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

List of Tables

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 2041-2046ISBN: 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 " *20thInternational 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.

Chapter 1: Introduction

1.1. Problem Statement

Preeclampsia and preterm birth are major complications of late pregnancy, and are the leading causes of maternal and perinatal death (Kramer, 2003; King et al., 2004; Cnossen et al., 2006; Said, 2006). In Australia, there is an estimated maternal mortality ratio of 8.4 deaths per 100,000 confinements during 2003-2005 (Sullivan et al., 2008), and a perinatal death rate of 9.3 deaths per 1000 births in 2010 (Li et al., 2010).

Currently, there is a need for screening tests that accurately predict pregnancy outcomes, especially during early stages of pregnancy prior to symptoms, which can be used to monitor and assess the risk for complications on regular antenatal visits. Since pregnancy complications often present suddenly, the standard intervals between antenatal visits may result in delays in diagnosis with an increased chance of severe complications (SCOPE Consortium, 2004). Screening tests and accurate prediction of pregnancy outcomes prior to their clinical onset is vital, as high risk women could benefit from intensive monitoring and preventative treatment (Dekker et al., 2001; Mostello et al., 2003; Said, 2006).

Due to the vast number of pregnancy complications, this study will focus on preeclampsia and preterm birth, but the methods applied in this study may also be applicable for other pregnancy outcomes.

1.2. Aim

Since the factors related or associated with preeclampsia and preterm birth are vast and complex, the primary aim of this study is to develop potential prediction tests for pregnancy outcomes using statistical or data mining methods, through modelling, classification and clustering techniques, which examines the data structure and relationship between variables..

As there is increasing evidence of genetic predisposition, i.e. family history is often a risk factor, along with clinical and environmental factors, this study will also aim at developing models incorporating clinical, environmental and single-nucleotide polymorphism (SNP) factors.

Each model will be verified and the accuracy of the prediction models will be evaluated using sensitivity, specificity and Receiver Operating Characteristic (ROC) curves. The best model for preeclampsia and preterm birth will be selected based on these accuracy measures and model penalty functions including Akaike Information Criterion (AIC) and Elastic-Net penalty. This will be further discussed in Chapter 5.

The ultimate aim is to develop an integrative prediction method for pregnancy complications that can be practically used to identify women at risk from as early as pre-conception to 20 weeks of gestation, which may assist in providing tailored care for individuals, and initiate prevention strategies as early as possible.

1.3. Outline

This thesis consists of both theoretical development and practical application of prediction models, with a literature review of current approaches published, along with mathematical approaches used, followed by practical applications.

An overview of the two pregnancy complications studied, preeclampsia and spontaneous preterm birth, is provided in Chapter 2. This contains a review on the epidemiology, health impacts, as well as associated clinical and genetic factors, and current prediction approaches.

Chapter 3 discusses the features and structure of the Screening fOr Pregnancy Endpoints (SCOPE) database used in this study. This includes data summary, along with discussion on real-life statistical analysis issues, followed by a summary of results from exploratory analysis.

Details of main methodology approaches are discussed in Chapter 4, which includes conventional data mining approaches such as Classification and Clustering. The mathematical theory behind each approach applied in this study is discussed, accompanied by an illustrative scenario in context to this study. Critical issues including methods of validation and model selection are also discussed in this chapter.

An additional methodology chapter (Chapter 5) discusses both the theoretical and application of a novel approach, i.e. the tiered modelling approach. This includes the mathematical development of multi-model integration, as well as the concept of Process of Elimination in the application of the tiered approach.

Chapters 6 to 8 are analysis results and final models, presented in manuscript format.

The first paper (Chapter 6) discuss on the lifestyle factors for pregnancy complications, which analyze the interactions between BMI, smoking, and marijuana usage. Final models for preeclampsia and preterm birth are presented in Chapters 7 and 8, using a tiered prediction approach.

The final Chapter 9 provides a summary of results from this study, followed by discussions on future improvements.

References

Cnossen, J. S., van der Post, J. A., Mol, B. W., Khan, K. S., Meads, C. A. and ter Riet, G. (2006). "Prediction of pre-eclampsia: a protocol for systematic reviews of test accuracy." *BMC Pregnancy Childbirth***6**: 29.

Dekker, G. and Sibai, B. (2001). "Primary, secondary, and tertiary prevention of pre-eclampsia." *The Lancet* **357**(9251): 209-215.

King, J. F., Slaytor, E. K. and Sullivan, E. A. (2004). "Maternal deaths in Australia, 1997-1999." *Med J Aust* **181**(8): 413-414.

Kramer, M. S. (2003). "The epidemiology of adverse pregnancy outcomes: an overview." *J Nutr* **133**(5 Suppl 2): 1592S-1596S.

Li, Z., Zeki, R., Hilder, L. and Sullivan, E. A. (2010). Australia's mothers and babies 2010. *Perinatal Statistics Series*. Canberra, Australian Institute of Health and Welfare National Perinatal Statistics Unit.

Mostello, D., Droll, D. A., Bierig, S. M., Cruz-Flores, S. and Leet, T. (2003). "Tertiary care improves the chance for vaginal delivery in women with preeclampsia." *Am J Obstet Gynecol* **189**(3): 824-829.

Said, J. (2006). "Managing pre-eclampsia." *O&G Magazine* **8**(4): 43-44.

SCOPE Consortium (2004). "The SCOPE Pregnancy Research Study." Retrieved 5 March 2010, from http://www.scopestudy.net/.

Sullivan, E. A., Hall, B. and King, J. (2008). Maternal deaths in Australia, 2003-2005. *Maternal Deaths Series*. Canberra, AIHW National Perinatal Statistics Unit.

Chapter 2: Literature Review

This chapter will discuss the definitions of Preeclampsia and Preterm birth, along with possible clinical and genetic risk factors that have been identified, and also, current prediction methods.

2.1. Preeclampsia

Preeclampsia is one of hypertensive disorders in pregnancy, along with eclampsia and gestational hypertension (Fig. 2.1.1). Hypertension occurs when women have a blood pressure of 140/90 mmHg or greater after 20 weeks of pregnancy. This study followed the research definition (Brown et al., 2001), where Preeclampsia is defined as hypertension accompanied by proteinuria of 300 mg or greater on 24-hour urine collection, or a spot Protein to Creatinine ratio of 30 mg/mmol creatinine or greater, with any organ manifestation. This is consistent with the new definition recently published by the International Society for the Study of Hypertension in Pregnancy (ISSHP).

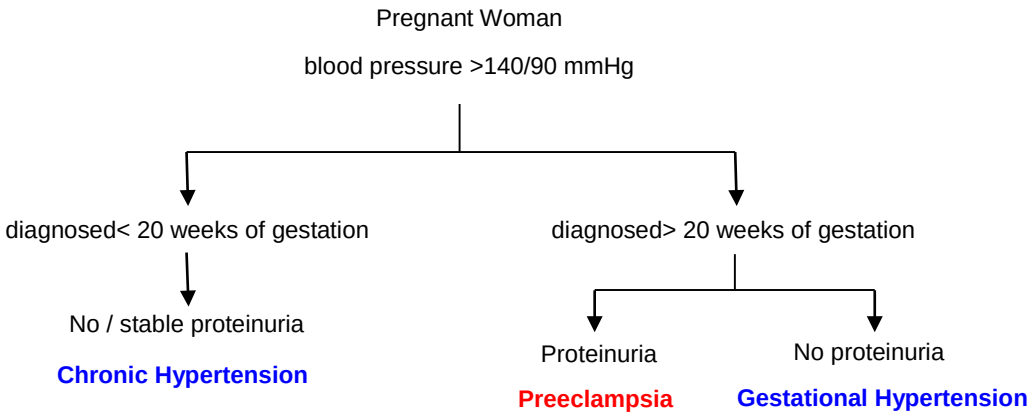

Fig. 2.1.1: Differentiating hypertensive disorders in pregnant women (Wagner, 2004)

It has been estimated that preeclampsia affects approximately 3-5% of pregnancies worldwide (Verlohren et al., 2010; Wang et al., 2010). Amongst women who gave birth in South Australia during 2010, 7% had pregnancy-induced hypertension, which is the highest in Australia. Furthermore, hypertension or preeclampsia was the main reason for 12.8% of birth by induction, and 2.5% of caesarean sections, which is the second highest compared to New South Wales, Victoria, Queensland and Tasmania (Li et al., 2010).

2.1.1. Complications

Severe preeclampsia is associated with placental abruption (Stekkinger et al., 2009), disseminated intravascular coagulation (DIC) (Ducloy-Bouthors, 2010), renal failure (Brown et al., 1992; Ansari et al., 2008; Goplani et al., 2008), hepatic failure (Rahman et al., 2002; Hay, 2008), central nervous system haemorrhage (Ahmed, 2002; Moodley, 2008) and stroke in the mother (Aali et al., 2004; Hacker et al., 2004; Irminger-Finger et al., 2008). It is a major cause of maternal and perinatal morbidity and mortality, and the third leading cause of direct maternal death (King et al., 2004; Cnossen et al., 2006; Said, 2006). Every year, over 50,000 mothers around the world die from eclampsia following preeclampsia (Brennecke, 2010).

Table 2.1.1: Maternal & fetal complications associated with preeclampsia (Sibai et al., 2005)

	Complications	Occurrence
Maternal Complications	Abruptio placentae	1-4%
	Disseminated coagulopathy / HELLP syndrome	10-20%
	Pulmonary oedema / aspiration	2-5%
	Acute renal failure	1-5%
	Eclampsia	< 1%
	Liver failure or haemorrhage	< 1%
	Stroke	rare
	Death	rare
	Long-term cardiovascular morbidity	
Neonatal Complications	Preterm delivery	15-67%
	Fetal growth restriction	10-25%
	Hypoxia-neurologic injury	< 1%
	Perinatal death	1-2%
	Long-term cardiovascular morbidity associated with low birth weight (developmental origins of adult disease)	

Patients with preeclampsia may experience headache and visual disturbance due to hypertension (Hacker et al., 2004). These symptoms, along with epigastric pain, may indicate progression towards eclampsia.

Renal involvement in preeclampsia relates to glomeruloendotheliosis, which is swelling of the glomerular capillary endothelium that decreases glomerular perfusion and glomerular filtration rate; a characteristic lesion of preeclampsia (DeCherney et al., 2002). In a minority of patients, preeclampsia may lead to acute renal failure on the basis of tubular necrosis or cortical necrosis.

Severe preeclampsia can be complicated by disseminated intravascular coagulation (DIC) (Ducloy-Bouthors, 2010). In severe preeclampsia cases (in particular in developing countries), pulmonary oedema may occur (DeCherney et al., 2002; Zhang

et al., 2003; Hanretty, 2009). Moreover, preeclampsia can cause fetal and neonatal complications, such as growth restriction, prematurity, and perinatal death (Hacker et al., 2004). It is estimated that 15% of preterm deliveries are due to preeclampsia (Walsh, 2007).

Table 2.1.1 shows some maternal multisystem disorders and fetal complications associated with preeclampsia reviewed by Sibai et al. (Sibai et al., 2005).

The most common cause of death among women with preeclampsia/eclampsia is intracranial haemorrhage. Other causes include renal or hepatic failure, pulmonary oedema and preeclampsia with hepatic and haematological abnormalities (HELLP syndrome). Of the 5 direct maternal deaths from severe hypertensive disease during 2003-2005 in Australia, more than half of the cases had intracranial hemorrhage related to preeclampsia or pregnancy-induced hypertension (Sullivan et al., 2008). Every year, there are approximately 200 perinatal deaths that result from preeclampsia in Australia, and many of these are a consequence of induced premature delivery rather than the disease itself, as the only cure for preeclampsia is delivery (Brennecke, 2010).

One point to note is that the number of mothers and babies being severely affected is expected to be higher than the estimate, as there are many more cases of women with preeclampsia related complications who had permanent morbidity, e.g. in intensive care or long-term health problems, who are not being accounted for in the death statistics published in 2008 (Sullivan et al., 2008). A study by Tuffnell et al. in Yorkshire analyzed 1087 women who delivered between 1991 and 2003 and were diagnosed with preeclampsia/eclampsia. There were no maternal deaths. However, 151 (around 14%) had serious complications and 32% of those cases required ICU admission (Tuffnell et al., 2005).

2.1.2. Screening and Prediction

Due to the serious health impacts, screening tests and accurate prediction of preeclampsia prior to its clinical onset is vital, as high risk women could benefit from intensive monitoring and preventative treatment (Dekker et al., 2001; Mostello et al., 2003; Said, 2006). Hence, many studies have been undertaken to investigate possible clinical and genetic risk factors associated with preeclampsia, and also protein markers in maternal blood at different times in pregnancy. Moreover, many statistical models are being developed based on clinical and genotype data (Yu et al., 2008; North et al., 2011; Wright et al., 2012; Akolekar et al., 2013). This will be further discussed in Section 2.1.2.3.

2.1.2.1. Clinical Risk Factors

Currently, the aetiology of preeclampsia is unknown, which creates complexity when investigating methods of prediction. However, many theories have been proposed, and it is accepted that the starting point of preeclampsia, in particular when associated with IUGR, is in the placental bed (Hanretty, 2009). Based on clinical experience and statistics, a number of risk factors have been investigated. Some common risk factors include obesity, age, obstetric history and family history (Farag et al., 2004; Carty et al., 2008; Briceno-Perez et al., 2009; Steegers et al., 2010). The risk of preeclampsia for women who have a BMI of 26 is estimated to be double that of those who have a BMI of 21 which triples at a BMI of 30, and increases further with severe obesity (Ros et al., 1998; Bodnar et al., 2005; Stone et al., 2007). Extremes of age, e.g. less than 20 years old or older than 35, also appears to increase risk (King et al., 2004). A study by Tubbergen et al. (Tubbergen et al., 1999) showed that the incidence of preeclampsia in

first pregnancy is higher than subsequent pregnancies and that change of partner raises the risk of preeclampsia in subsequent pregnancies. Furthermore, family history of hypertension has an estimated odds ratio of 1.7 for preeclampsia based on a case-control study in America (Eskenazi et al., 1991). In addition, maternal low birth weight and preterm birth have also been found to increase risk for preeclampsia (Innes et al., 1999).

Smoking has an interesting association with preeclampsia. Although smoking during pregnancy has been found to be associated with a variety of adverse pregnancy outcomes, such as an increased risk of intrauterine growth restriction and perinatal death, the incidence of preeclampsia is lower amongst women who smoke (Ananth et al., 1996; Zhang et al., 1999; Tsai et al., 2008). Moreover, a study by Conde-Agudelo et al. (Conde-Agudelo et al., 1999) who performed a meta-analysis on 35 studies on the effect of smoking on preeclampsia confirmed that the risk of preeclampsia among pregnant women who smoked was 32% lower than that among non-smoking pregnant women (Spinillo et al., 1994; Cnattingius et al., 1997; Xiong et al., 2000; Stone et al., 2007; Pipkin, 2008).

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)

	Risk Factors
Couple-related	Limited sperm exposure
	Primipaternity
	Pregnancies after donor insemination, oocyte donation embryo donation
	Protective effect of partner change in the case of previous preeclamptic pregnancy
Maternal or pregnancy-related	Extremes of maternal age
	Multifetal gestation
	Preeclampsia in a previous pregnancy
	Chronic hypertension or renal disease
	Rheumatic disease
	Maternal low birthweight
	Obesity and insulin resistance
	Pre-gestational diabetes mellitus
	Maternal infections
	Pre-existing thrombophilia
	Maternal susceptibility genes
	Family history of preeclampsia
	Smoking (reduced risk)
	Hydropic degeneration of placenta

Season may also be a potential risk factor. A study by Bodnar et al. (Bodnar et al., 2007) investigated the monthly variation of preeclampsia incidence, and reported that there is a lower incidence of preeclampsia when the baby is due during the summer months amongst Caucasian women in the US.

Studies by Klonoff-Cohen et al. (Klonoff-Cohen et al., 1996) and more recently by SCOPE consortium (McCarthy et al., 2013) have shown that alcohol consumption does not appear to have a significant association with preeclampsia. Alcohol has been known to be associated with non-pregnant hypertension. In fact, 5% of non-pregnant hypertension is due to alcohol consumption and 30-60% of alcoholics have

hypertension (James, 2010).

2.1.2.2. Genetic Risk Factors

Interestingly, there is growing evidence of familial tendency in preeclampsia. A classic study by Chesley and Cooper (Chesley et al., 1986) has found that preeclampsia occurs in 26% of the daughters and 16% of the granddaughters of women who had preeclampsia. Subsequent studies have also estimated that the incidence of preeclampsia is nearly tripled amongst women with a family history of preeclampsia (Cincotta et al., 1998; Esplin et al., 2001). In addition, studies have found that there is a difference in the prevalence of preeclampsia in different ethnicities. This might be because allele frequencies at many polymorphisms differ between ethnic groups (Chappell et al., 2006), for instance, it is known that the Angiotensinogen AGT rs699 and AGT rs5409 polymorphisms, which have been linked to hypertension and preeclampsia, are more common in African American women, and these women have higher rates of preeclampsia than white American women (Medica et al., 2007; Jenkins et al., 2008; Zafarmand et al., 2008). These may suggest a genetic predisposition (Esplin et al., 2001; Cnossen et al., 2006).

Moreover, a study by Cnattingius et al. (Cnattingius et al., 2004) has estimated that genetic factors contribute more than 50% of the total phenotypic variance in the incidence of preeclampsia. Approximately 35% of the genetic contribution originated from the mother and 20% are fetal effects combining genetic factors originating from both parents (Peterson, 2010). Other explanations for familial pattern could be the higher incidence of preeclampsia in women with a low birthweight (Innes et al., 1999).

