

Predicting Risk for Pregnancy Complications

Shalem Yiner-Lee Leemaqz

Discipline of Obstetrics and Gynaecology
School of Paediatrics and Reproductive Health
The University of Adelaide
Australia

Thesis submitted for the degree of
Doctor of Philosophy

June 2015



Abstract

For years, it has been a challenge to identify women at risk of Preeclampsia (PE) and Preterm Birth (PTB), one of the leading causes of maternal and perinatal morbidity and mortality. Despite an increasing number of clinical and statistical prediction models being developed, which have been shown to outperform traditional approaches based on maternal history, due to complex underlying relationships and gene-environment interactions, identifying women at risk based on a single time-point, especially during early stages of pregnancy, remains a challenge.

Therefore, this study not only aims to identify potential predictors for pregnancy outcomes and develop prediction models based on combinations of clinical measurements and Single-nucleotide polymorphisms (SNP) predictors, but also to establish a tiered prediction system by integrating risk estimates at various stage of pregnancy.

This thesis contains both theoretical development and practical application of the models, with results of best models written as manuscripts for future publication. Critical issues in real-life statistical analysis, including subgroup differences, and model and variable selection (with FDR control) were discussed, as well as novel strategies on the tiered prediction model development.

The results from tiered models provide prediction for PE and spontaneous preterm birth (SPTB) that not only outperform traditional approaches, but also provide an earlier prediction applicable to all pregnant women, including healthy nulliparous women. This approach also allows for regular monitoring and revision of predicted risk throughout pregnancy. This may assist in providing tailored antenatal care or interventions that could benefit both the mother and child, and to avoid unnecessary interventions for low-risk individuals, while modifiable predictors could also be addressed to reduce the risk or severity of PE or PTB.

Declaration

I, Shalem Y. Leemaqz, hereby declare that this work contains no material that has been accepted for the award of any other degree or diploma in any university or other tertiary institution. To the best of my knowledge and belief, it contains no material previously published or written by any other person, except where due reference is made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

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

- Leemaqz S.Y., Dekker G.A., Roberts C.T. (2013) "Tiered Prediction System for Preeclampsia: an integrative application of multiple models " International Congress on Modelling and Simulation (MODSIM) 2013. pp 2041-2046 ISBN: 978-0-9872143-3-1.
- Dekker G.A., Lee S.Y., North R.A., McCowan L.M., Simpson N.A.B., Roberts C.T., (2012) "Risk factors for preterm birth in an international prospective cohort of nulliparous women" PLoS ONE. 7(7):e39154. doi:10.1371/journal.pone.0039154
- Lee S.Y., Lee S.X., Dekker G.A., Roberts C.T. (2012) "Multivariate Visual Clustering of Single Nucleotide Polymorphisms and Clinical Predictors using Chernoff Faces" Proceedings of the 5th Annual Conference in Applied Statistics Education and Research Collaboration (ASEARC 2012). pp 56-59

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, and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Acknowledgement

I would like to express my deepest gratitude to my principal supervisors Professor Claire Roberts for her constant guidance and unwavering support during my candidature and Professor Gus Dekker for his expertise and breadth of knowledge in clinical research.

I am greatly indebted to and extremely grateful for your dedicated help, insightful expertise guidance, continuous encouragement, understanding, patience, mentorship and friendship throughout all the years.

I am also grateful for the generous financial support for interstate and international trips to attend and present at various workshops and conferences. Your sincere care and selfless support is truly a blessing to all your students and members of the Placenta lab.

I would also like to thank Dr. Stephen Bent for his supervision, support and technical assistance.

My gratefulness also extends to my parents for their understanding and invaluable contribution to both my studies and daily chores throughout this time, and also to my brother and sister, Dr. K.L. Leemaqz and Dr. S.X. Leemaqz, for taking interest to involve in innumerable discussions with me.

