additional Discussion	93
6	Relative Risk Estimation in Cluster Randomised Trials: A Comparison of Generalised Estimating Equation Methods	96
6.1	Preface	96
6.2	Statement of Authorship	97
6.3	Article	98
6.3.1	Abstract	98
6.3.2	Introduction	98
6.3.3	Methods	99
6.3.4	Simulation Results	105
6.3.5	Example	113
6.3.6	Discussion	116
6.4	Additional Discussion	121
7	Performance of the Modified Poisson Regression Approach for Estimating Relative Risks with Clustered Prospective Data	124
7.1	Preface	124

7.2	Statement of Authorship.....	125
7.3	Article	126
7.3.1	Abstract.....	126
7.3.2	Introduction.....	126
7.3.3	Methods.....	128
7.3.4	Results.....	130
7.3.5	Illustrative Examples.....	138
7.3.6	Discussion.....	141
7.4	Additional Discussion.....	143
8	Analysis of Binary Outcomes from Randomised Trials Including Multiple Births: When Should Clustering be Taken Into Account?	145
8.1	Preface	145
8.2	Statement of Authorship.....	146
8.3	Article	147
8.3.1	Abstract.....	147
8.3.2	Introduction.....	147
8.3.3	Methods.....	149
8.3.4	Results.....	154
8.3.5	Discussion.....	168
8.4	Additional Discussion.....	174
8.4.1	Relative Risks.....	174
9	Summary and Conclusions	185
9.1	Key Findings and Contributions	185
9.1.1	Specific Aim 1	185
9.1.2	Specific Aim 2	186

9.1.3	Specific Aim 3	187
9.1.4	Specific Aim 4	188
9.2	Limitations and Future Directions	189
9.3	Concluding Remarks.....	191
10	References.....	192

Abstract

Background: Binary outcomes have traditionally been analysed using logistic regression which estimates odds ratios. A popular alternative is to estimate relative risks using log binomial regression. Due to convergence problems with this model, alternative methods have been proposed for estimating relative risks. Comparisons between methods are limited and guidance on which method(s) should be used in practice is lacking. These methods are often applied to clustered data, despite the absence of evidence supporting their use in this setting.

Comparison of methods in the clustered data setting via simulation is difficult. The simulation model requires specification of the random effects variance on the log scale, but the intraclass correlation coefficient (ICC) on the probability scale is the preferred measure of dependence. The relationship between the ICC and the random effects variance has been defined under the logistic model but not the log binomial model.

The appropriate method for analysing binary outcomes from perinatal trials which include infants from multiple births is a matter of debate, and relative risks have received little attention in this context.

Aim: To investigate statistical issues associated with the analysis of binary outcomes in randomised controlled trials (RCTs) when the effect measure of interest is the relative risk. Specifically, the aims are:

- To compare the performance of methods for estimating relative risks in RCTs with independent and clustered observations;
- To determine the relationship between the ICC on the probability scale and the between cluster variance on the log scale;
- To provide guidance on the analysis of binary outcomes from perinatal trials including infants from multiple births.

Methods: Simulation studies are conducted to compare methods for estimating relative risks using independent and clustered data. To determine the ICC in the latter scenario, the relationship between the ICC on the probability scale and the random effects variance on the log scale is derived. Additional simulation studies are conducted to determine how different analytical methods compare in perinatal trials with multiple births. Example datasets are analysed for illustration.

Results: Some methods for estimating relative risks are associated with large bias and poor coverage. Others fail to overcome the convergence problems of log binomial regression. Several methods perform well across a wide range of independent and clustered data settings, including modified Poisson regression.

When simulating clustered data, the ICC can be determined from the random effects variance on the log scale based on a Taylor series expansion or properties of the lognormal distribution.

Failure to account for clustering in perinatal trials including multiple births leads to inflated type I errors and undercoverage, unless both the ICC and the multiple birth rate are low.

Conclusion: Relative risks are a useful measure of effect for binary outcomes. Difficulties in estimating relative risks due to convergence problems with log binomial regression can be overcome using one of several alternatives, including the popular modified Poisson regression approach. This method works well for both independent and clustered data. Clustering should be taken into account in the analysis of perinatal trials including multiple births.

Declaration

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution to Lisa Yelland and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this thesis (as listed below) resides with the copyright holder(s) of those works.

- Yelland LN, Salter AB, Ryan P. Relative risk estimation in randomized controlled trials: a comparison of methods for independent observations. *The International Journal of Biostatistics* 2011; 7(1):5. DOI: 10.2202/1557-4679.1278. © 2011 Berkeley Electronic Press.
- Yelland LN, Salter AB, Ryan P, Laurence CO. Adjusted intraclass correlation coefficients for binary data: methods and estimates from a cluster randomized trial in primary care. *Clinical Trials* 2011; 8:48-58. DOI: 10.1177/1740774510392256. © The Author(s), 2011.
- Yelland LN, Salter AB, Ryan P. Relative risk estimation in cluster randomized trials: a comparison of generalized estimating equation methods. *The International Journal of Biostatistics* 2011; 7(1):27. DOI: 10.2202/1557-4679.1323. © 2011 Berkeley Electronic Press.
- Yelland LN, Salter AB, Ryan P. Performance of the modified Poisson regression approach for estimating relative risks with clustered prospective data. *American Journal of Epidemiology*: first published online August 12, 2011. DOI: 10.1093/aje/kwr183. © The Author 2011.