Hence, many studies, such as Genetics Of Pre-EClampsia (GOPEC) (GOPEC Consortium, 2005), have searched for genetic factors, attempting to identify

chromosomal regions or candidate genes whose variants may be related to high preeclampsia susceptibility. Three loci have been identified to have significant linkage to preeclampsia based on Genome-wide linkage analysis. These are identified on chromosomal 2p13, 2p25 and 9p13 (Arngrimsson et al., 1999; Moses et al., 2000; Laivuori et al., 2003; Farag et al., 2004; Peterson, 2010).

Fig. 2.1.2 shows the susceptibility regions identified for preeclampsia based on genome-wide studies (Mutze et al., 2008).

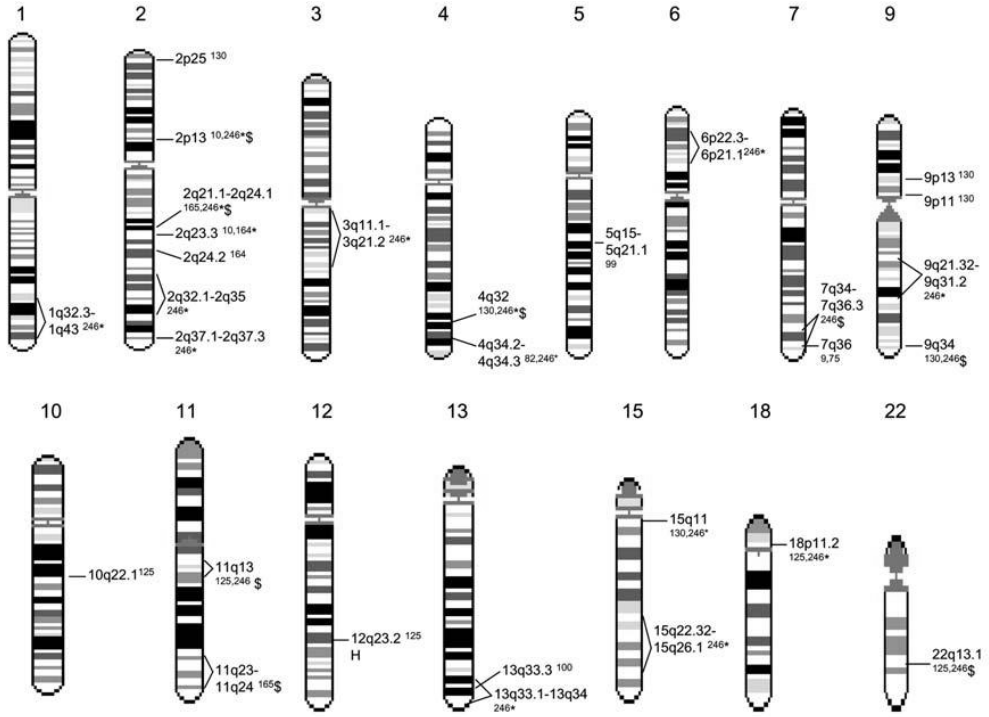

Fig. 2.1.2: Susceptibility regions for preeclampsia (Mutze et al., 2008)

Candidate gene studies have also been undertaken. The majority of these studies are carried out by comparing frequencies of genetic variants in cases and controls. Many investigated a single polymorphism in a single candidate gene, and some tested several genes, or multiple polymorphisms in one or more genes (Chappell et al., 2006).

Amongst the 50 or more candidate genes studied, only 8 genes account for about 70% of research published on the topic (Mutze et al., 2008). These include Factor V Leiden

(F5) (Dizon-Townson et al., 1996; De Groot et al., 1999; Currie et al., 2002; D'Elia et al., 2002; Dalmaz et al., 2006), Methylenetetrahydrofolatereductase (MTHFR) (Powers et al., 1999; Kaiser et al., 2000; Driul et al., 2004; Vucic et al., 2009), Prothrombin (F2) (De Maat et al., 2004; Romero et al., 2010), Angiotensin converting enzyme (ACE) (Zhou et al., 1999; Steegers et al., 2010), Angiotensin type 1 and type 2 receptors (AGTR1, AGTR2) (Bouba et al., 2003; Kobashi et al., 2004), Angiotensinogen (AGT) (Arngrimsson et al., 1993; Ward et al., 1993), Endothelial nitric oxide synthetase 3 (eNOS3) (Brennecke et al., 1997; Yoshimura et al., 2000; Bashford et al., 2001), and Tumor necrosis factor α (TNFα) (Farag et al., 2004; Chappell et al., 2006; Peterson, 2010).

Wilson et al. (Wilson et al., 2003) suggested that candidate genes can be subdivided into six categories based on their hypothesized role in preeclampsia etiology. Fig. 2.1.3 shows a scheme of pathophysiological relevant factors in preeclampsia and corresponding candidate genes reviewed by Mutze et al. (Mutze et al., 2008).

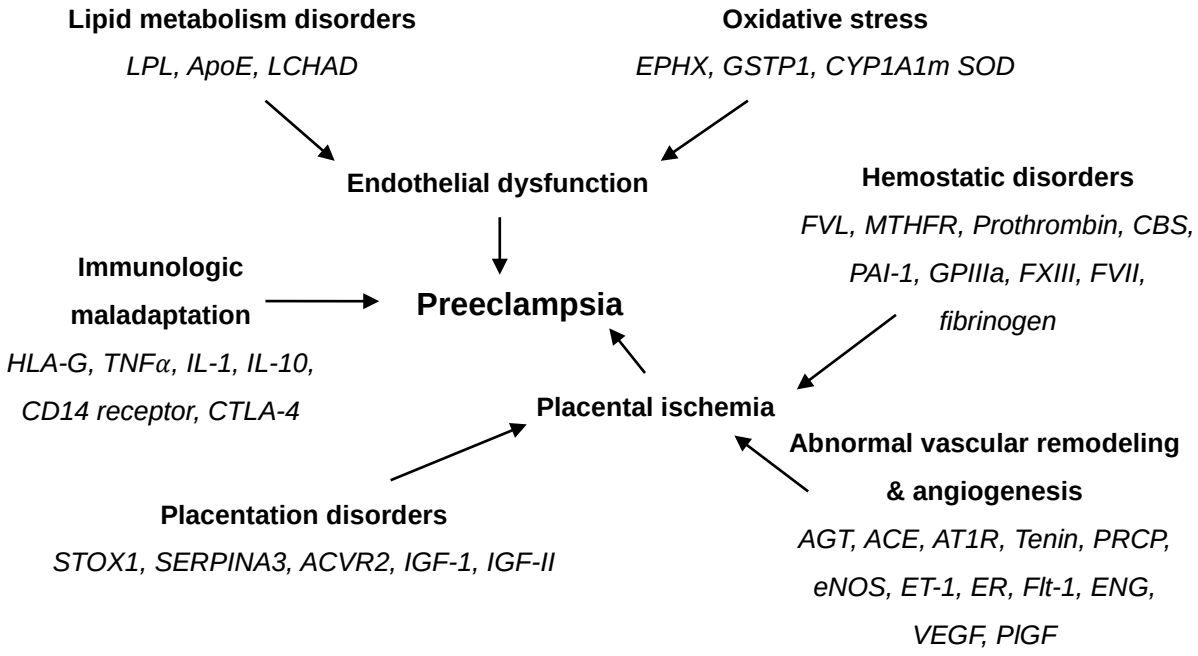

Fig. 2.1.3: Scheme of pathophysiological relevant factors in preeclampsia and corresponding candidate genes (Mutze et al., 2008)

2.1.2.3. Current Prediction Methods

Currently, there are no simple screening or prediction tests available for preeclampsia and the detection of preeclampsia continues to depend on increasingly frequent antenatal visits in late pregnancy for blood pressure measurement and urinalysis (Conde-Agudelo et al., 2004; Farag et al., 2004). Nevertheless, uterine artery Doppler has also been widely used, but it has been shown to have limited value as a single screening test (Farag et al., 2004; Papageorghiou et al., 2006; Herraiz et al., 2009). Maternal history is also widely used, yet, it is estimated that only 30% to 40% of the preeclampsia cases are successfully predicted (Papageorghiou et al., 2005).

More and more statistical prediction models have been developed, and some of them have been shown to obtain a more accurate prediction result than clinical screenings alone. Interestingly, the majority of statistical models result from logistic regression. Table 2.1.3 summarizes some recent publications on potential prediction models of preeclampsia.

However, one point to note is that most of the results are based on single studies from independent data, with variable definitions of preeclampsia. Hence, the validity of the available tests is difficult to evaluate as a result of the absence of a ‘gold standard’ to confirm the diagnosis (Briceno-Perez et al., 2009).

Although a number of prediction models published have been shown to have remarkable predictive results, yet, many of the prediction models are for early-onset preeclampsia and are not performed in early stages of pregnancy, and early prediction for term disease still remains a challenge. Prevention strategies in women identified at risk need to be initiated as early as possible in pregnancy. The new analysis on low-

dose Aspirin (Bujold et al., 2014) indicate that low-dose Aspirin needs to start prior to 16 weeks' gestation.

In addition, it may not be applicable or accurate enough to predict preeclampsia based on a single clinical or statistical method. Nevertheless, accurate prediction methods for specific groups of women, e.g. from the same ethnicity background, will also be valuable.

Table 2.1.3: Summary of current prediction methods for preeclampsia (sorted by accuracy)

Method	Details	Accuracy	Reference
Statistical Logistic regression	**2nd trimester** Placental growth factor (PlGF) Vascular endothelial growth factor (VEGF) Anti-angiogenic factors	**Sensitivity:** 100% **Specificity:** 99%	(Kusanovic et al., 2009)
Clinical Urinary angiogenic factors	**24-40 weeks** Cutoff > 2.1 in ratio log[sFlt-1/PlGF]	**Sensitivity:** 88.2% **Specificity:** 100%	(Buhimschi et al., 2005)
Statistical Logistic regression	**11-13 weeks** Maternal factors Uterine artery PI MAP PAPP-A PlGF	**Sensitivity:** 93.1% **Specificity:** 95%	(Poon et al., 2009)
Clinical sFlt -1, PlGF	**1st & 2nd trimester** sFlt -1/PlGF ratio	**Area Under Curve:** 0.97	(Verlohren et al., 2010)
Statistical Logistic regression Two-dimensional analysis	**1st & 2nd trimester** Urinary creatinine Systolic BP Urinary inorganic phosphorus	**Sensitivity:** 75% **Specificity**: 95%	(Kuromoto et al., 2010)
Statistical Logistic regression	**1st & 2nd trimester** Maternal factors PlGF and sFlt	**Sensitivity:** 52% **AUC:** 0.84	(Myers et al., 2013)
Statistical Logistic regression	**15-20 weeks** Demographic characteristics	**Sensitivity:** 75.5% **Specificity:** 86.9%	(von Dadelszen et al., 2010)

	Obstetric history Fetal assessments		
Statistical Logistic regression	**1st trimester** Maternal history Uterine artery pulsatility index	**Sensitivity:** 100% **Specificity:** 60.7%	(Herraiz et al., 2009)
Statistical Logistic regression	**15-20 weeks** Soluble vascular endothelial growth factor receptor-1 (Flt-1) Endoglin (ENG)	**Sensitivity:** 66% **Specificity:** 90%	(Sekizawa et al., 2010)
Statistical Univariate regression Logistic regression	**1st trimester** Serum markers Uterine artery resistance index	**Sensitivity:** 69.2%	(Thilaganathan et al., 2010)
Statistical Logistic regression	**11-13 weeks** Glycerol, carnitine 1-methylhistidine	**Sensitivity:** 56.7% **Specificity:** 95% **AUC:** 0.783	(Bahado-Singh et al., 2013)
Clinical Cardiovascular risk factor	Frequency and timing of blood pressure measurement Cholesterol and glucose measurements Vascular diagnosis	**Sensitivity:** 51.4% **Specificity:** 100%	(Nijdam et al., 2009)
Clinical	**34-40 weeks** Urinary soluble endoglin	**Sensitivity:** 70% **Specificity:** 80%	(Buhimschi et al., 2010)
Statistical Functional network analysis	**16-19 weeks** Soluble endoglin (sEndoglin) Soluble frms-like tyrosine kinase receptor-1 (sFLT-1) Leptin Adiponectin Endothelin 1	**Area Under Curve:** 0.753	(Wang et al., 2010)
Clinical Plasma fibronectin Advanced oxidative protein products (AOPP)	**19-25 weeks** Total fibronectin≥ 360 mg/l	**Sensitivity:** 57% **Specificity:** 92%	(Dane et al., 2009)
Clinical	**2nd trimester** Uterine artery Doppler	**Sensitivity:** 60% **Specificity:** 75%	(Farag et al., 2004; Jacquemyn et al., 2010)

Clinical Renal Interlobar Vein Impedance Index (RIV II)	**2nd trimester** RIV II higher in preeclampsia than in uncomplicated pregnancy < 34 weeks gestation	**Sensitivity:** 60% **Specificity:** 64.3%	(Gyselaers et al., 2009)
Statistical Logistic regression	**1st trimester** Demographic characteristics Maternal history	**Area Under Curve:** 0.67	(Direkvand-Moghadam et al., 2012)
Clinical	**Pre-pregnancy** Maternal history	**Sensitivity:** 30-40% **Specificity:** 60-70%	(Papageorghiou et al., 2005)

** measurements are in maternal blood unless otherwise stated*

2.2. Spontaneous Preterm Birth

Preterm birth is defined as onset of labour before 37 completed weeks, or less than 259 days, of gestation. It occurs in approximately 5-10% of births (Pfeifer, 2007; Hanretty, 2009). In Australia, 8.3% of mothers had preterm labour during 2010, a 0.9% increase since 2007 (Laws et al., 2009; Li et al., 2010). Preterm births may be indicated or spontaneous. Indicated preterm birth is generally a result of medical or obstetric complications, such as hypertension, diabetes, or intrauterine growth restriction (Villar et al., 2004; Murphy, 2007). Spontaneous preterm births accounts for 60-70% of preterm births, these cases occur naturally, and are most likely due to covert or subclinical infective processes, cervical dysfunction, poor placentation, multiple gestation, and possibly, nutritional and environmental factors (Lumley, 2003; Honest et al., 2009). This study will focus on prediction for spontaneous preterm birth.

Preterm birth can be categorized based on gestational age. Onset of labour prior to 24 weeks of gestation is considered as pre-viable, in which the survival chance of the neonate is very low. The majority (60-70%) of preterm labour occurs between 34 to 37 weeks of gestation. They are referred to as Late preterm. Onset of labour between 25 and 28 weeks of gestation is considered as Extreme preterm. It accounts for approximately 5% of preterm births (Goldenberg et al., 2008). Fig. 2.2.1 shows the stages of preterm birth.

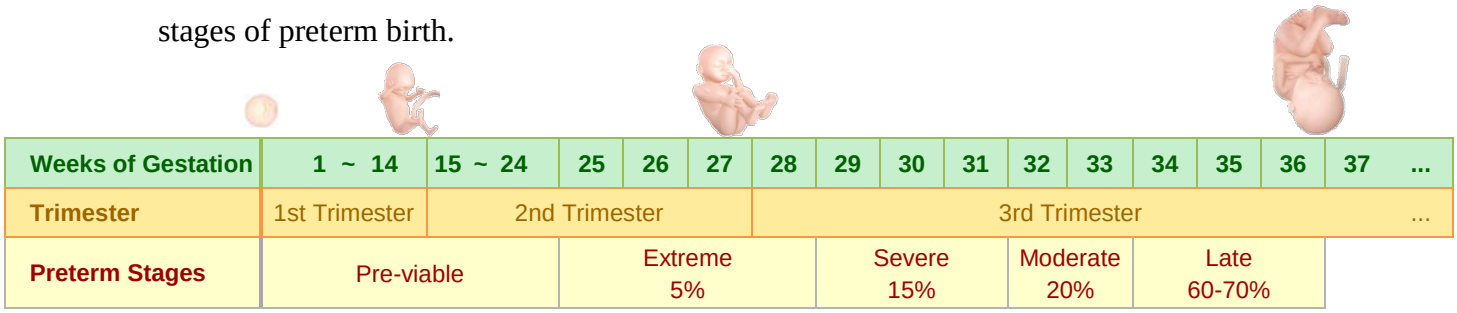

Fig. 2.2.1: Fetal development timeline (weeks 1 to 37). *Images from 3D Pregnancy (Nickelodeon Parents and Preschool Network, 2010).*

2.2.1. Health Impacts

Preterm infants, especially those born before 34 weeks of gestation, have a high risk of short-term or long-term morbidity, and even death. It is estimated that 75% of neonatal mortality is due to preterm birth, and 50% of children who have long-term neurological impairment were born preterm (Den Ouden et al., 1996; Kramer et al., 2000;Mikkola et al., 2005). Moreover, preterm birth resulted in approximately 500,000 deaths per year worldwide (Child Health Research Project, 1999). In South Australia, 15% of perinatal deaths were due to spontaneous preterm birth in 2008, which is the second leading cause of perinatal death, following congenital abnormalities (Maternal Perinatal and Infant Mortality Committee, 2009).

Preterm infants are likely to suffer serious morbidities such as respiratory distress syndrome (Platzker, 1972; Lepercq et al., 2004; Hibbard et al.), bronchopulmonary dysplasia (Woynarowska et al., 2008; Doyle et al.; Farstad et al.), intraventricular haemorrhage (2010; Lee et al., 2010), and retinopathy of prematurity (Shah et al., 2005; Shinsato et al., 2010; VanStone, 2010). Recently, Navaei et al. (Navaei et al., 2010) have studied 194 newborns with a gestational age of 30 weeks or less. Approximately 76% suffered respiratory distress syndrome, 30.9% had septicemia, 10.3% had bronchopulmonary dysplasia, and 7.2% had intraventricular haemorrhage. Unfortunately, only 35.6% of preterm infants in this study survived. Table 2.2.1 summarizes the morbidities associated with preterm birth.