Table of Contents

LIST OF FIGURES.....	i
LIST OF TABLES	ii
PUBLICATIONS ARISING FROM THIS THESIS.....	iii
CONFERENCE PRESENTATIONS AND ABSTRACT PUBLICATIONS ARISING FROM THIS THESIS...	iv
CHAPTER 1: INTRODUCTION.....	1
1.1. PROBLEM STATEMENT	1
1.2. AIM.....	2
1.3. OUTLINE.....	3
REFERENCES.....	5
CHAPTER 2: LITERATURE REVIEW	6
2.1. PREECLAMPSIA	6
2.1.1. <i>Complications</i>	7
2.1.2. <i>Screening and Prediction</i>	10
2.1.2.1. Clinical Risk Factors	10
2.1.2.2. Genetic Risk Factors	13
2.1.2.3. Current Prediction Methods.....	16
2.2. SPONTANEOUS PRETERM BIRTH	20
2.2.1. <i>Health Impacts</i>	21
2.2.2. <i>Screening and Prediction</i>	23
2.2.2.1. Clinical Risk Factors	23
2.2.2.2. Genetic Risk Factors	26
2.2.2.3. Current Prediction Methods.....	28
REFERENCES.....	30
CHAPTER 3: SCOPE DATABASE	46
3.1. OVERVIEW	46
3.1.1. <i>Outcomes</i>	47
3.1.2. <i>Data Quality Control</i>	48

3.2.	DATA STRUCTURE	49
3.3.	SUBGROUP DIFFERENCES	51
3.3.1.	<i>Statistical Tests on Significant Subgroup Differences</i>	52
3.3.2.	<i>Feasibility of Subgroup Analysis</i>	55
3.3.3.	<i>Cochran-Mantel-Haenszel Test</i>	56
3.4.	EXPLORATORY DATA ANALYSIS	58
3.4.1.	<i>Univariate Analysis</i>	58
3.4.2.	<i>Subgroup Effects</i>	61
	REFERENCES.....	62

CHAPTER 4: METHODOLOGY..... 65

4.1.	OVERVIEW	65
4.2.	CLASSIFICATION.....	68
4.2.1.	<i>Logistic Regression</i>	68
4.2.1.1.	Binary Logistic Regression	68
4.2.1.2.	Advantages and Disadvantages.....	70
4.2.1.3.	Applications	71
4.2.1.4.	Illustrative Example - Prediction Model for PE	72
4.2.2.	<i>Discriminant Analysis</i>	75
4.2.2.1.	Classification Rule	75
4.2.2.2.	Advantages and Disadvantages.....	77
4.2.2.3.	Applications	77
4.2.2.4.	Illustrative Example - Discriminant Functions for PE.....	78
4.3.	CLUSTERING	81
4.3.1.	<i>K-means Clustering</i>	81
4.3.1.1.	K-means Algorithm	82
4.3.1.2.	Advantages and Disadvantages.....	83
4.3.1.3.	Applications	84
4.3.1.4.	Illustrative Example – Clustering of PTB	85
4.3.2.	<i>Chernoff Faces</i>	88
4.3.2.1.	Advantages and Disadvantages.....	88
4.3.2.2.	Illustrative Example - Visual Clustering of PTB	89
4.4.	MODEL AND VARIABLE SELECTION	92
4.4.1.	<i>Penalty Functions</i>	92
4.4.2.	<i>Regularization Methods</i>	93
4.5.	VALIDATION METHODS	95
4.5.1.	<i>Sensitivity and Specificity</i>	95

4.5.2. Predictive Values.....	97
4.5.3. Receiver Operating Characteristic Curve.....	97
4.6. SUMMARY.....	100
REFERENCES.....	101
CHAPTER 5: TIERED MODELLING APPROACH.....	108
5.1. OVERVIEW.....	108
5.2. MODEL INTEGRATION.....	110
5.2.1. Bayes Theorem.....	110
5.2.2. Risk Classification.....	112
5.3. PROCESS OF ELIMINATION.....	113
REFERENCES.....	117
CHAPTER 6: MANUSCRIPT I – INTERACTION EFFECT OF MARIJUANA.....	118
CHAPTER 7: MANUSCRIPT II – PREECLAMPSIA MODEL.....	151
CHAPTER 8: MANUSCRIPT III – PRETERM BIRTH MODEL.....	176
CHAPTER 9: DISCUSSION AND FUTURE WORK.....	209
9.1. FINAL DISCUSSION.....	209
9.1.1. Predictors for PE and SPTB.....	210
9.1.2. Prediction with Tiered Models.....	216
9.1.3. Model Comparison.....	218
9.2. STRENGTHS AND LIMITATIONS.....	222
9.3. FUTURE WORK.....	223
REFERENCES.....	224
APPENDIX.....	226
I. LIST OF SNPS.....	226
II. SIGNIFICANT SUBGROUP DIFFERENCES BETWEEN ADELAIDE AND AUCKLAND SCOPE PREGNANCIES.....	228
III. EXPLORATORY ANALYSIS OF PREECLAMPSIA.....	252
IV. SIGNIFICANT COCHRAN-MANTEL-HAENSZEL (CMH) TEST RESULTS FOR PE.....	279
V. EXPLORATORY ANALYSIS OF PRETERM BIRTH.....	282