- Yelland LN, Salter AB, Ryan P, Makrides M. Analysis of binary outcomes from randomised trials including multiple births: when should clustering be taken into account? *Paediatric and Perinatal Epidemiology* 2011; 25:283-297. DOI: 10.1111/j.1365-3016.2011.01196.x. © 2011 Blackwell Publishing Ltd.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue, the Australasian Digital Theses Program (ADTP) and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed:
Lisa Yelland (Candidate)

Date:

Manuscripts Contributing to this Thesis

- Yelland LN, Salter AB, Ryan P. Relative risk estimation in randomized controlled trials: a comparison of methods for independent observations. *The International Journal of Biostatistics* 2011; 7(1):5. DOI: 10.2202/1557-4679.1278
- Yelland LN, Salter AB, Ryan P, Laurence CO. Adjusted intraclass correlation coefficients for binary data: methods and estimates from a cluster randomized trial in primary care. *Clinical Trials* 2011; 8:48-58. DOI: 10.1177/1740774510392256
- Yelland LN, Salter AB, Ryan P. Relative risk estimation in cluster randomized trials: a comparison of generalized estimating equation methods. *The International Journal of Biostatistics* 2011; 7(1):27. DOI: 10.2202/1557-4679.1323
- Yelland LN, Salter AB, Ryan P. Performance of the modified Poisson regression approach for estimating relative risks with clustered prospective data. *American Journal of Epidemiology*: first published online August 12, 2011. DOI: 10.1093/aje/kwr183
- Yelland LN, Salter AB, Ryan P, Makrides M. Analysis of binary outcomes from randomised trials including multiple births: when should clustering be taken into account? *Paediatric and Perinatal Epidemiology* 2011; 25:283-297. DOI: 10.1111/j.1365-3016.2011.01196.x

Presentations Arising out of this Thesis

- Yelland LN. Analysis of Perinatal Trials Including Multiple Births: When Should Clustering be Taken Into Account? Statistical Society of Australia and the Australasian Epidemiological Association Joint Seminar Series. Adelaide, July 2011.
- Yelland LN. Estimating Relative Risks from Clustered Data. School of Population Health and Clinical Practice Seminar Series. Adelaide, December 2010.
- Yelland LN. Clustering in Perinatal Trials: What is it and when does it matter? School of Population Health and Clinical Practice HDR Research Symposium. Adelaide, 2010.
- Yelland LN, Salter AB, Ryan P. Relative Risk Estimation in Randomised Controlled Trials: A Comparison of Methods for Independent Observations. International Biometrics Society Australasian Region Conference. Taupo, New Zealand, December 2009.
- Yelland LN. Relative Risk Estimation in Randomised Controlled Trials: A Comparison of Methods for Independent Observations. School of Population Health and Clinical Practice Seminar Series. Adelaide, July 2009.
- Yelland LN. Statistical Issues Associated with Binary Outcome Data in Randomised Controlled Trials. School of Population Health and Clinical Practice Seminar Series. Adelaide, August 2008.

Awards Arising out of this Thesis

- Overall best presentation, School of Population Health and Clinical Practice HDR Research Symposium, Adelaide, 2010.
- Runner up, best presentation by a young statistician, International Biometric Society Australasian Region conference in Taupo, New Zealand, 2009.
- CSIRO scholarship to attend the International Biometric Society Australasian Region conference in Taupo, New Zealand, 2009.

Acknowledgements

I would like to sincerely thank the following people for helping to make this thesis possible.

To my supervisors, Phil Ryan and Amy Salter, thank you for all your advice, support and encouragement throughout my candidature. You have both taught me many things in the years we have worked together that I know will stay with me throughout my career. Special thanks must also go to Phil for possibly being the fastest PhD supervisor on the planet at providing feedback!

To Maria Makrides, thank you for your involvement with the multiple births paper, and for giving me the opportunity to be involved with some fantastic projects outside this thesis. My interest in and passion for perinatal trials has been born out of these experiences and I am grateful for your role in that.

To Caroline Laurence, thank you for your involvement with the ICC paper and for all you taught me during our years working on the PoCT Trial, particularly about writing papers. Those lessons have been helpful throughout this thesis and no doubt will continue to be of great use in future.

To Justin Beilby on behalf of the PoCT Trial Management Committee, Maria Makrides on behalf of the DINO Trial Steering Committee and Carmel Collins, thank you for granting me permission to use your datasets as examples in this thesis.

To my fellow PhD candidates, thank you for sharing this exciting and challenging journey with me. Special thanks to Nicole for going first and sharing your wisdom with me, to Jesia and Oana for always being available for a chat, and to my office mates George and Marianne for your support and for creating a quiet and productive work environment (most of the time!).

To my friends and family, thank you for your support and the much needed distractions throughout this journey.

To my parents, thank you for encouraging me to do my best, always believing in me, and giving me all the best opportunities in life. I would not be where I am today without your love and support.

Finally to Tom, thank you for sharing every high and low of this rollercoaster ride I have been on for the last three years. You have talked through the issues with me, listened to my frustrations, reminded me there's more to life than this thesis and celebrated all the wins with me. I am so lucky to have shared this journey with you.

Abbreviations

ACR	Albumin-Creatinine Ratio
AGHQ	Adaptive Gauss-Hermite Quadrature
ANOVA	Analysis of Variance
CS	Conditional Standardisation
DEFF	Design Effect
DHA	Docosahexaenoic Acid
GEE	Generalised Estimating Equation
GLM	Generalised Linear Model
HbA1c	Glycated Haemoglobin
HDL	High Density Lipoprotein
ICC	Intraclass/Intracluster Correlation Coefficient
INR	International Normalised Ratio
IQR	Interquartile Range
MEM	Mixed Effects Model
MQL	Marginal Quasi-Likelihood
MS	Marginal Standardisation
PoCT	Point of Care Testing
PQL	Penalised Quasi-Likelihood
RCT	Randomised Controlled Trial