Table 2.2.1: Fetal morbidities associated with Preterm birth

Systems at risk	Medical Conditions
Central Nervous System	Hypoxic-ischemic encephalopathy (HIE)
	Developmental disability
	Cerebral palsy
	Intraventricular haemorrhage (IVH)
Cardiovascular System	Patent ductus arteriosus (PDA)
	Hypotension
	Bradycardia
Pulmonary System	Respiratory distress syndrome (RDS / IRDS)
	Bronchopulmonary dysplasia (BPD)
Gastrointestinal System	Hypoglycemia
	Gastroesophageal reflux (GER)
	Inguinal hernia
	Necrotizing enterocolitis (NEC)
Hematologic System	Anemia of prematurity
	Thrombocytopenia
	Hyperbilirubinemia
Auditory System	Hearing disorders (congenital or perinatal)
	Impairment of speech and language development
Ophthalmic System	Retinopathy of prematurity (ROP)
	Myopia
	Strabismus
	Amblyopia
	Optic nerve atrophy
	Cataracts
	Cortical visual impairment
Other Complications	Sepsis
	Pneumonia
	Urinary tract infection

Infants born prior to 32 weeks of gestation, i.e. severe or extreme preterm, have the greatest risk of poor health outcomes (Murphy, 2007). It has been shown that neonatal morbidity and mortality is inversely proportional to gestational age (Pfeifer, 2007; Shinsato et al., 2010). That is, the more preterm, the risk of neonatal morbidity and

mortality increases.

Furthermore, even preterm infants who did not have serious complications after birth have a higher risk of developing long-term health and developmental problems (2007; Honest et al., 2009). A recent study by Luu et al. (Luu et al., 2010) compared the healthcare use of 254 preterm infants during 18 months from neonatal discharge. A re-hospitalisation rate of 49% occurred in extreme preterm infants. More than half (59%) required physical or occupational therapy, and 17% were enrolled in a long-term rehabilitation program. This is not only a life-long burden to the child, but also an emotional, psychological and financial burden for their families. Moreover, a recent Offspring study by Abraham (Abraham, 2009) concluded that preterm birth has impacts on at least two generations.

2.2.2. Screening and Prediction

Considering the significant long-term and short-term effects, ways to identify or predict high-risk individuals are valuable, as it may assist clinicians to provide appropriate care or antenatal interventions that could benefit both the mother and child, and also, to avoid unnecessary, costly, and possibly hazardous interventions for low-risk individuals (Committee on Understanding Premature Birth and Assuring Healthy Outcomes, 2007; Honest et al., 2009).

2.2.2.1. Clinical Risk Factors

Since preterm birth has multiple aetiologies, such as infections or obstetric complications, determining the associated risk factors may be an efficient approach, as they encompass the possible causes. A variety of medical, nutritional, environmental and socioeconomic risk factors have been found. These include age, history of preterm birth, low socioeconomic status, and smoking (McCowan et al., 2009; Bhattacharya et

al., 2010). Table 2.2.2 summarizes some common risk factors for preterm birth reviewed by Murphy (Murphy, 2007).

Table 2.2.2: Risk factors for Preterm birth (Murphy, 2007)

	Risk Factors
Demographic	African - American / Aboriginal / Hispanic races Low BMI / poor weight gain / excess weight gain Young maternal age
Obstetric	Previous early pregnancy loss - induced / miscarriage Previous preterm birth - indicated or spontaneous Short inter-pregnancy interval (< 12 months)
Medical	Procedures including Large loop excision of transformation zone (LLETZ) / amniocentesis
Fetal	Fetal gender - Male Multiple pregnancy Assisted conception
Environmental	Periodontal infection Bacterial vaginosis / sexually transmitted infection
Socioeconomic / psychosocial	Social inequality / poverty / neighbourhood disadvantage Physical violence Stressful / traumatic life events / anxiety / depression
Nutritional	Elevated homocysteine / suboptimal vitamin B-12 and B-6 Unbalanced polyunsaturated fatty acids (PUFA) Multivitamin (non-use)
Substance use / toxins	Excess alcohol - ≥ 3 drinks/day or ≥ 7 drinks/week Smoking Cocaine Pollutants - sulphur dioxide, particulate matter

Extremes of maternal age and weight have shown to increase risk of preterm birth. Women aged 35 or older have an estimated odds of 1.8 for preterm birth (Martius et al., 1998). Also, very low maternal weight gain is strongly associated with preterm birth, with an adjusted odds ratio of 9.8. The odds of extreme or severe preterm birth is doubled in women with a high weight gain (Murphy, 2007). In addition, the incidence of preterm birth is higher amongst women with lower income and lower educational

status, with an estimated odds ratio of 2.73 for mothers who had less than 5 years of education (Begum et al., 2003). Also, it is estimated that black women are three to four times more likely to have extreme or severe preterm delivery compared to other ethnic groups (Goldenberg et al., 2008).

Previous history of preterm birth has long been a predictor of preterm birth, as the recurrence rate is high. The relative risk of preterm birth in the second pregnancy is 3.9, and this increases to 6.5 for the next pregnancy (Hacker et al., 2004). Moreover, not only mothers who had a previous history of preterm birth have an increased risk, mothers who were born preterm themselves are also at risk. A study by Porter et al. (Porter et al., 1997) concluded that mothers who were born preterm have a significantly higher risk than those born at term, with an odds ratio of 1.18. Moreover, women with siblings born preterm also have an increased risk of giving birth preterm (Bhattacharya et al., 2010).

Smoking is typically mentioned as a common risk factor for preterm birth. Although the incidence of preeclampsia is lower amongst women who smoke, it is estimated that smoking 10 cigarettes a day will triple the risk of preterm birth compared to non-smoking women (Braillon et al., 2010). Moreover, smoking in pregnancy is also a risk factor for adverse neurodevelopmental outcome in preterm infants. Nevertheless, a study by McCowan et al. (McCowan et al., 2009) concluded that the rate of spontaneous preterm birth in women who quit smoking before 15 weeks of gestation is similar to that of non-smokers. Also, although (Dekker et al., 2012) found an association between smoking and SPTB (as a univariate analysis), smoking was not found to be an independent risk factor for SPTB.

Preterm births are also common in multiple pregnancies. In fact, preterm birth occurs

in approximately 60% of twins, in which 40% were born following spontaneous labour or PPROM. In addition, the majority of women with higher multiple gestations have preterm delivery (Goldenberg et al., 2008).

2.2.2.2. Genetic Risk Factors

In view of the fact that family history and ethnicity contributes to preterm birth, it provides evidence of inherited predisposition for preterm birth. The pro-inflammatory Tumor necrosis factor α (TNFα) (Crider et al., 2005; Murphy, 2007; Liang et al., 2010), Interleukin-1 β (IL1β) (Genc et al., 2002) and Interleukin-6 (IL6) (Simhan et al., 2003; Engel et al., 2005; Goldenberg et al., 2008) appear to be the most consistent genes that are associated with preterm birth (Varner et al., 2005; 2007). Fig. 2.2.2shows a scheme of pathophysiological relevant factors in preterm birth and corresponding candidate genes reviewed by Esplin and Varner (Esplin et al., 2005).

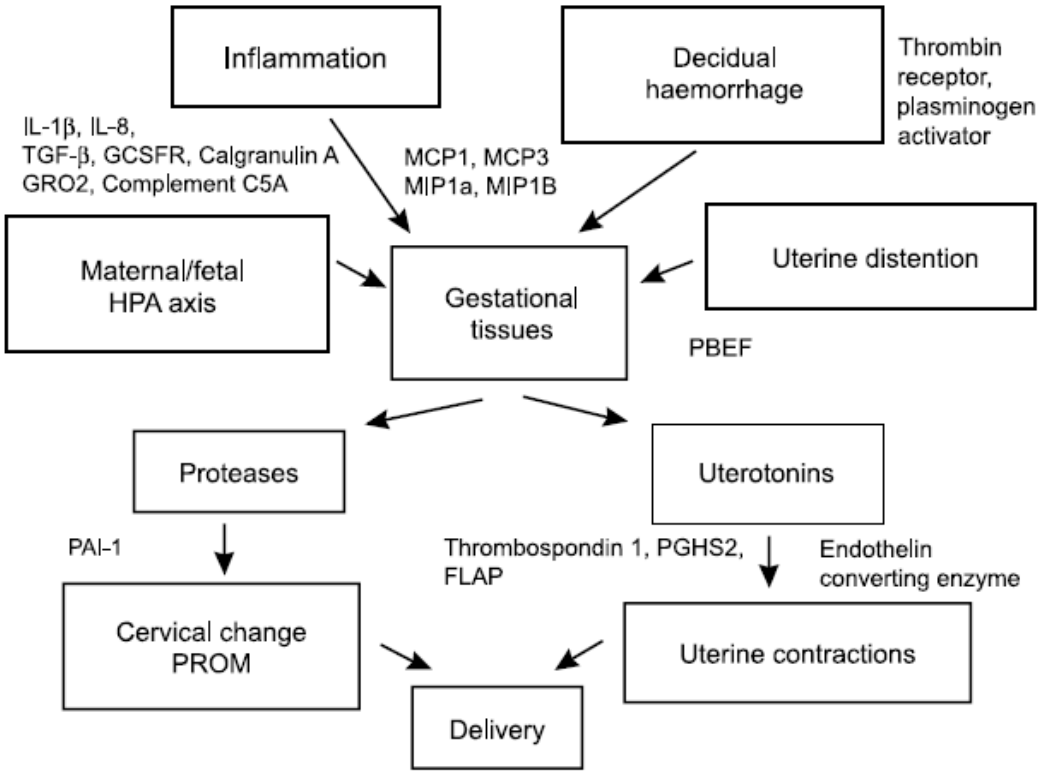

Fig. 2.2.2: Scheme of pathophysiological relevant factors in preterm birth and corresponding candidate genes (Esplin et al., 2005)

Interestingly, a number of gene-environment interactions have been identified where the SNP itself has not been found to be independently associated with preterm birth. For instance, black women who carry the IL6 rs1800795 allele and have bacterial vaginosis appear to have a two-fold increased risk of preterm birth compared to women who did not have bacterial vaginosis but carried the IL6 variant (Goldenberg et al., 2008). However, maternal carriage of the polymorphism in IL6 has not been found to be independently associated with preterm birth (Engel et al., 2005). Also, with the presence of bacterial vaginosis, TNFα rs1800629 has been found to have a gene-environment interaction that increases the risk of preterm birth with an odds ratio of 6.1 (Macones et al., 2004). Moreover, a study by Tsai et al. (Tsai et al., 2008) concluded that for women with high-risk Cytochrome P-4501A1 (CYP1A1 rs1048943) and Glutathione S-transferase Theta 1 (GSTT1 rs71748309) genotypes, the effect of smoking on preterm birth was significantly increased.

On the other hand, genes that are involved in blood clotting (thrombosis), such as Factor V Leiden (F5), Factor VII, Factor XIII and Prothrombin rs1799963 mutation have also been studied. Preterm birth is associated with the maternal carrier status of Factor VII rs5742910 polymorphism, with an odds ratio of 1.7, and with a lower frequency in neonatal Factor XIII rs5985 polymorphism (Murphy, 2007). Fetal factor V has also been identified as a risk factor for preterm labour (O'Callaghan et al., 2013).

Furthermore, Dihydrofolate reductase (DHFR rs70991108) allele has been associated with an increased risk of preterm birth, with an adjusted odds ratio of 3. The risk increases further to an odds ratio of 5.5 for women who also had low folate intake of less than 400 μg/day (Johnson et al., 2005).

2.2.2.3. Current Prediction Methods

Presently, no simple screening or prediction tests that have an optimal predictive value are available for preterm birth (Committee on Understanding Premature Birth and Assuring Healthy Outcomes, 2007). Nevertheless, apart from identifying high-risk women based on risk factors, a number of clinical tests are available. These include vaginal examination, fetal fibronectin, cervical length measurement, and periodontal assessment. A few statistical prediction methods have also been developed. Table 2.2.3 summarizes some publications on potential prediction models for spontaneous preterm birth.

Table 2.2.3: Summary of potential prediction methods for spontaneous preterm birth

Method	Details	Accuracy	Reference
Statistical Logistic regression	**Mid-trimester** Amniotic macrophage inflammatory protein-1 beta Cervical interferon-gamma Monocyte chemotactic protein-1	**Sensitivity:** 91% **Specificity:**84%	(Holst et al., 2009)
Clinical Maternal serum	**2nd trimester** α-fetoprotein (MSAFP)	**Sensitivity:** 95% **Specificity:** 70%	(Hamilton et al., 1985)
Clinical	**Mid-trimester** Fetal fibronectin	**Sensitivity:** 73.7% **Specificity:** 90.9%	(Pieta-Dolinska et al., 2005)
Statistical Neural networks	**1st trimester** Vaginal bleeding	**Sensitivity:** 70%	(Elaveyini et al., 2010)
Clinical Pregnancy history	**Pre-pregnancy** Previous preterm birth	**Sensitivity:** 67% **Specificity:** 73%	(Iams et al., 1998)
Statistical Logistic regression	**Mid-trimester** Phosphorylated isoform of insulin-like growth factor binding protein-1 (phIGFBP-1) Cervical length	**Sensitivity:** 70% **Specificity:** 69%	(Paternoster et al., 2009)

Clinical Maternal	**2nd trimester** Estriol measurement	**Sensitivity:** 57% **Specificity:** 78%	(Heine et al., 2000)
Clinical	**Mid-trimester** Interleukin-6 (IL6) in cervical and vaginal secretions	**Sensitivity:** 50% **Specificity:** 85%	(Lockwood et al., 1994)
Clinical Ultrasound	**15 - 26 weeks** Cervical length	**Sensitivity:** 52% **Specificity:** 82%	(Fuchs et al., 2010)
Statistical Logistic regression	**Mid-trimester** Fetal fibronectin	**Sensitivity:** 61% **Specificity:** 72%	(Lockwood et al., 1991)
Clinical	**Mid-trimester** β-human chorionic gondotrophin (β-hCG)	**Sensitivity:** 74% **Specificity:** 53%	(Lieppman et al., 1993)
Clinical Maternal serum	**2nd trimester** Cortisol-releasing hormone (CRH)	**Sensitivity:** 80% **Specificity:** 44%	(Holzman et al., 2001)
Clinical	**2nd trimester** Interleukin-8 (IL8)	**Sensitivity:** 44% **Specificity:** 80%	(Sakai et al., 2004)
Clinical Digital examination	**2nd trimester** Vaginal examination	**Sensitivity:** 87% **Specificity:** 30%	(Blondel et al., 1990)
Clinical Maternal serum	**1st trimester** Relaxin	**Sensitivity:** 67% **Specificity:** 45%	(Weiss et al., 1993)
Clinical	**1st & 2nd trimester** Periodontal health assessment	**Sensitivity:** 78% **Specificity:** 27%	(Offenbacher et al., 2001)

Many of the prediction models developed are based on predictors from second trimester, or late in pregnancy. Since preterm birth can occur as early as 24 weeks of gestation, the predictive value for these tests may be ineffective and may be too late to provide preventative interventions or treatments.

References

Aali, B. S., Ghafoorian, J., et al. (2004). "Severe preeclampsia and eclampsia in Kerman, Iran: complications and outcomes." *Med Sci Monit* **10**(4): CR163-167.

Abraham, S. (2009). Preterm birth has impact on the long term health of two generations. *Department of Paediatrics*. Auckland, The University of Auckland.

Ahmed, A. S. (2002). "Pre-eclampsia: multisystem involvement and maternal risks." *J R Soc Promot Health* **122**(4): 211-212.

Akolekar, R., Syngelaki, A., et al. (2013). "Competing risks model in early screening for preeclampsia by biophysical and biochemical markers." *Fetal Diagn Ther* **33**(1): 8-15.

Ananth, C. V., Savitz, D. A., et al. (1996). "Maternal cigarette smoking as a risk factor for placental abruption, placenta previa, and uterine bleeding in pregnancy." *Am J Epidemiol* **144**(9): 881-889.

Ansari, M. R., Laghari, M. S., et al. (2008). "Acute renal failure in pregnancy: one year observational study at Liaquat University Hospital, Hyderabad." *J Pak Med Assoc* **58**(2): 61-64.

Arngrimsson, R., Purandare, S., et al. (1993). "Angiotensinogen: a candidate gene involved in preeclampsia?" *Nat Genet* **4**(2): 114-115.

Arngrimsson, R., Sigurard ttir, S., et al. (1999). "A genome-wide scan reveals a maternal susceptibility locus for pre-eclampsia on chromosome 2p13." *Hum Mol Genet* **8**(9): 1799-1805.

Bahado-Singh, R. O., Akolekar, R., et al. (2013). "First-trimester metabolomic detection of late-onset preeclampsia." *Am J Obstet Gynecol* **208**(1): 58 e51-57.

Bashford, M. T., Hefler, L. A., et al. (2001). "Angiotensinogen and endothelial nitric oxide synthase gene polymorphisms among Hispanic patients with preeclampsia." *Am J Obstet Gynecol* **184**(7): 1345-1350; discussion 1350-1341.

Begum, F., Buckshe, K., et al. (2003). "Risk factors associated with preterm labour." *Bangladesh Med Res Counc Bull* **29**(2): 59-66.

Bhattacharya, S., Raja, E. A., et al. (2010). "Inherited predisposition to spontaneous preterm delivery." *Obstet Gynecol* **115**(6): 1125-1133.

Blondel, B., Le Coutour, X., et al. (1990). "Prediction of preterm delivery: is it substantially improved by routine vaginal examinations?" *Am J Obstet Gynecol* **162**(4): 1042-1048.

Bodnar, L. M., Catov, J. M., et al. (2007). "Racial/ethnic differences in the monthly variation of preeclampsia incidence." *Am J Obstet Gynecol* **196**(4): 324 e321-325.