VI. SIGNIFICANT COCHRAN-MANTEL-HAENSZEL (CMH) TEST RESULTS FOR PTB	322
VII. PAPER I: TIERED PREDICTION SYSTEM FOR PREECLAMPSIA	326
VIII. PAPER II: RISK FACTORS FOR PRETERM BIRTH.....	333
IX. PAPER III: MULTIVARIATE VISUAL CLUSTERING.....	342

List of Figures

Fig. 2.1.1: Differentiating hypertensive disorders in pregnant women (Wagner, 2004)	6
Fig. 2.1.2: Susceptibility regions for preeclampsia (Mutze et al., 2008).....	14
Fig. 2.1.3: Scheme of pathophysiological relevant factors in preeclampsia and corresponding candidate genes (Mutze et al., 2008)	15
Fig. 2.2.1: Fetal development timeline (weeks 1 to 37). Images from 3D Pregnancy (Nickelodeon Parents and Preschool Network, 2010).	20
Fig. 2.2.2: Scheme of pathophysiological relevant factors in preterm birth and corresponding candidate genes (Esplin et al., 2005)	26
Fig. 3.1.1: Pregnancy Outcomes in SCOPE	48
Fig. 3.2.1: Clinical measurements database layout	49
Fig. 3.2.2: Single 3D array structure.....	50
Fig. 3.2.3: Multiple 3D array structure	51
Fig. 3.3.1: Outcome differences between (a) Adelaide and (b) Auckland SCOPE women.....	53
Fig. 3.3.2: Common clinical measurement differences between Adelaide (shaded in blue) and Auckland (shaded in red) SCOPE pregnancies, with overall distribution (shown in white).....	54
Fig. 4.1.1: Methodology overview.....	67
Fig. 4.2.1: Partition plot (white=Uncomplicated pregnancy, yellow=PE)	80
Fig. 4.3.1: K-means Clustering	82
Fig. 4.3.2: K-means cluster map (blue=term birth, red=Preterm birth)	86
Fig. 4.3.3: Chernoff face	88
Fig. 4.3.4: Chernoff faces displaying 11 clinical and 4 genetic predictors (yellow=PTB cases).....	91
Fig. 4.4.1: Elastic-Net variable shrinkage pathway	94
Fig. 4.5.1: Distribution of test results and optimal cutoff point.....	98
Fig. 4.5.2: Perfect, conventional and baseline ROC curves.....	99
Fig. 5.1.1: Tiered prediction approach.....	108
Fig. 5.2.1: Model integration overview.....	110
Fig. 5.2.2: Tiered model risk classification.....	112
Fig. 5.3.1: Process of elimination	115
Fig. 9.1.1: ROC curves for a) PE and b) SPTB models	215
Fig. 9.1.2: Venn diagram of patients predicted as high risk for PE or SPTB.....	218
Fig. 9.1.3: Top 40 models developed in this study for preeclampsia (sorted by r+s) compared with current approaches	220
Fig. 9.1.4: Top 40 models developed in this study for preterm birth (sorted by r+s) compared with current approaches	221