Bodnar, L. M., Ness, R. B., et al. (2005). "The risk of preeclampsia rises with increasing prepregnancy body mass index." *Ann Epidemiol* **15**(7): 475-482.

Bouba, I., Makrydimas, G., et al. (2003). "Interaction between the polymorphisms of the renin-angiotensin system in preeclampsia." *Eur J Obstet Gynecol Reprod Biol* **110**(1): 8-11.

Braillon, A. and Bewley, S. (2010). "The enigma of spontaneous preterm birth." *N Engl J Med* **362**(21): 2032; author reply 2033-2034.

Brennecke, S. (2010). "Australian Action on Pre-eclampsia (AAPEC)." *How dangerous is Pre Eclampsia?* Retrieved 18 July 2010, from http://www.aapec.org.au/Pre-Eclampsia.

Brennecke, S. P., Gude, N. M., et al. (1997). "Reduction of placental nitric oxide synthase activity in pre-eclampsia." *Clin Sci (Lond)* **93**(1): 51-55.

Briceno-Perez, C., Briceno-Sanabria, L., et al. (2009). "Prediction and prevention of preeclampsia." *Hypertension in Pregnancy* **28**(2): 138-155.

Brown, M. A., Lindheimer, M. D., et al. (2001). "The classification and diagnosis of the hypertensive disorders of pregnancy: statement from the International Society for the Study of Hypertension in Pregnancy (ISSHP)." *Hypertens Pregnancy* **20**(1): IX-XIV.

Brown, M. A. and Whitworth, J. A. (1992). "The kidney in hypertensive pregnancies--victim and villain." *Am J Kidney Dis* **20**(5): 427-442.

Buhimschi, C. S., Baumbusch, M. A., et al. (2010). "The role of urinary soluble endoglin in the diagnosis of pre-eclampsia: comparison with soluble fms-like tyrosine kinase 1 to placental growth factor ratio." *Bjog* **117**(3): 321-330.

Buhimschi, C. S., Norwitz, E. R., et al. (2005). "Urinary angiogenic factors cluster hypertensive disorders and identify women with severe preeclampsia." *Am J Obstet Gynecol* **192**(3): 734-741.

Bujold, E., Roberge, S., et al. (2014). "Low-dose aspirin for prevention of adverse outcomes related to abnormal placentation." *Prenat Diagn*.

Carty, D. M., Delles, C., et al. (2008). "Novel biomarkers for predicting preeclampsia." *Trends Cardiovasc Med* **18**(5): 186-194.

Chappell, S. and Morgan, L. (2006). "Searching for genetic clues to the causes of pre-eclampsia." *Clin Sci (Lond)* **110**(4): 443-458.

Chesley, L. C. and Cooper, D. W. (1986). "Genetics of hypertension in pregnancy: possible single gene control of pre-eclampsia and eclampsia in the descendants of eclamptic women." *Br J Obstet Gynaecol* **93**(9): 898-908.

Child Health Research Project (1999). Reducing Perinatal and Neonatal Mortality *Special Report*. Black, R. E. and Kelley, L.

Cincotta, R. B. and Brennecke, S. P. (1998). "Family history of pre-eclampsia as a predictor for pre-eclampsia in primigravidas." *Int J Gynaecol Obstet* **60**(1): 23-27.

Cnattingius, S., Mills, J. L., et al. (1997). "The paradoxical effect of smoking in preeclamptic pregnancies: smoking reduces the incidence but increases the rates of perinatal mortality, abruptio placentae, and intrauterine growth restriction." *Am J Obstet Gynecol* **177**(1): 156-161.

Cnattingius, S., Reilly, M., et al. (2004). "Maternal and fetal genetic factors account for most of familial aggregation of preeclampsia: a population-based Swedish cohort study." *Am J Med Genet A* **130A**(4): 365-371.

Cnossen, J. S., van der Post, J. A., et al. (2006). "Prediction of pre-eclampsia: a protocol for systematic reviews of test accuracy." *BMC Pregnancy Childbirth* **6**: 29.

Committee on Understanding Premature Birth and Assuring Healthy Outcomes (2007). *Preterm Birth: Causes, Consequences, and Prevention*. Washington DC, National Academies Press.

Conde-Agudelo, A., Althabe, F., et al. (1999). "Cigarette smoking during pregnancy and risk of preeclampsia: a systematic review." *Am J Obstet Gynecol* **181**(4): 1026-1035.

Conde-Agudelo, A., Villar, J., et al. (2004). "World Health Organization systematic review of screening tests for preeclampsia." *Obstet Gynecol* **104**(6): 1367-1391.

Crider, K. S., Whitehead, N., et al. (2005). "Genetic variation associated with preterm birth: a HuGE review." *Genet Med* **7**(9): 593-604.

Currie, L., Peek, M., et al. (2002). "Is there an increased maternal-infant prevalence of Factor V Leiden in association with severe pre-eclampsia?" *Bjog* **109**(2): 191-196.

D'Elia, A. V., Driul, L., et al. (2002). "Frequency of factor V, prothrombin and methylenetetrahydrofolate reductase gene variants in preeclampsia." *Gynecol Obstet Invest* **53**(2): 84-87.

Dalmaz, C. A., Santos, K. G., et al. (2006). "Relationship between polymorphisms in thrombophilic genes and preeclampsia in a Brazilian population." *Blood Cells Mol Dis* **37**(2): 107-110.

Dane, C., Buyukasik, H., et al. (2009). "Maternal plasma fibronectin and advanced oxidative protein products for the prediction of preeclampsia in high risk pregnancies: a prospective cohort study." *Fetal Diagn Ther* **26**(4): 189-194.

De Groot, C. J., Bloemenkamp, K. W., et al. (1999). "Preeclampsia and genetic risk factors for thrombosis: a case-control study." *Am J Obstet Gynecol* **181**(4): 975-980.

De Maat, M. P., Jansen, M. W., et al. (2004). "Preeclampsia and its interaction with common variants in thrombophilia genes." *J Thromb Haemost* **2**(9): 1588-1593.

DeCherney, A. H. and Nathan, L. (2002). *CURRENT Obstretric & Gynecologic Diagnosis & Treatment*, McGraw-Hill Medical.

Dekker, G. and Sibai, B. (2001). "Primary, secondary, and tertiary prevention of pre-eclampsia." *The Lancet* **357**(9251): 209-215.

Dekker, G. A., Lee, S. Y., et al. (2012). "Risk factors for preterm birth in an international prospective cohort of nulliparous women." *PLoS One* **7**(7): e39154.

Den Ouden, A. L., Kok, J. H., et al. (1996). "The relation between neonatal thyroxine levels and neurodevelopmental outcome at age 5 and 9 years in a national cohort of very preterm and/or very low birth weight infants." *Pediatr Res* **39**(1): 142-145.

Direkvand-Moghadam, A., Khosravi, A., et al. (2012). "Predictive factors for preeclampsia in pregnant women: a unvariate and multivariate logistic regression analysis." *Acta Biochim Pol* **59**(4): 673-677.

Dizon-Townson, D. S., Nelson, L. M., et al. (1996). "The factor V Leiden mutation may predispose women to severe preeclampsia." *Am J Obstet Gynecol* **175**(4 Pt 1): 902-905.

Doyle, L. W. and Anderson, P. J. (2010). "Pulmonary and neurological follow-up of extremely preterm infants." *Neonatology* **97**(4): 388-394.

Driul, L., Damante, G., et al. (2004). "Screening for pre-eclampsia in a low-risk population at 24 weeks: uterine artery Doppler flow velocimetry and genetic variants of factor V, prothrombin and methylenetetrahydrofolate reductase." *Minerva Ginecol* **56**(5): 385-390.

Ducloy-Bouthors, A. S. (2010). "Clotting disorders and preeclampsia." *Ann Fr Anesth Reanim* **29**(5): e121-134.

Elaveyini, U., Devi, S. P., et al. (2010). "Neural networks prediction of preterm delivery with first trimester bleeding." *Arch Gynecol Obstet* **283**(5): 971-979.

Engel, S. A., Erichsen, H. C., et al. (2005). "Risk of spontaneous preterm birth is associated with common proinflammatory cytokine polymorphisms." *Epidemiology* **16**(4): 469-477.

Eskenazi, B., Fenster, L., et al. (1991). "A multivariate analysis of risk factors for preeclampsia." *Jama* **266**(2): 237-241.

Esplin, M. S., Fausett, M. B., et al. (2001). "Paternal and maternal components of the predisposition to preeclampsia." *N Engl J Med* **344**(12): 867-872.

Esplin, M. S. and Varner, M. W. (2005). "Genetic factors in preterm birth--the future." *Bjog* **112 Suppl 1**: 97-102.

EXPRESS Group (2010). "Incidence of and risk factors for neonatal morbidity after active perinatal care: extremely preterm infants study in Sweden (EXPRESS)." *Acta Paediatr* **99**(7): 978-992.

Farag, K., Hassan, I., et al. (2004). "Prediction of preeclampsia: can it be achieved?" *Obstet Gynecol Surv* **59**(6): 464-482.

Farstad, T., Bratlid, D., et al. (2010). "Bronchopulmonary dysplasia - prevalence, severity and predictive factors in a national cohort of extremely premature infants." *Acta Paediatr* **100**(1): 53-58.

Fuchs, F., Bouyer, J., et al. (2010). "Ultrasound cervical length measurement for prediction of delivery before 32 weeks in women with emergency cerclage for cervical insufficiency." *Int J Gynaecol Obstet* **110**(3): 245-248.

Genc, M. R., Gerber, S., et al. (2002). "Polymorphism in the interleukin-1 gene complex and spontaneous preterm delivery." *Am J Obstet Gynecol* **187**(1): 157-163.

Goldenberg, R. L., Culhane, J. F., et al. (2008). "Epidemiology and causes of preterm birth." *Lancet* **371**(9606): 75-84.

GOPEC Consortium (2005). "Disentangling fetal and maternal susceptibility for pre-eclampsia: a British multicenter candidate-gene study." *Am J Hum Genet* **77**(1): 127-131.

Goplani, K. R., Shah, P. R., et al. (2008). "Pregnancy-related acute renal failure: A single-center experience." *Indian J Nephrol* **18**(1): 17-21.

Gyselaers, W. and Mesens, T. (2009). "Renal interlobar vein impedance index: a potential new Doppler parameter in the prediction of preeclampsia?" *J Matern Fetal Neonatal Med* **22**(12): 1219-1221.

Hacker, N. F., Moore, J. G., et al. (2004). *Essentials of Obstetrics and Gynecology*, Saunders.

Hamilton, M. P., Abdalla, H. I., et al. (1985). "Significance of raised maternal serum alpha-fetoprotein in singleton pregnancies with normally formed fetuses." *Obstet Gynecol* **65**(4): 465-470.

Hanretty, K. (2009). *Obstetrics Illustrated*, Churchill Livingstone.

Hay, J. E. (2008). "Liver disease in pregnancy." *Hepatology* **47**(3): 1067-1076.

Heine, R. P., McGregor, J. A., et al. (2000). "Serial salivary estriol to detect an increased risk of preterm birth." *Obstet Gynecol* **96**(4): 490-497.

Herraiz, I., Arbues, J., et al. (2009). "Application of a first-trimester prediction model for pre-eclampsia based on uterine arteries and maternal history in high-risk pregnancies." *Prenat Diagn* **29**(12): 1123-1129.

Hibbard, J. U., Wilkins, I., et al. (2010). "Respiratory morbidity in late preterm births." *Jama* **304**(4): 419-425.

Holst, R. M., Hagberg, H., et al. (2009). "Prediction of spontaneous preterm delivery in women with preterm labor: analysis of multiple proteins in amniotic and cervical fluids." *Obstet Gynecol* **114**(2 Pt 1): 268-277.

Holzman, C., Jetton, J., et al. (2001). "Second trimester corticotropin-releasing hormone levels in relation to preterm delivery and ethnicity." *Obstet Gynecol* **97**(5 Pt 1): 657-663.

Honest, H., Forbes, C. A., et al. (2009). "Screening to prevent spontaneous preterm birth: systematic reviews of accuracy and effectiveness literature with economic modelling." *Health Technol Assess* **13**(43): 1-627.

Iams, J. D., Goldenberg, R. L., et al. (1998). "The Preterm Prediction Study: recurrence

risk of spontaneous preterm birth. National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network." *Am J Obstet Gynecol* **178**(5): 1035-1040.

Innes, K. E., Marshall, J. A., et al. (1999). "A woman's own birth weight and gestational age predict her later risk of developing preeclampsia, a precursor of chronic disease." *Epidemiology* **10**(2): 153-160.

Irminger-Finger, I., Jastrow, N., et al. (2008). "Preeclampsia: a danger growing in disguise." *Int J Biochem Cell Biol* **40**(10): 1979-1983.

Jacquemyn, Y. and Zemtsova, O. (2010). "Risk factors and prediction of preeclampsia." *Acta Clinica Belgica* **65**(1): 1-12.

James, D. C. S. (2010). "Hypertension." Retrieved 28 July 2010, from http://www.faqs.org/nutrition/Hea-Irr/Hypertension.html.

Jenkins, L. D., Powers, R. W., et al. (2008). "Preeclampsia risk and angiotensinogen polymorphisms M235T and AGT -217 in African American and Caucasian women." *Reprod Sci* **15**(7): 696-701.

Johnson, W. G., Scholl, T. O., et al. (2005). "Common dihydrofolate reductase 19-base pair deletion allele: a novel risk factor for preterm delivery." *Am J Clin Nutr* **81**(3): 664-668.

Kaiser, T., Brennecke, S. P., et al. (2000). "Methylenetetrahydrofolate reductase polymorphisms are not a risk factor for pre-eclampsia/eclampsia in Australian women." *Gynecol Obstet Invest* **50**(2): 100-102.

King, J. F., Slaytor, E. K., et al. (2004). "Maternal deaths in Australia, 1997-1999." *Med J Aust* **181**(8): 413-414.

Klonoff-Cohen, H. S. and Edelstein, S. L. (1996). "Alcohol Consumption During Pregnancy and Preeclampsia." *Journal of Women's Health* **5**(3): 225-230.

Kobashi, G., Hata, A., et al. (2004). "A1166C variant of angiotensin II type 1 receptor gene is associated with severe hypertension in pregnancy independently of T235 variant of angiotensinogen gene." *J Hum Genet* **49**(4): 182-186.

Kramer, M. S. (2003). "The epidemiology of adverse pregnancy outcomes: an overview." *J Nutr* **133**(5 Suppl 2): 1592S-1596S.

Kramer, M. S., Demissie, K., et al. (2000). "The contribution of mild and moderate preterm birth to infant mortality. Fetal and Infant Health Study Group of the Canadian Perinatal Surveillance System." *Jama* **284**(7): 843-849.

Kuromoto, K., Watanabe, M., et al. (2010). "Increases in urinary creatinine and blood pressure during early pregnancy in pre-eclampsia." *Ann Clin Biochem* **47**(4): 336-342.

Kusanovic, J. P., Romero, R., et al. (2009). "A prospective cohort study of the value of maternal plasma concentrations of angiogenic and anti-angiogenic factors in early pregnancy and midtrimester in the identification of patients destined to develop preeclampsia." *J Matern Fetal Neonatal Med* **22**(11): 1021-1038.

Laivuori, H., Lahermo, P., et al. (2003). "Susceptibility loci for preeclampsia on chromosomes 2p25 and 9p13 in Finnish families." *Am J Hum Genet* **72**(1): 168-177.

Laws, P. and Sullivan, E. A. (2009). Australia's mothers and babies 2007. *Perinatal Statistics Series*. Unit, A. N. P. S. Sydney, Australian Institute of Health and Welfare, University of New South Wales.

Lee, J. Y., Kim, H. S., et al. (2010). "Risk factors for periventricular-intraventricular hemorrhage in premature infants." *J Korean Med Sci* **25**(3): 418-424.

Lepercq, J., Coste, J., et al. (2004). "Factors associated with preterm delivery in women with type 1 diabetes: a cohort study." *Diabetes Care* **27**(12): 2824-2828.

Li, Z., Zeki, R., et al. (2010). Australia's mothers and babies 2010. *Perinatal Statistics Series*. Canberra, Australian Institute of Health and Welfare National Perinatal Statistics Unit.

Liang, M., Wang, X., et al. (2010). "Association of combined maternal-fetal TNF-alpha gene G308A genotypes with preterm delivery: a gene-gene interaction study." *J Biomed Biotechnol* **2010**: 396184.

Lieppman, R. E., Williams, M. A., et al. (1993). "An association between elevated

levels of human chorionic gonadotropin in the midtrimester and adverse pregnancy outcome." *Am J Obstet Gynecol* **168**(6 Pt 1): 1852-1856; discussion 1856-1857.

Lockwood, C. J., Ghidini, A., et al. (1994). "Increased interleukin-6 concentrations in cervical secretions are associated with preterm delivery." *Am J Obstet Gynecol* **171**(4): 1097-1102.

Lockwood, C. J., Senyei, A. E., et al. (1991). "Fetal fibronectin in cervical and vaginal secretions as a predictor of preterm delivery." *N Engl J Med* **325**(10): 669-674.

Lumley, J. (2003). "Defining the problem: the epidemiology of preterm birth." *Bjog* **110 Suppl 20**: 3-7.

Luu, T. M., Lefebvre, F., et al. (2010). "Continuing utilisation of specialised health services in extremely preterm infants." *Arch Dis Child Fetal Neonatal Ed* **95**(5): F320-325.

Macones, G. A., Parry, S., et al. (2004). "A polymorphism in the promoter region of TNF and bacterial vaginosis: preliminary evidence of gene-environment interaction in the etiology of spontaneous preterm birth." *Am J Obstet Gynecol* **190**(6): 1504-1508; discussion 1503A.

Martius, J. A., Steck, T., et al. (1998). "Risk factors associated with preterm (<37+0 weeks) and early preterm birth (<32+0 weeks): univariate and multivariate analysis of 106 345 singleton births from the 1994 statewide perinatal survey of Bavaria." *Eur J Obstet Gynecol Reprod Biol* **80**(2): 183-189.

Maternal Perinatal and Infant Mortality Committee (2009). Maternal, Perinatal and Infant Mortality in South Australia 2008. Adelaide, Maternal, Perinatal and Infant Mortality Committee. SA Health, Government of South Australia.

McCarthy, F. P., O'Keeffe, L. M., et al. (2013). "Association between maternal alcohol consumption in early pregnancy and pregnancy outcomes." *Obstet Gynecol* **122**(4): 830-837.

McCowan, L. M., Dekker, G. A., et al. (2009). "Spontaneous preterm birth and small for gestational age infants in women who stop smoking early in pregnancy: prospective cohort study." *Bmj* **338**: b1081.

Medica, I., Kastrin, A., et al. (2007). "Genetic polymorphisms in vasoactive genes and preeclampsia: a meta-analysis." *Eur J Obstet Gynecol Reprod Biol* **131**(2): 115-126.

Mikkola, K., Ritari, N., et al. (2005). "Neurodevelopmental outcome at 5 years of age of a national cohort of extremely low birth weight infants who were born in 1996-1997." *Pediatrics* **116**(6): 1391-1400.

Moodley, J. (2008). "Maternal deaths due to hypertensive disorders in pregnancy." *Best Pract Res Clin Obstet Gynaecol* **22**(3): 559-567.

Moses, E. K., Lade, J. A., et al. (2000). "A genome scan in families from Australia and New Zealand confirms the presence of a maternal susceptibility locus for pre-eclampsia, on chromosome 2." *Am J Hum Genet* **67**(6): 1581-1585.

Mostello, D., Droll, D. A., et al. (2003). "Tertiary care improves the chance for vaginal delivery in women with preeclampsia." *Am J Obstet Gynecol* **189**(3): 824-829.

Murphy, D. J. (2007). "Epidemiology and environmental factors in preterm labour." *Best Practice & Research Clinical Obstetrics & Gynaecology* **21**(5): 773-789.

Mutze, S., Rudnik-Schoneborn, S., et al. (2008). "Genes and the preeclampsia syndrome." *J Perinat Med* **36**(1): 38-58.

Myers, J. E., Kenny, L. C., et al. (2013). "Angiogenic factors combined with clinical risk factors to predict preterm pre-eclampsia in nulliparous women: a predictive test accuracy study." *Bjog* **120**(10): 1215-1223.

Navaei, F., Aliabady, B., et al. (2010). "Early outcome of preterm infants with birth weight of 1500 g or less and gestational age of 30 weeks or less in Isfahan city, Iran." *World J Pediatr*: ahead of print.

Nickelodeon Parents and Preschool Network (2010). "3D Pregnancy." Retrieved 11 August 2010, from http://3dpregnancy.parentsconnect.com/.

Nijdam, M. E., Timmerman, M. R., et al. (2009). "Cardiovascular risk factor assessment after pre-eclampsia in primary care." *BMC Fam Pract***10**: 77.

North, R. A., McCowan, L. M. E., et al. (2011). "Clinical risk prediction for pre-eclampsia in nulliparous women: development of model in international prospective cohort." *Bmj* **342**.

O'Callaghan, M. E., MacLennan, A. H., et al. (2013). "Single-nucleotide polymorphism associations with preterm delivery: a case-control replication study and meta-analysis." *Pediatr Res* **74**(4): 433-438.

Offenbacher, S., Lieff, S., et al. (2001). "Maternal periodontitis and prematurity. Part I: Obstetric outcome of prematurity and growth restriction." *Ann Periodontol***6**(1): 164-174.

Papageorghiou, A. T. and Campbell, S. (2006). "First trimester screening for preeclampsia." *Curr Opin Obstet Gynecol* **18**(6): 594-600.

Papageorghiou, A. T., Yu, C. K., et al. (2005). "Assessment of risk for the development of pre-eclampsia by maternal characteristics and uterine artery Doppler." *British Journal of Obstetrics and Gynaecology* **112**(6): 703-709.

Paternoster, D., Riboni, F., et al. (2009). "Phosphorylated insulin-like growth factor binding protein-1 in cervical secretions and sonographic cervical length in the prediction of spontaneous preterm delivery." *Ultrasound Obstet Gynecol* **34**(4): 437-440.

Peterson, H. (2010). Genetic Studies of Pre-eclampsia. *Department of Biosciences and Nutrition*. Stockholm, Sweden, Karolinska Institute**:** 74.

Pfeifer, S. M. (2007). *NMS Obstetrics and Gynecology*, Lippincott Williams & Wilkins.

Pieta-Dolinska, A., Jaczewski, B., et al. (2005). "[Assessment of fetal fibronectin concentration in identifying patients at risk for preterm delivery]." *Ginekol Pol* **76**(6): 431-435.

Pipkin, F. B. (2008). "Smoking in moderate/severe preeclampsia worsens pregnancy outcome, but smoking cessation limits the damage." *Hypertension* **51**(4): 1042-1046.

Platzker, A. C. (1972). "Assessment of the risk of the respiratory distress syndrome with preterm delivery." *Calif Med* **117**(2): 49.

Poon, L. C., Kametas, N. A., et al. (2009). "First-trimester prediction of hypertensive disorders in pregnancy." *Hypertension* **53**(5): 812-818.

Porter, T. F., Fraser, A. M., et al. (1997). "The risk of preterm birth across generations." *Obstet Gynecol* **90**(1): 63-67.

Powers, R. W., Minich, L. A., et al. (1999). "Methylenetetrahydrofolate reductase polymorphism, folate, and susceptibility to preeclampsia." *J Soc Gynecol Investig* **6**(2): 74-79.

Rahman, T. M. and Wendon, J. (2002). "Severe hepatic dysfunction in pregnancy." *Qjm* **95**(6): 343-357.

Romero, R., Velez Edwards, D. R., et al. (2010). "Identification of fetal and maternal single nucleotide polymorphisms in candidate genes that predispose to spontaneous preterm labor with intact membranes." *Am J Obstet Gynecol* **202**(5): e1-34.

Ros, H. S., Cnattingius, S., et al. (1998). "Comparison of risk factors for preeclampsia and gestational hypertension in a population-based cohort study." *Am J Epidemiol* **147**(11): 1062-1070.

Said, J. (2006). "Managing pre-eclampsia." *O&G Magazine* **8**(4): 43-44.

Sakai, M., Ishiyama, A., et al. (2004). "Relationship between cervical mucus interleukin-8 concentrations and vaginal bacteria in pregnancy." *Am J Reprod Immunol* **52**(2): 106-112.

SCOPE Consortium (2004). "The SCOPE Pregnancy Research Study." Retrieved 5 March 2010, from http://www.scopestudy.net/.

Sekizawa, A., Purwosunu, Y., et al. (2010). "Prediction of pre-eclampsia by an analysis of placenta-derived cellular mRNA in the blood of pregnant women at 15-20 weeks of gestation." *Bjog* **117**(5): 557-564.

Shah, V. A., Yeo, C. L., et al. (2005). "Incidence, risk factors of retinopathy of prematurity among very low birth weight infants in Singapore." *Ann Acad Med Singapore* **34**(2): 169-178.

Shinsato, R. N., Paccola, L., et al. (2010). "[Frequency of retinopathy of prematurity in newborns at the Clinical Hospital, Ribeirao Preto Medical School, University of Sao Paulo]." *Arq Bras Oftalmol* **73**(1): 60-65.

Sibai, B., Dekker, G., et al. (2005). "Pre-eclampsia." *The Lancet* **365**(9461): 785-799.

Simhan, H. N., Krohn, M. A., et al. (2003). "Interleukin-6 promoter -174 polymorphism and spontaneous preterm birth." *Am J Obstet Gynecol* **189**(4): 915-918.

Spinillo, A., Capuzzo, E., et al. (1994). "Cigarette smoking in pregnancy and risk of pre-eclampsia." *J Hum Hypertens* **8**(10): 771-775.

Steegers, E. A. P., von Dadelszen, P., et al. (2010). "Pre-eclampsia." *The Lancet* **376**(9741): 631-644.

Stekkinger, E., Zandstra, M., et al. (2009). "Early-onset preeclampsia and the prevalence of postpartum metabolic syndrome." *Obstet Gynecol* **114**(5): 1076-1084.

Stone, C. D., Diallo, O., et al. (2007). "The combined effect of maternal smoking and obesity on the risk of preeclampsia." *J Perinat Med* **35**(1): 28-31.

Sullivan, E. A., Hall, B., et al. (2008). Maternal deaths in Australia, 2003-2005. *Maternal Deaths Series*. Canberra, AIHW National Perinatal Statistics Unit.

Thilaganathan, B., Wormald, B., et al. (2010). "Early-pregnancy multiple serum markers and second-trimester uterine artery Doppler in predicting preeclampsia." *Obstet Gynecol* **115**(6): 1233-1238.

Tsai, H. J., Liu, X., et al. (2008). "Maternal cigarette smoking, metabolic gene polymorphisms, and preterm delivery: new insights on GxE interactions and pathogenic pathways." *Hum Genet* **123**(4): 359-369.

Tubbergen, P., Lachmeijer, A. M., et al. (1999). "Change in paternity: a risk factor for preeclampsia in multiparous women?" *J Reprod Immunol* **45**(1): 81-88.

Tuffnell, D. J., Jankowicz, D., et al. (2005). "Outcomes of severe pre-eclampsia/eclampsia in Yorkshire 1999/2003." *Bjog* **112**(7): 875-880.

VanStone, W. (2010). "Retinopathy of prematurity: an example of a successful screening program." *Neonatal Netw* **29**(1): 15-21.

Varner, M. W. and Esplin, M. S. (2005). "Current understanding of genetic factors in preterm birth." *Bjog* **112 Suppl 1**: 28-31.

Verlohren, S., Galindo, A., et al. (2010). "An automated method for the determination of the sFlt-1/PIGF ratio in the assessment of preeclampsia." *Am J Obstet Gynecol* **202**(2): change ce161-171.

Villar, J., Abalos, E., et al. (2004). "Heterogeneity of perinatal outcomes in the preterm delivery syndrome." *Obstet Gynecol* **104**(1): 78-87.

von Dadelszen, P., Payne, B., et al. (2010). "Prediction of adverse maternal outcomes in pre-eclampsia: development and validation of the fullPIERS model." *Lancet*.

Vucic, N., Frleta, M., et al. (2009). "Thrombophilia, preeclampsia and other pregnancy complications." *Acta Med Croatica* **63**(4): 297-305.

Wagner, L. K. (2004). "Diagnosis and management of preeclampsia." *Am Fam Physician* **70**(12): 2317-2324.

Walsh, S. W. (2007). "Obesity: a risk factor for preeclampsia." *Trends Endocrinol Metab* **18**(10): 365-370.

Wang, C. N., Chang, S. D., et al. (2010). "Change in amniotic fluid levels of multiple anti-angiogenic proteins before development of preeclampsia and intrauterine growth restriction." *J Clin Endocrinol Metab* **95**(3): 1431-1441.

Ward, K., Hata, A., et al. (1993). "A molecular variant of angiotensinogen associated with preeclampsia." *Nat Genet* **4**(1): 59-61.

Weiss, G., Goldsmith, L. T., et al. (1993). "Elevated first-trimester serum relaxin concentrations in pregnant women following ovarian stimulation predict prematurity risk and preterm delivery." *Obstet Gynecol* **82**(5): 821-828.

Wilson, M. L., Goodwin, T. M., et al. (2003). "Molecular epidemiology of

preeclampsia." *Obstet Gynecol Surv* **58**(1): 39-66.

Woynarowska, M., Rutkowska, M., et al. (2008). "Risk factors, frequency and severity of bronchopulmonary dysplasia (BPD) diagnosed according to the new disease definition in preterm neonates." *Med Wieku Rozwoj* **12**(4 Pt 1): 933-941.

Wright, D., Akolekar, R., et al. (2012). "A competing risks model in early screening for preeclampsia." *Fetal Diagn Ther* **32**(3): 171-178.

Xiong, X., Wang, F. L., et al. (2000). "Maternal smoking and preeclampsia." *J Reprod Med* **45**(9): 727-732.

Yoshimura, T., Yoshimura, M., et al. (2000). "Association of the missense Glu298Asp variant of the endothelial nitric oxide synthase gene with severe preeclampsia." *J Soc Gynecol Investig* **7**(4): 238-241.

Yu, C. K., Khouri, O., et al. (2008). "Prediction of pre-eclampsia by uterine artery Doppler imaging: relationship to gestational age at delivery and small-for-gestational age." *Ultrasound Obstet Gynecol* **31**(3): 310-313.

Zafarmand, M. H., Franx, A., et al. (2008). "The M235T variant of the angiotensinogen gene is related to development of self-reported hypertension during pregnancy: the Prospect-EPIC cohort study." *Hypertens Res* **31**(7): 1299-1305.

Zhang, J., Klebanoff, M. A., et al. (1999). "The puzzling association between smoking and hypertension during pregnancy." *Am J Obstet Gynecol* **181**(6): 1407-1413.

Zhang, J., Meikle, S., et al. (2003). "Severe maternal morbidity associated with hypertensive disorders in pregnancy in the United States." *Hypertens Pregnancy* **22**(2): 203-212.

Zhou, N., Yu, P., et al. (1999). "Detection of insertion/deletion polymorphism of angiotensin converting enzyme gene in preeclampsia." *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* **16**(1): 29-31.

Chapter 3: Scope Database

3.1. Overview

The data used in this study have been obtained from the Screening fOr Pregnancy Endpoints (SCOPE) project (SCOPE Consortium, 2004), which aims at building a clinical database and pregnancy biobank to screen candidate markers of pregnancy complications.

Nulliparous women with singleton pregnancies were recruited to the SCOPE study between November 2004 and February 2011 in Adelaide, Australia, and Auckland, New Zealand. Women were invited to participate prior to 15 weeks' gestation when attending hospital antenatal clinics, obstetricians, general practitioners or community midwives, and were interviewed and examined by a research midwife at 15±1 and 20±1 weeks of gestation. All women provided written informed consent.

The exclusion criteria included women who were considered to be at high risk of preeclampsia, delivering preterm or small for gestational age babies due to underlying medical conditions (e.g. chronic hypertension requiring antihypertensive medication or diabetes), previous cervical knife cone biopsy, 3 terminations or 3 miscarriages, current ruptured membranes, or their pregnancy was complicated by a known major fetal anomaly or abnormal karyotype, or if they received interventions that may modify pregnancy outcome(e.g. aspirin, cervical suture).

A total of 3201 participants were included, in which 1169 were from Adelaide and 2032 were from Auckland. Maternal and paternal demographic information (including age, ethnicity, immigration details, education, work, socioeconomic index, income level,

living situation), health conditions (including BMI and pre-existing health conditions), medical, pregnancy and family history, along with dietary and lifestyle questionnaires at 15 weeks' and 20 weeks' gestation (including self-reported fruit, vegetables, cigarette and alcohol consumption, legal and illicit drug use), as well as details of antenatal visits (e.g. blood pressure and ultrasound), and neonatal records (e.g. birth weight) were recorded into an internet accessed, central database with a complete audit trail (Medscinet^{AB}, Stockholm, Sweden).

In addition to the clinical measurements and pregnancy outcomes, maternal, paternal and neonatal genotypes for 100 Single-nucleotide polymorphisms (SNPs) were also recorded. Analyses will be performed on combinations of clinical measurements and SNPs.

3.1.1. Outcomes

More than nine pregnancy outcomes are recorded. These include preeclampsia, small/large for gestational age babies, preterm birth, neonatal morbidity, gestational hypertension, gestational diabetes, placental abruption and maternal death. Fig. 3.1.1**:** shows the proportions of pregnancy outcomes in the SCOPE database. This thesis will focus on Preeclampsia (PE) and Spontaneous Preterm birth (SPTB), which accounts for 6% and 5% of the outcomes respectively.

In exploratory analysis (Section 3.4) PE and SPTB will be compared against uncomplicated pregnancies (Controls) in an attempt to obtain a list of potential variables that are unique to PE or SPTB. However, the final models will be classifying PE or SPTB with non-cases for prediction purpose. It is worth noting that while this thesis only focus on two complications, the statistical analyses and methods discussed may still be applicable to other pregnancy outcomes.

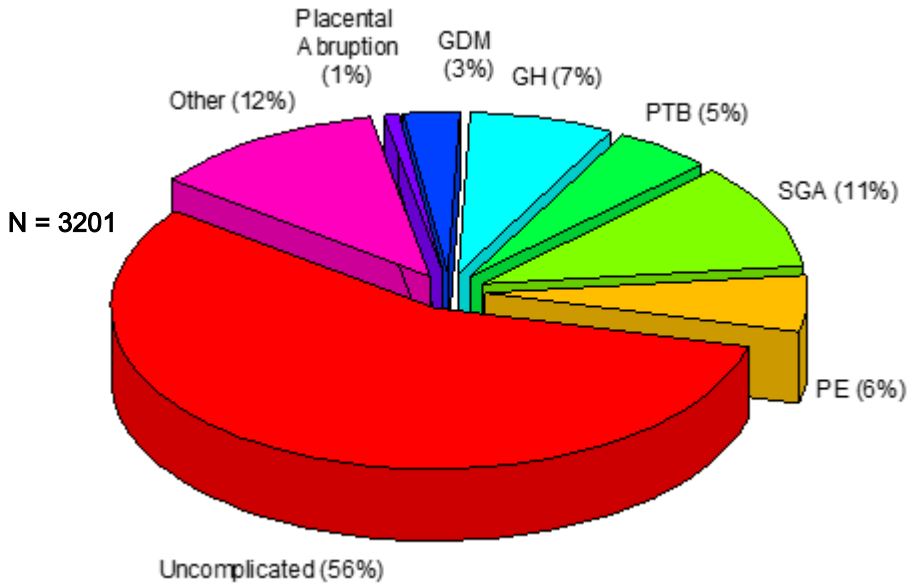

Fig. 3.1.1: Pregnancy Outcomes in SCOPE

PE=preeclampsia; SGA=small for gestational age; SPTB=spontaneous preterm birth; GH=gestational hypertension; GDM=gestational diabetes mellitus

3.1.2. Data Quality Control

All data collected data were monitored and checked for each participant, including checks for transcription errors of lifestyle questionnaire, and detection of illogical or inconsistent data. Each component of the database are closely monitored and updates are made when an error was found or when new information becomes available (SCOPE Consortium, 2007). All genotype data have been tested for Hardy-Weinberg equilibrium, and SNP trios were checked for Mendelian errors.