List of Tables

Table 2.1.1: Maternal & fetal complications associated with preeclampsia (Sibai et al., 2005).....	8
Table 2.1.2: Risk factors for preeclampsia that can be measured at the first antenatal appointment, reviewed by Sibai et al. (Sibai et al., 2005).....	12
Table 2.1.3: Summary of current prediction methods for preeclampsia (sorted by accuracy).....	17
Table 2.2.1: Fetal morbidities associated with Preterm birth.....	22
Table 2.2.2: Risk factors for Preterm birth (Murphy, 2007).....	24
Table 2.2.3: Summary of potential prediction methods for spontaneous preterm birth.....	28
Table 3.4.1: Univariate analysis of common factors associated with PE	59
Table 3.4.2: Univariate analysis of common factors associated with SPTB	59
Table 3.4.3: SNPs associated with PE in univariate analysis at 5% significance level.....	60
Table 3.4.4: SNPs associated with SPTB in univariate analysis at 5% significance level	61
Table 4.2.1: Logistic regression model for PE at 15 weeks of gestation.....	73
Table 4.2.2: Example clinical and genotype record.....	74
Table 4.2.3: Group means for PE and Uncomplicated pregnancy	79
Table 5.3.1: Pre-test and post-test accuracy measures	114
Table 5.3.2: Final risk classification	116
Table 9.1.1: Final model predictors for preeclampsia and spontaneous preterm birth	211
Table 9.1.2: Predicted risk vs. true cases of preeclampsia and preterm birth	217

Publications arising from this thesis

- **Leemaqz** S.Y., Dekker G.A., Roberts C.T. (2013) "Tiered Prediction System for Preeclampsia: an integrative application of multiple models " *International Congress on Modelling and Simulation (MODSIM) 2013*. pp 2041-2046 ISBN: 978-0-9872143-3-1.
- Dekker G.A., **Lee**[‡] S.Y., North R.A., McCowan L.M., Simpson N.A.B., Roberts C.T., (2012) "Risk factors for preterm birth in an international prospective cohort of nulliparous women" *PLoS ONE*. 7(7):e39154. doi:10.1371/journal.pone.0039154
- **Lee**[‡] S.Y., Lee S.X., Dekker G.A., Roberts C.T. (2012) "Multivariate Visual Clustering of Single Nucleotide Polymorphisms and Clinical Predictors using Chernoff Faces" *Proceedings of the 5th Annual Conference in Applied Statistics Education and Research Collaboration (ASEARC 2012)*. pp 56-59

[‡] Last name changed to Leemaqz from Lee in 2013.

Conference presentations and abstract publications arising from this thesis

- **Leemaqz** S.Y., Dekker G.A., McCowan L.M.E., Roberts C.T. (2014) "Prediction Model for Spontaneous Preterm at 15 weeks of Gestation" *American Journal of Obstetrics & Gynaecology*. 120(1) pp S380-381
- **Leemaqz** S.Y., Dekker G.A., McCowan L.M.E., Roberts C.T. (2014) "Prediction for Spontaneous Preterm Birth: A Three-tiered Approach" *Perinatal Society of Australia and New Zealand (PSANZ) 18th Annual Congress*. 6-9 April 2014.
- **Leemaqz** S.Y., Dekker G.A., McCowan L.M.E., Roberts C.T. (2013) "A model at 15 weeks gestation to discriminate between uncomplicated pregnancies and those destined to develop preeclampsia" *Placenta*. 34(9) pp A54
- **Leemaqz** S.Y., Dekker G.A., Roberts C.T. (2013) "Tiered Prediction System for Preeclampsia: an integrative application of multiple models " *20th International Congress on Modelling and Simulation (MODSIM)*. 1-6 December 2013.
- **Leemaqz** S.Y., Dekker G.A., Roberts C.T. (2013) "Preeclampsia prediction model at 15 weeks of gestation using clinical and SNP predictors" *Perinatal Society of Australia and New Zealand (PSANZ) 17th Annual Congress*. 14-17 April 2013.
- **Lee** S.Y., Lee S.X., Dekker G.A., Roberts C.T. (2012) "Multivariate Visual Clustering of Single Nucleotide Polymorphisms and Clinical Predictors using Chernoff Faces" *5th Annual Conference in Applied Statistics Education and Research Collaboration (ASEARC 2012)*. 2-3 February 2012.