3.2. Data Structure

The SCOPE database is made up of 2 components: clinical measurements and lifestyle records for each patient and genotypes for 100 SNPs in pregnancy trios (mother, partner and baby). This section will discuss the structure of each component of the database and how they are merged and prepared for analysis.

More than 1000 variables on clinical measurements, maternal history, family history, and lifestyles are recorded in the clinical measurements database across 3 time-points (pre-pregnancy, 1st visit at 15 weeks of gestation, and 2nd visit at 20 weeks of gestation). Lifestyle questionnaires were used to collect information on dietary practices, including self-reported drug use. Socio-economic Index (SEI) were obtained as a continuum ranging from 10 to 90, with 10 being the lowest and 90 the highest (Davis et al., 2003). Paternal information is also recorded, as well as neonatal data. The layout of the clinical measurements database is shown in Fig. 3.2.1.

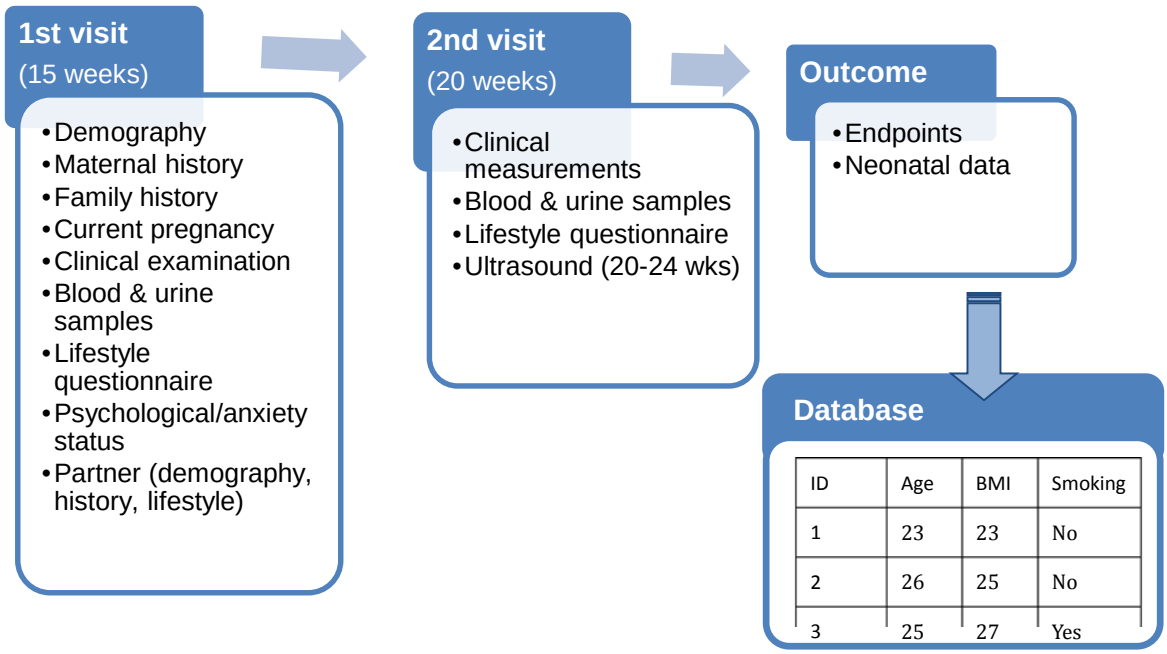

Fig. 3.2.1: Clinical measurements database layout

A more complex structure is used to store and manage the SNP database, as the raw data and updates are identified based on the plate position they are genotyped, and there are a total of 95 plates, with each unique ID having three records (i.e. the trios) for 100 SNPs (a complete list of SNPs in the database is shown in Appendix).

Hence, information are stored into 3D arrays (shown in Fig. 3.2.2), with the first two dimensions as the plate position (i.e. $n_i = 8$ and $p_i = 12$, where $i = 1, \dots, 95$), and the third dimension are 3 data identifiers and 100 SNPs recorded for the corresponding patient (i.e. $q_i = 103$, where $i = 1, \dots, 95$). There are a total of 95 arrays, and each entry is identified by its Plate ID, plate position, registration ID, and Patient (i.e. maternal, paternal, or neonatal).

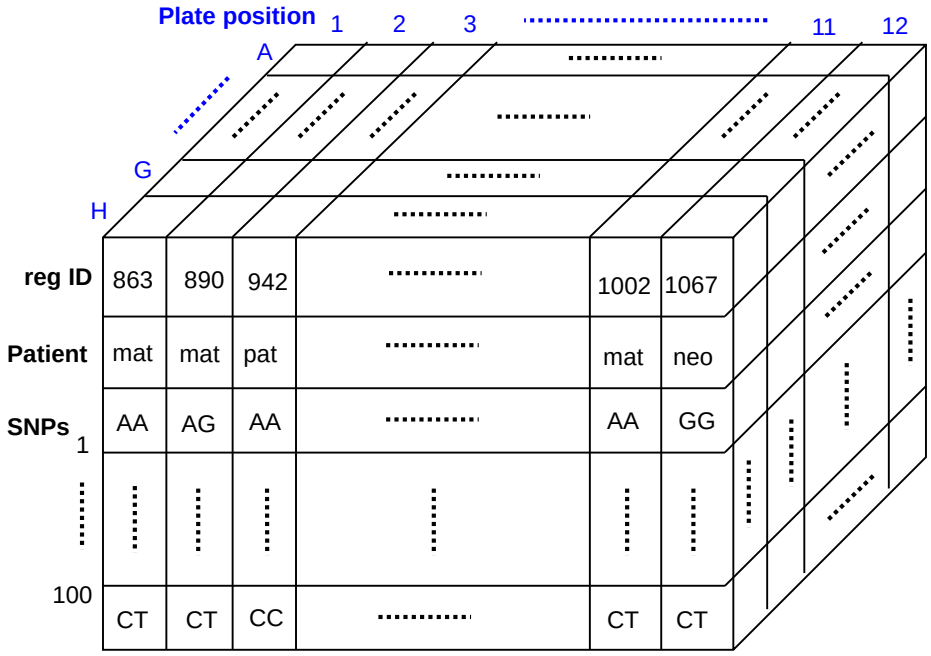

Fig. 3.2.2: Single 3D array structure

The forty arrays are then merged together to form an $8 \times 480 \times 103$ array, sorted by maternal, paternal and neonatal data. Since the Plate position is not needed for further matching, the 3D array structure is no longer needed. Hence, three 103×3201 matrices (containing SNPs data for mothers, fathers, and babies separately) are

extracted and further transposed to obtain a similar layout as the clinical measurements database, where the two databases are merged together by matching the registration ID (Fig. 3.2.3).

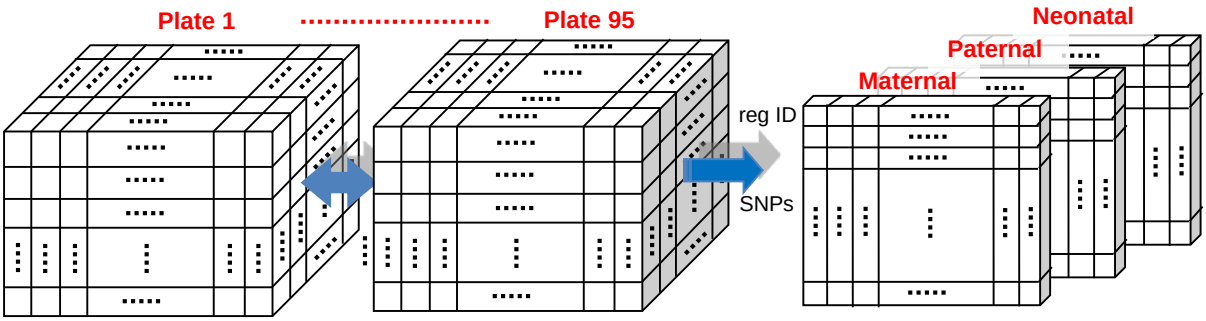

Fig. 3.2.3: Multiple 3D array structure

3.3. Subgroup Differences

Prior to exploratory analysis for potential predictors of pregnancy complications, one precaution that needs to be considered is the regional differences (between Adelaide, Australia and Auckland, New Zealand). Since there are known demographic differences between the two populations, determining whether the results of the analyses are real predictor-outcome effects or just the effects of demographic differences is crucial, as it will severely affect the generalization of the prediction models obtained.

This section will first explore the natural differences between the two subgroups, and discuss the use of stratified and subgroup analysis in the context of this study, followed by exploratory data analysis results.

3.3.1. Statistical Tests on Significant Subgroup Differences

Statistical tests on the differences between the two subgroups for all variables (including outcomes) are done using proportional test (for categorical variables) and Kruskal-Wallis test (for continuous variables).

The proportional test (Wilson, 1927; Newcombe, 1998) is used to test the difference in proportions of a categorical variable. Its null hypothesis is given by $H_0: p_1 = p_2$, where p is the probability of success in a subgroup of a given variable. In other words, it tests whether the proportions of a variable between two subgroups are equal.

For continuous variables, the Kruskal-Wallis test (Hollander et al., 1973) is used as a non-parametric alternative to the t-test, as some variables do not satisfy the normality assumption. The null hypothesis is given by H_0: all sample distribution functions are equal. In other words, it tests whether the two subgroups are from an identical population for a given variable.

Approximately 470 variables (including outcomes) are significantly different between the Adelaide and Auckland population, with a P-value less than 0.05. A full list of variables that are statistically different is given in Appendix.

One apparent difference is the prevalence of adverse pregnancy outcomes between Adelaide and Auckland SCOPE women (Fig. 3.3.1). Compared to Auckland women, Adelaide women had a greater proportion of complicated pregnancies, and the prevalence of all outcomes recorded in the SCOPE database are also higher. The prevalence of preeclampsia (PE) (8% in Adelaide vs. 4% in Auckland; $P = 0.000$), spontaneous preterm birth (SPTB) (6% in Adelaide vs 4% in Auckland; $P = 0.033$), gestational hypertension (GH) (10% in Adelaide vs 6%in Auckland; $P = 0.000$), and

gestational diabetes mellitus (GDM) (4% in Adelaide vs 2%in Auckland; $P = 0.000$) are significantly different between the two subgroups, as well as the percentage of uncomplicated pregnancies (44% in Adelaide vs 63% in Auckland; $P = 0.000$).

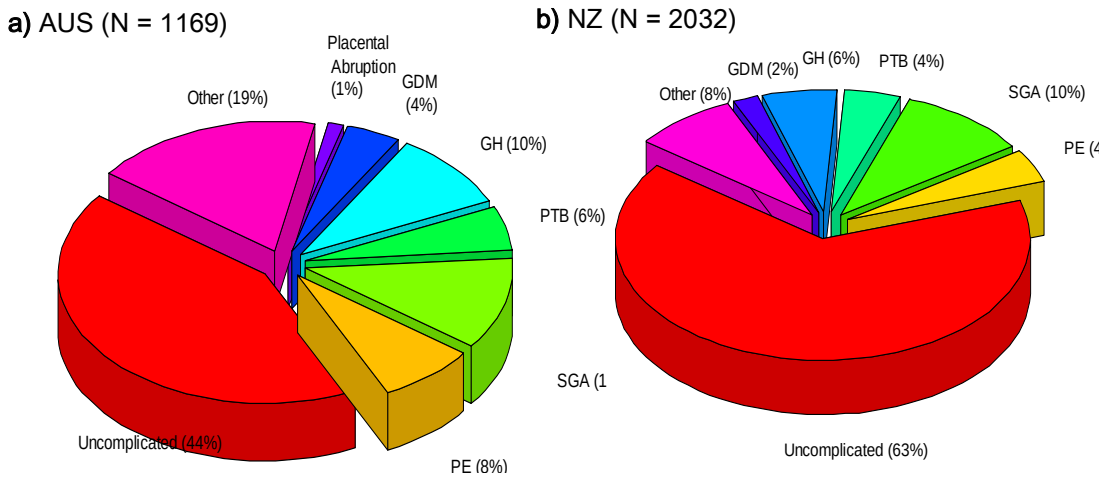

Fig. 3.3.1: Outcome differences between (a) Adelaide and (b) Auckland SCOPE women

A number of family history and dietary variables are also significantly different under the proportional test, including common factors such as family history of PE (14% in Adelaide vs. 7.9% in Auckland; $P = 0.000$), SPTB (20% in Adelaide vs 14.5% in Auckland; $P = 0.000$), CH (34.8% in Adelaide vs 42% in Auckland; $P = 0.000$), IHD (11% in Adelaide vs 18.2% in Auckland; $P = 0.0000$), fruit consumption more than 3 times per day during 1st trimester (15.7% in Adelaide vs. 54.5% in Auckland; $P = 0.0000$), high vegetable consumption of more than 3 times per day during 1st trimester (3.2% in Adelaide vs 12.5% in Auckland; $P = 0.000$), and any folate (31% in Adelaide vs 66.8% in Auckland; $P = 0.000$) or multivitamin supplementation (16.6% in Adelaide vs. 33% in Auckland; $P = 0.000$) 3 months prior to pregnancy. (Refer to the Appendix for a comprehensive list of variables with subgroup differences)

For continuous variables, common demographics such as age (average 24 yrs old in Adelaide vs. 30 yrs old in Auckland; $P = 0.000$), level of education (average 11 yrs of schooling in Adelaide vs 13 yrs in Auckland; $P = 0.000$), socio-economic index

(average 27.8 in Adelaide vs 47.9 in Auckland; $P = 0.000$), and BMI (average 27.1 kg/m^2 in Adelaide vs. 24.8 kg/m^2 in Auckland; $P = 0.000$) are significantly different.

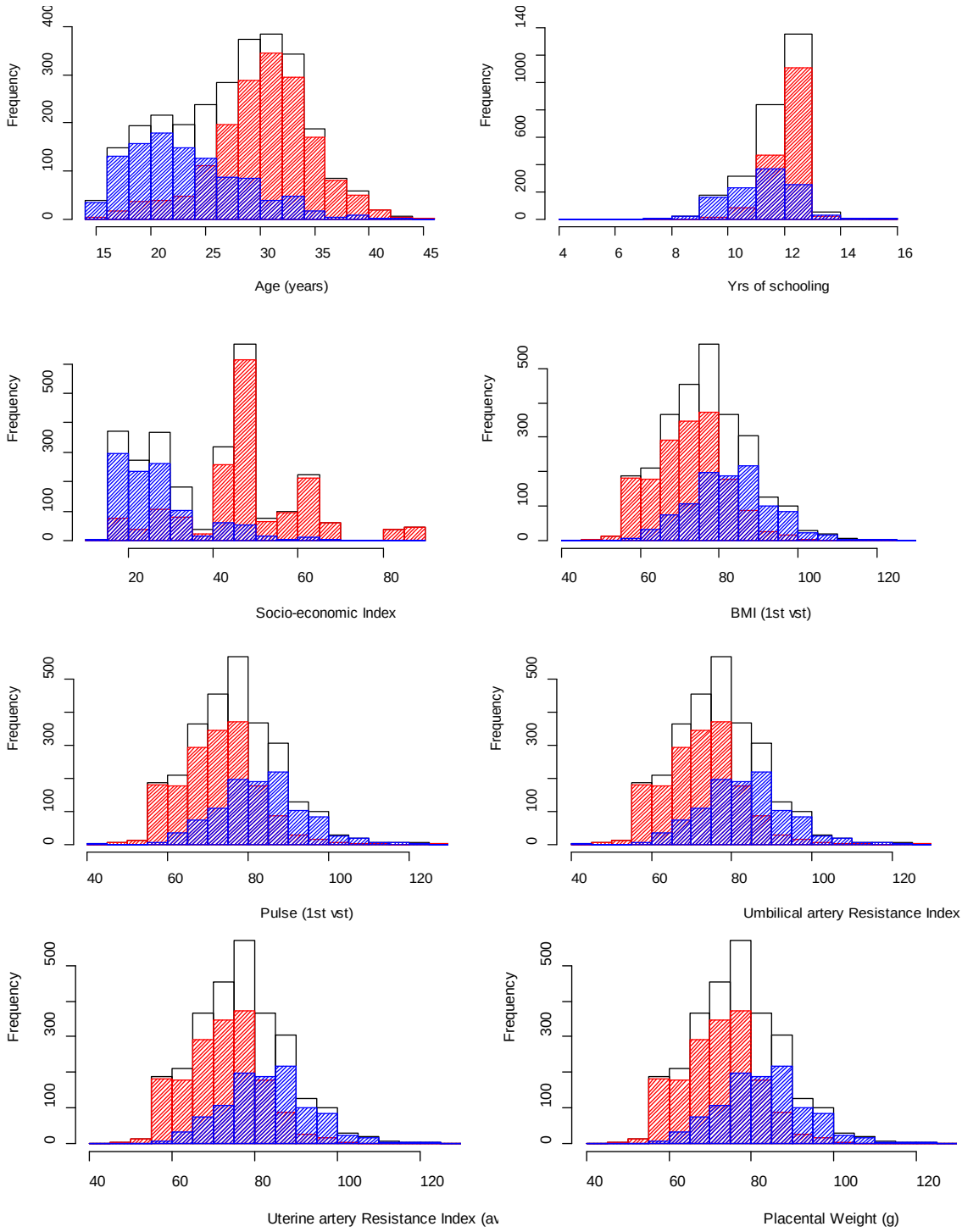

Fig. 3.3.2: Common clinical measurement differences between Adelaide (shaded in blue) and Auckland (shaded in red) SCOPE pregnancies, with overall distribution (shown in white)

In addition, clinical measurements including pulse rate at first visit (average 83.7 per minute in Adelaide vs. 73.5 per minute in Auckland; $P = 0.000$), umbilical artery resistance index at 20 weeks (average 0.75 in Adelaide vs. 0.72 in Auckland; $P = 0.000$), uterine artery resistance index at 20 weeks (average 0.60 in Adelaide vs 0.54 in Auckland; $P = 0.000$), and placental weight (average 573g in Adelaide vs 641g in Auckland; $P = 0.000$) are also significantly different (Fig. 3.3.2).

Based on the test results, nearly half of the variables collected for Adelaide and Auckland SCOPE pregnancies have different prevalence and distributions. Interestingly, despite the Auckland patients being older, they seem to perform better in terms of pregnancy outcomes (with a lower prevalence of pregnancy complications) and live a healthier lifestyle (with greater fruit and vegetable consumption and fewer cigarette smokers).

3.3.2. Feasibility of Subgroup Analysis

Since there are a large number of differences, the subgroup effects cannot be ignored, as the data resembles two populations with diverse characteristics. This raises the question of whether it is appropriate to include the whole cohort in analyses, where predictor effects are examined on the combined distribution, or to perform a stratified analysis, in which the effects are assessed separately.

While stratified analysis is commonly used as a conventional method to control for confounding factors, concerns have been raised in the literature on overemphasized or misleading subgroup effects resulting from inappropriate analyses (Cui et al., 2002; Fletcher, 2007; Wang et al., 2007; Sun et al., 2012).

Issues of increased false-discovery rate due to more tests being performed (i.e. multiple testing scenario) and decreased statistical power due to smaller sample size in each individual stratum increases the chance of having false or misleading subgroup effects (Moyé, 2012; Pandis, 2013). While these issues can be partly addressed by correcting for false discovery rates, a study by Yusuf et al. (Yusuf et al., 1991) showed that even very conservative tests may produce unreliable and non-replicable subgroup effects, and that more reliable estimates of effects are usually obtained using combined or non-stratified data.

Hence, alternative statistical tests on interactions of subgroups based on non-stratified data provide a more robust approach (Guillemin, 2007; Klebanoff, 2007). Such tests can be performed with interaction terms added to common methods including ANCOVA and linear or logistic regressions. However, as the aim of this study is to develop a general prediction, having location as an interaction effect or predictor is not practical. For this reason, the Cochran-Mantel-Haenszel test is chosen as an exploratory analysis to examine whether the effects of predictors for patients in Adelaide and Auckland are equal.

3.3.3. Cochran-Mantel-Haenszel Test

The Cochran-Mantel-Haenszel (CMH) test (Mantel et al., 1959; Agresti, 2002) compares two groups on a binary response (i.e. with complications or not) adjusting for control variables or stratifiers. In other words, it tests whether predictors, within each stratum, are independent from each other. This is achieved by examining the differences in odds ratios, which describe the strength of association or non-independence between two binary data values, for two stratified groups.

The CMH null hypothesis is given by $H_0: \theta_1 = \theta_2 = 1$, and the test statistic is obtained using $CMH = \frac{[\sum_k (n_{11k} \widehat{\mu_{11k}})]^2}{\sum_k Var(n_{11k})}$, where $\mu_{11k} = \frac{n_{1+k} n_{+1k}}{n_{++k}}$ and $Var(n_{11k}) = \frac{n_{1+k} n_{2+k} n_{+1k} n_{+2k}}{n_{++K}^2 (n_{++K} - 1)}$. The p-values are calculated for all variables comparing complications (i.e. Preeclampsia and Preterm birth) across the two locations (i.e. Adelaide and Auckland).

One point to note is that it has been acknowledged that the MH estimates of odds ratio has an assumption of homogeneous odds ratios, nevertheless, since it is only used for preliminary analyses to examine whether a difference is present or not (i.e. the MH estimate is not analyzed), further testing of interest can be done using Breslow-Day test (for testing homogeneity) followed by Woolf's test (Woolf, 1955; Liu, 2005).

3.4. Exploratory Data Analysis

This section shows the basic univariate tests of association results for all variables in the database on Preeclampsia and Preterm Birth. A simple regression is used to examine the relationship of continuous variables and complications, while for categorical variables and SNPs, a Fisher's exact test is used.

Again, while this is only an exploratory analysis, results shown in this section have not been controlled for false discovery rates. The aim of this analysis is to scale down the list of variables specifically for each complication, and issues of false discovery rates will be addressed in further analysis during the model selection process (discussed in Chapter 4).

3.4.1. Univariate Analysis

Many common factors, including BMI, SEI and family history, are consistent with the literature (discussed in Chapter 2). A list of common factors associated with Preeclampsia and Preterm birth are shown in Table 3.4.1 and Table 3.4.2. Full lists of test results are provided in Appendix III. All variables with $P<0.05$ are selected, with Fisher’s exact test for categorical variables or t-test for continuous variables.

Table 3.4.1: Univariate analysis of common factors associated with PE

Variable	N (Control)	N (PE)	Mean±SD / % (Control)	Mean±SD / % (PE)	P
SEI	1984	178	41.7 ± 16.41	36.7 ± 15.56	0.0001
Folate supplementation in 1st trimester (μg/day)	1984	178	730 ± 551	622 ± 262	0.0042
MAP at 15 wks (mmHg)	1984	178	78.2 ± 7.80	84.2 ± 8.46	0.0000
BMI (kg/m^2)	1984	178	25.0 ± 4.66	28.3 ± 6.83	0.0000
Family history (GH)	141	21	7.11%	11.80%	0.0244
Family history (recurrent PE)	38	9	1.92%	5.06%	0.0082
Family history (PTB)	292	37	14.72%	20.79%	0.0319
Family history (CH)	749	82	37.75%	46.07%	0.0295
Family history (IHD)	278	40	14.01%	22.47%	0.0025
Eat fruit >3 times/day (1m pre-preg)	512	27	25.81%	15.17%	0.0020
Eat vege>3 times/day (1m pre-preg)	222	9	11.19%	5.06%	0.0137
Any alcohol in 1st trim	972	63	48.99%	35.39%	0.0006
Abnormal umbrilical doppler	190	27	9.95%	15.52%	0.0227

** GH=gestational hypertension; PE=preeclampsia; PTB=preterm birth; CH=chronic hypertension; IHD=ischaemic heart disease*

Table 3.4.2: Univariate analysis of common factors associated with SPTB

Variable	N (Control)	N (SPTB)	Mean±SD / % (Control)	Mean±SD / % (SPTB)	P
Mother's birthweight (g)	1872	141	3333 ± 530	3229 ± 619	0.0263
Gravidity	1984	156	1.3 ± 0.62	1.5 ± 0.76	0.0015
Number of vaginal bleeds before 15 weeks' gestation	1984	156	0.3 ± 0.61	0.4 ± 0.82	0.0045
Folate supplementation in 1st trimester (μg/day)	1984	156	730 ± 551	632 ± 455	0.0211
Any cigarettes (3 mths prior preg)	181	23	9.12%	14.74%	0.0134
Any cigarettes (1st trimester)	109	15	5.49%	9.62%	0.0205
Marijuana (number of times 3 months pre-preg)	1982	156	15.4 ± 179.05	87.9 ± 459.71	0.0013
Marijuana (number of times in 1st trimester)	1983	156	9.0 ± 131.45	41.4 ± 239.64	0.0342
MAP at15 wks (mmHg)	1984	156	77.8 ± 7.53	79.4 ± 8.30	0.0079
Maternal height (cm)	1984	156	165.5 ± 6.55	164.0 ± 6.82	0.0052

Tests of association of SNPs showed that SNPs in genes such as IL10, IGF1, IGF2, MTHFR and PGF (Powers et al., 1999; Kaiser et al., 2000; Driul et al., 2004; Mutze et

al., 2008; Vucic et al., 2009) are consistent with the literature as candidate genes for Preeclampsia (Table 3.4.3). Similarly, SNPs in IL1β and IL6 (Genc et al., 2002; Simhan et al., 2003; Engel et al., 2005; Goldenberg et al., 2008) have also been identified to be associated with Preterm Birth (Table 3.4.4). One point to note is that since the aim of this exploratory analysis is to obtain a draft list of potential SNPs, a relaxed threshold ($P < 0.1$) have been used, as there may be interacting SNPs that are not independently significant.

Table 3.4.3: SNPs associated with PE in univariate analysis at 5% significance level

Gene	RS no.	Patient	P	Gene	RS no.	Patient	P
AGT	rs699	Part	0.0515	CYP11A1	rs8039957	Part	0.0212
IL10	rs1800896	Mum	0.0787	COX2	rs20417	Mum	0.0893
PTEN	rs1234220	Mum	0.0000	IGF1	rs7965399	Mum	0.0741
TP53	rs1042522	Part	0.0473	IGF1R	rs11247361	Mum	0.0679
MTHFR	rs1801131	Mum	0.0381	INSR	rs2059806	Mum	0.0135
BCL2	rs2279115	Part	0.0044	MMP2	rs2285053	Mum	0.0350
GSTP1	rs1695	Part	0.0850	COX2	rs5275	Neo	0.0195
TGFB	rs1800469	Mum	0.0800	COX2	rs20417	Neo	0.0335
PGF	rs1042886	Part	0.0933	IGF1R	rs2229765	Neo	0.0672
IGF2R	rs2274849	Mum	0.0938	IGF2	rs680	Neo	0.0435
IL1RN	rs454078	Mum	0.0868	IGF2	rs3741204	Neo	0.0416
PAI1	rs1799768	Mum	0.0798	IGF2AS	rs1004446	Neo	0.0845
PAI1	rs1799889	Mum	0.0402	INSR	rs2059806	Neo	0.0066
IL1B	rs16944	Neo	0.0169	LIN28	rs12747426	Neo	0.0641
IL1RN	rs454078	Neo	0.0551	MMP2	rs243865	Part	0.0540

** Mum=maternal; Part=paternal; Neo=neonatal*

Table 3.4.4: SNPs associated with SPTB in univariate analysis at 5% significance level

Gene	RS no.	Patient	P	Gene	RS no.	Patient	P
AGT	rs4762	Mum	0.0382	ENG	rs10987759	Part	0.0241
AGTR1	rs5186	Part	0.0547	IGF2R	rs2274849	Mum	0.0066
FTO	rs9939609	Mum	0.0845	IL1A	rs17561	Neo	0.0900
PTEN	rs1234220	Mum	0.0410	uPA	rs2227564	Neo	0.0826
ADRB2	rs1042714	Neo	0.0635	IL1B	rs16944	Part	0.0830
ADD1	rs4961	Part	0.0490	IGF1	rs12579108	Mum	0.0821
IL6	rs1800795	Part	0.0005	IGF1R	rs2229765	Mum	0.0587
BCL2	rs2279115	Mum	0.0683	MMP2	rs243865	Mum	0.0100
ANXA5	rs17551751	Neo	0.0217	COL4A2	rs41315048	Neo	0.0681
F5	rs6025	Neo	0.0685	COX2	rs20417	Neo	0.0866
MTR	rs1805087	Neo	0.0866	TIMP3	rs5749511	Neo	0.0874
NAT2	rs1208	Neo	0.0419	COX2	rs5275	Part	0.0413
TCN2	rs1801198	Neo	0.0721	COX2	rs20417	Part	0.0485
MTHFD1	rs2236225	Part	0.0532	INSR	rs2059806	Part	0.0247
MTHFR	rs1801131	Part	0.0518				

** Mum=maternal; Part=paternal; Neo=neonatal*

3.4.2. Subgroup Effects

As models will be developed for PE and SPTB separately, CMH tests were performed on all variables between Adelaide and Auckland for PE and SPTB separately. The effects for 63 variables are different for PE between Adelaide and Auckland, and 129 for SPTB, which means that less than 20% of variables are independent between Adelaide and Auckland. A full list of significant differences under CMH for PE and SPTB are provided in Appendix.

Interestingly, common risk factors including age, BMI and SEI do not appear to have an effect difference for PE and SPTB between the two locations, despite their diverse distributions. Hence, this study will analyze the combined Adelaide and Auckland cohort, and more detailed analysis will be performed later whenever a subgroup effect is present in models developed.

References

Agresti, A. (2002). Categorical Data Analysis. New Jersey, John Wiley & Sons, Inc.

Cui, L., Hung, J. H. M., et al. (2002). "Issues related to Subgroup analysis in clinical trials." Journal of Biopharmaceutical Statistics **12**(3): 347-358.

Davis, P., Jenkin, G., et al. (2003). New Zealand Socio-economic Index 1996. Wellington, New Zealand, Statistics New Zealand.

Driul, L., Damante, G., et al. (2004). "Screening for pre-eclampsia in a low-risk population at 24 weeks: uterine artery Doppler flow velocimetry and genetic variants of factor V, prothrombin and methylenetetrahydrofolate reductase." Minerva Ginecol **56**(5): 385-390.

Engel, S. A., Erichsen, H. C., et al. (2005). "Risk of spontaneous preterm birth is associated with common proinflammatory cytokine polymorphisms." Epidemiology **16**(4): 469-477.

Fletcher, J. (2007). "Subgroup analyses: how to avoid being misled." Bmj **335**(7610): 96-97.

Genc, M. R., Gerber, S., et al. (2002). "Polymorphism in the interleukin-1 gene complex and spontaneous preterm delivery." Am J Obstet Gynecol **187**(1): 157-163.

Goldenberg, R. L., Culhane, J. F., et al. (2008). "Epidemiology and causes of preterm birth." Lancet **371**(9606): 75-84.

Guillemin, F. (2007). "Primer: the fallacy of subgroup analysis." Nat Clin Pract Rheumatol **3**(7): 407-413.

Hollander, M. and Douglas, A. (1973). Nonparametric Statistical Methods. New York, John Wiley & Sons.

Kaiser, T., Brennecke, S. P., et al. (2000). "Methylenetetrahydrofolate reductase polymorphisms are not a risk factor for pre-eclampsia/eclampsia in Australian women." Gynecol Obstet Invest **50**(2): 100-102.

Klebanoff, M. A. (2007). "Subgroup analysis in obstetrics clinical trials." Am J Obstet Gynecol **197**(2): 119-122.

Liu, I. M. (2005). Breslow–Day Test. Encyclopedia of Biostatistics, John Wiley & Sons, Ltd.

Mantel, N. and Haenszel, W. (1959). "Statistical aspects of the analysis of data from retrospective studies of disease." J Natl Cancer Inst **22**(4): 719-748.

Moyé, L. A. (2012). "Rudiments of Subgroup Analyses." Progress in Cardiovascular Diseases **54**(4): 338-342.

Mutze, S., Rudnik-Schoneborn, S., et al. (2008). "Genes and the preeclampsia syndrome." J Perinat Med **36**(1): 38-58.

Newcombe, R. G. (1998). "Two-sided confidence intervals for the single proportion: comparison of seven methods." Stat Med **17**(8): 857-872.

Pandis, N. (2013). "Multiplicity 1: Subgroup analyses." American Journal of Orthodontics and Dentofacial Orthopedics **143**(3): 439-441.

Powers, R. W., Minich, L. A., et al. (1999). "Methylenetetrahydrofolate reductase polymorphism, folate, and susceptibility to preeclampsia." J Soc Gynecol Investig **6**(2): 74-79.

SCOPE Consortium (2004). "The SCOPE Pregnancy Research Study." Retrieved 5 March 2010, from http://www.scopestudy.net/.

SCOPE Consortium (2007). SCOPE Midwives SOP, Confidential SCOPE Consortium.

Simhan, H. N., Krohn, M. A., et al. (2003). "Interleukin-6 promoter -174 polymorphism and spontaneous preterm birth." Am J Obstet Gynecol **189**(4): 915-918.

Sun, X., Briel, M., et al. (2012). "Credibility of claims of subgroup effects in randomised controlled trials: systematic review." Bmj **344**: e1553.

Vucic, N., Frleta, M., et al. (2009). "Thrombophilia, preeclampsia and other pregnancy complications." Acta Med Croatica **63**(4): 297-305.

Wang, R., Lagakos, S. W., et al. (2007). "Statistics in Medicine — Reporting of Subgroup Analyses in Clinical Trials." New England Journal of Medicine **357**(21): 2189-2194.

Wilson, E. B. (1927). "Probable Inference, the Law of Succession, and Statistical Inference." Journal of the American Statistical Association **22**(158): 209-212.

Woolf, B. (1955). "On estimating the relation between blood group and disease." Ann Hum Genet **19**(4): 251-253.

Yusuf, S., Wittes, J., et al. (1991). "Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials." Jama **266**(1): 93-98.

Chapter 4: Methodology

4.1. Overview

Pregnancy complications may lead to serious maternal or fetal morbidity and mortality. Models that predict their occurrence need to be trained with caution to avoid false or misleading results, and thorough validation is crucial to ensure their reliability.

To develop reliable screening tools, various statistical and data mining methods, including two commonly used approaches, classification and clustering, are used in this study to not only explore the use of these methods, but also taking into account the advantages and limitations of each of these methods and cross-validate the results obtained.

This chapter will discuss the Classification and Clustering approaches used to develop individual models to be integrated into a tiered model. The individual models will be developed based on 2 time-points in pregnancy: 15 weeks and 20 weeks of gestation, with specific model requirements for each tier. The best models for each pregnancy complication are selected based on penalty measures for model over-fitting, including Akaike Information Criterion and Elastic-net penalization, as well as validation, a measure for goodness of classification based on sensitivity, specificity, and Receiver-Operating Characteristic Curve (ROC). These will be further discussed in Section 4.4 and 4.5.

The best models are then integrated into a tiered prediction system (discussed in Chapter 5), which will monitor and update the predicted risk for individuals throughout pregnancy when new predictors are available or when changes occur.

An overview of the methodology is shown in Fig. 4.1.1. Initial models are trained on the SCOPE database using classification, clustering and visualization methods. Best models are then selected based on penalty function and accuracy measures for each tier. These are used to build a tiered model system with different sets of predictors for pregnancy complications from before and during early pregnancy, and provide a predicted risk at each stage, and hence, allow tailored antenatal care for women not at risk and for those at risk.

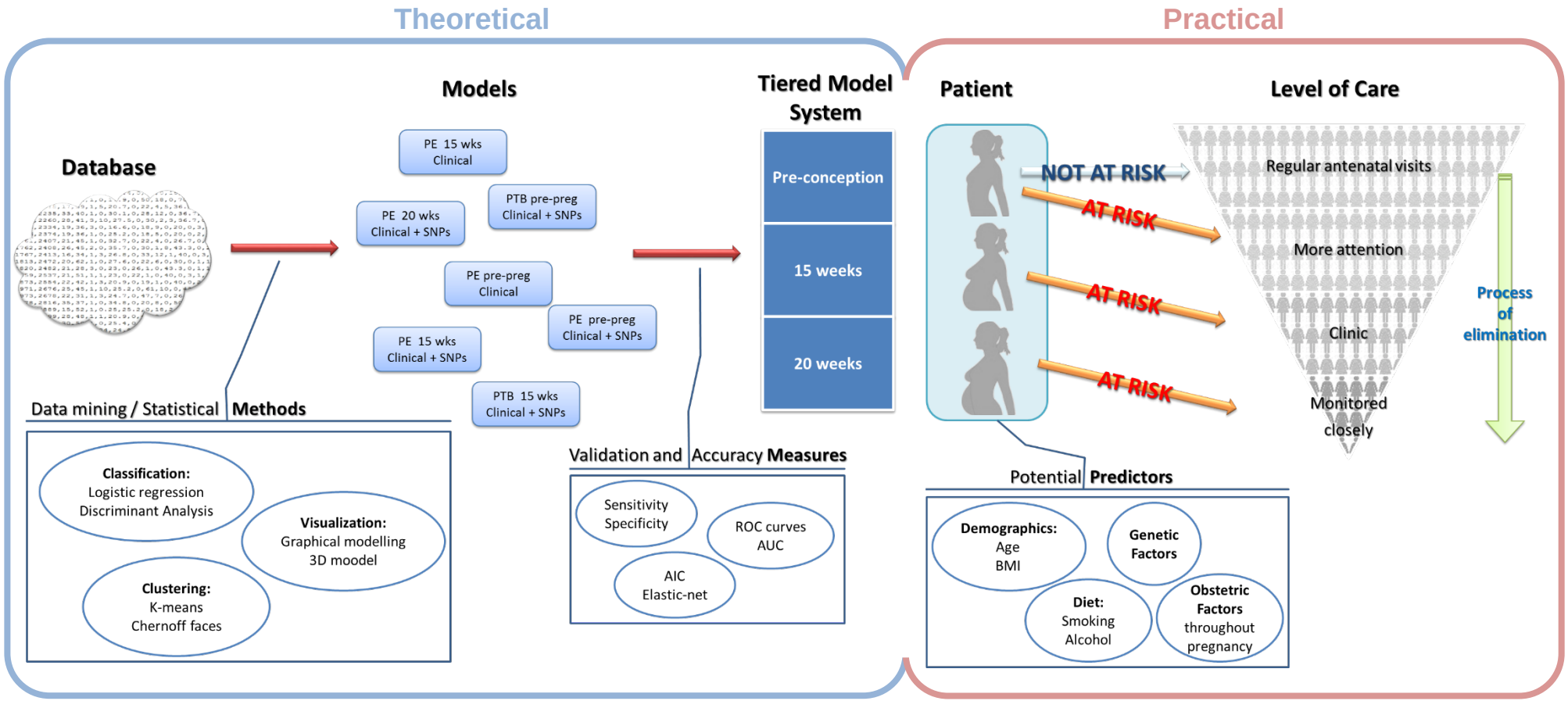

Fig. 4.1.1: Methodology overview

Separate models were trained from database using data mining methods, in which the best models based on accuracy measures were chosen to integrate into the tiered model. The predictors in the final models would then be used to classify women into three levels of risk.

4.2. Classification

Classification is a statistical approach that aims to separate distinct sets of observations (or cases) and allocate new observations to previously defined groups (Wichern and Johnson, 2007). Hence, classification methods require a training dataset, from which a classification rule can be developed. This is similar to 'learning' from known samples, and is often referred to as 'supervised learning' in data mining.

4.2.1. Logistic Regression

Logistic regression is a Classification method in statistical analysis. It is often used to develop predictive models with categorical outcome variables. Although logistic regression has been developed for multinomial or polytomous outcomes (Tabachnick and Fidell, 1996), i.e. an outcome of more than two categories, most disease predictions are interested in predicting binary outcomes, e.g. 'disease' or 'non-disease', this section will therefore focus on Binary Logistic Regression.

4.2.1.1. Binary Logistic Regression

In binary logistic regression, the response variable, denoted Y, can take only two values: 0 or 1. It is customary to represent a 'success' or 'positive' as 1. For example, 1 could denote 'diseased' or 'have symptom'. Logistic regression classifies variables into the two groups (i.e. group 1 and 0) by modeling the posterior probability of class 1 membership via a linear function of the explanatory variables (Hastie et al., 2009).

Instead of directly modeling the posterior probability of class 1 membership, $p = E[Y|X]$, the *logit transformation* is used (Wichern and Johnson, 2007). This ensures the predicted response from the linear regression is bounded between 0 and 1. Let $X_1, X_2, \dots, X_k$denotes the k explanatory variables. Then the logit transformation of p,

which is defined as the logarithm of the odds ratio, is modeled by the linear function:

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

It is convenient to express the linear regression function in matrix form $\beta^T X$, where $\beta = [\beta_0, \beta_1, \dots, \beta_k]^T$ and $X = [1, X_1, X_2, \dots, X_k]^T$. The probability of 'diseased' can also be written as $p = \frac{1}{1+e^{-\beta^T X}}$ in matrix form.

An estimate of the coefficients β can be obtained via the maximum likelihood method. Conditional on the n observations of X (denoted $\boldsymbol{x_i} = [1, x_{i1}, x_{i2,} \dots, x_{ik}]$, $i = 1, 2, \dots, n$), the response variable is assumed to be Bernoulli-distributed with success probability p. Hence, the log likelihood can be specified

$$\log L(\beta) = \sum_{i=1}^{n} \log\left[p_i^{y_i}(1-p_i)^{1-y_i}\right]$$

$$= \sum_{i=1}^{n} [y_i \log(p_i) + (1-y_i)\log(1-p_i)]$$

$$= \sum_{i=1}^{n} \left[y_i \log\left(\frac{p_i}{1-p_i}\right) + \log(1-p_i)\right]$$

$$= \sum_{i=1}^{n} \left[y_i \beta^T \boldsymbol{x_i} - \log\left(1 + e^{\beta^T x_i}\right)\right]$$

The Maximum Likelihood Estimator (MLE) of β is obtained by maximizing the log likelihood. Equating the derivative of $\log L(\beta)$ to zero results in $k+1$ nonlinear equations in β, with no closed form.

$$\frac{\partial}{\partial \beta} \log L(\beta) = \sum_{i=1}^{n} [(y_i - p_i)\boldsymbol{x_i}] = 0$$

The solution can be deduced numerically using common iterative techniques such as the Newton-Raphson method. This requires the second derivative of the log likelihood

$$\frac{\partial}{\partial \beta \partial \beta^T} \log L(\beta) = \sum_{i=1}^{n} [\boldsymbol{x_i}^T \boldsymbol{x_i} p_i (1-p_i)]$$

The algorithm begins with an initial estimate of β, denoted $\beta^{(0)}$. It then generates new

estimates by subtracting the term $-\left[\frac{\partial \log L(\beta)}{\partial \beta \partial \beta^T}\right]^{-1}\left[\frac{\partial log L(\beta)}{\partial \beta}\right]$ iteratively. Hence, at the $(t+1)^{th}$ iteration, the estimate is given by

$$\beta^{(t+1)} = \beta^{(t)} - \frac{\sum_{i=1}^{n}[(y_i - p_i)\boldsymbol{x_i}]}{\sum_{i=1}^{n}[\boldsymbol{x_i^T x_i} p_i(1-p_i)]}$$

Note that the odds of a variable X is simply given by $\exp(\hat{\beta})$(Hosmer and Lemeshow, 2005). Consider a simple model with a single variable X_1, and the odds ratio is given by

$$OR = \frac{P(Y=1|X_1=1)}{P(Y=0|X_1=1)} \cdot \frac{P(Y=0|X_1=0)}{P(Y=1|X_1=0)}$$

Since $\log\left[\frac{P(Y=1|X_1=1)}{P(Y=0|X_1=1)}\right] = \text{logit}(p)$, the odds ratio (OR) can be written as

$$OR = \frac{\left(\frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\right)/\left(\frac{1}{1+e^{\beta_0+\beta_1}}\right)}{\left(\frac{e^{\beta_0}}{1+e^{\beta_0}}\right)/\left(\frac{1}{1+e^{\beta_0}}\right)}$$

$$= \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}}$$

$$= e^{(\beta_0+\beta_1)} \cdot e^{-\beta_0}$$

$$= e^{\beta_1}$$

4.2.1.2. Advantages and Disadvantages

The major advantage of Logistic regression is that it can develop prediction rules for dichotomous (or binary) outcomes and that it is also flexible enough to analyze a dataset that contains a mixture of continuous and categorical variables. In contrast, the prediction using linear regression on binary outcomes is not meaningful, while it violates the homoscedacity assumption (Dufty, 2007). Indeed, logistic regression develops prediction rules based on the logit (i.e. log of the odds) of the probability of 'success'p, as discussed previously in Section 4.2.1.1. By logit transforming p, a more sensible interpretation of the model coefficients is provided. In fact, they can be

interpreted as the natural logarithm of the odds ratio for a certain predictor in the model.

Another advantage of Logistic regression is that it does not have much of the assumptions that exist in methods like linear regression. It does not assume that independent variables (or predictors) are normally distributed or have an equal variance between the different groups in the data. In fact, it makes no assumption on the distribution of the variables (Tansey et al., 1996). Hence, it is more robust. Also, it does not assume a linear relationship between the predictors and outcome, and could handle non-linear effects.

However, one of the drawbacks for Logistic regression is that a much larger dataset is needed to achieve stable and meaningful results. It is estimated that at least 50 data points per predictor are necessary to achieve stable results (Dufty, 2007).

Multicollinearity is also an issue with logistic regression. This occurs when two or more predictors in the model are approximately a linear combination of each other, and these interrelationships can be shown by measures such as the Variance Inflation Factor (VIF). Multicollinearity may result in a lack of statistical significance of predictors for models which are shown to be strongly significant overall, and also highly unreliable estimate model coefficients and very large standard errors for coefficients may be produced.

4.2.1.3. Applications

As shown in Section 2.1.2.3, logistic regression is the most common statistical method used to predict preeclampsia, in which its convenient estimation of odds have provided meaningful results. A study by Austin et al. (Austin et al., 2010) has shown that logistic regression outperforms regression trees for predicting in-hospital mortality in patients hospitalized with heart failure.

In fact, logistic regression has been used in an increasingly wide range of areas, from risk prediction to scoring credit applications in marketing and business (Agresti, 2002). Recently, logistic regression has been applied in conjunction with geomorphological mapping to analyze the distribution of active rock glaciers in relation to altitude, aspect, slope, lithology and solar radiation in the Andes of San Juan (Angillieri, 2010). An interesting study by Ayuso et al. (Ayuso et al., 2010) has shown that traffic violations are associated with a higher probability of serious or fatal accidents based on logistic regression. Moreover, a study in ecological economics has used logistic regression to investigate the importance of economic factors, attitudes/values, knowledge/perceptions of resource scarcity, and social capital, on fostering conservation behaviour (Brooks, 2010).

4.2.1.4. Illustrative Example - Prediction Model for PE

This section gives an example of practical application and interpretation of Logistic regression for prediction of preeclampsia. The model was trained from a total of 906 patients, in which 99 of them developed preeclampsia cases. A combination of clinical and SNP predictors at 15 weeks of gestation were included in the model, including BMI, fruit and vegetable consumption, family history and several related genes.

A summary of model coefficients with the P-value and odds of each predictor is shown in Table 4.2.1. As expected, common risk factors such as BMI, MAP and family history of PE or GH increases the risk of PE, while eating fruits more than 3 times per day reduces the risk of PE.

Table 4.2.1: Logistic regression model for PE at 15 weeks of gestation

	Predictors	Odds	P-value
Current condition	BMI	1.07 (1.03,1.12)	***0.001***
	MAP (15 wks)	1.10 (1.06,1.14)	***0.000***
	Gastroenteritis (15 wks)	3.72 (1.51,9.17)	***0.004***
Family history	Mother's birthweight	0.9995 (0.9991,0.9999)	***0.011***
	Maternal father has IHD	3.55 (1.85,6.81)	***0.000***
	PE or GH	2.50 (1.42,4.39)	***0.001***
Dietary / lifestyle	Folate supplementation	0.9995 (0.9988,1.000)	0.123
	Fruit (>3x/day)	0.31 (0.14,0.70)	***0.004***
	Sex/month (pre-preg)	1.03 (1.01,1.04)	***0.001***
SNP	MTHFR (mat) [SNP]	1.63 (0.99,2.69)	0.053
	TGFb (mat) [SNP]	1.59 (0.97,2.62)	***0.066***
	PGF (mat) [SNP]	2.86 (1.35,6.06)	***0.006***
	IL10 (mat) [SNP]	1.86 (1.10,3.14)	***0.021***
	BCL2 (pat) [SNP]	1.60 (0.89,2.90)	0.119
	IGF2AS (pat) [SNP]	1.57 (0.96,2.58)	0.074
	NOS2A (pat) [SNP]	1.85 (1.08,3.17)	***0.024***
	PGF (pat) [SNP]	1.84 (1.02,3.31)	***0.043***

*IHD=Ischemic Heart Disease; PE=Preeclampsia; GH=Gestational Hypertension

To obtain a predicted risk for an individual patient, information for each predictor in the model must first be recorded, and then substituted into the model formula:

$$\begin{aligned}\text{logit}(\text{PE}) = {} & -12.71 + 0.07[\text{BMI}] + 0.093[\text{MAP}] + 1.31[\text{Gastroenteritis}] \\ & + 0.027[\text{Sex/month pre} - \text{preg}] - 0.0006[\text{Mother's birthweight}] \\ & + 1.268[\text{Father(IHD)}] + 0.914[\text{FH(PE/GH)}] \\ & - 1.167[\text{Fruit}(\geq 3\text{x/day})] - 0.0006[\text{Folate}] \\ & + 0.49[\text{MTHFR}_{\text{mat}}(\text{SNP})] + 0.62[\text{IL10}_{\text{mat}}(\text{SNP})] \\ & + 0.47[\text{TGFb}_{\text{mat}}(\text{SNP})] + 1.05\left[\text{PGF}_{\text{mat(SNP)}}\right] + 0.47\left[\text{BCL2}_{\text{pat(SNP)}}\right] \\ & + 0.62\left[\text{NOS2A}_{\text{pat}}(\text{SNP})\right] + 0.608\left[\text{PGF}_{\text{pat}}(\text{SNP})\right] \\ & + 0.45\left[\text{IGF2AS}_{\text{pat}}(\text{SNP})\right]\end{aligned}$$

For instance, an individual patient with information recorded shown in Table 4.1.2, the values recorded will be placed into the model formula to obtain her $\text{logit}(\text{PE})$, which can be transformed into the form of probability by:

$$\hat{p} = \frac{\exp\big(\text{logit}(\text{PE})\big)}{1 + \exp\big(\text{logit}(\text{PE})\big)} \approx 0.80$$

For categorical variables, dummy variables are created for each category, in which the presence of a category is indicated by 1 and absence is indicted by 0 (i.e.0/1 are substituted into the model formula to indicate the absence/presence of a categorical variable).

Table 4.2.2: Example clinical and genotype record

Clinical		**Genotype**	
BMI	35	MTHFR (mat)	Yes
MAP	79	IL10 (mat)	Yes
Gastroenteritis	No	TGFb (mat)	Yes
Sex/month pre-preg	14	PGF (mat)	Yes
Mother's birthweight	3287g	BCL2 (pat)	Yes
Father has IHD	Yes	NOS2A (pat)	No
Family history of PE/GH	Yes	PGF (pat)	Yes
Folate supplement	None	IGF2AS (pat)	No

*IHD=Ischemic Heart Disease; PE=Preeclampsia; GH=Gestational Hypertension

Hence, the predicted probability of PE for this individual patient, based on predictors available at 15 weeks of gestation, is 80%.

Interestingly, this model can also demonstrate the effect of modifiable factors. For the same patient described above, i.e. keeping the family history and genotypes constant, if her BMI reduces to 26, consume more fruits per day and take at least 526 μg/day of folate supplement, her predicted probability will be reduced to 33%.

4.2.2. Discriminant Analysis

Discriminant Analysis is a statistical classification method which aims to obtain rules that describe the separation between groups of observations (Hubert and Van Driessen, 2004). As a classification technique, the classification rule of discriminant analysis is based on predefined groups/clusters, e.g. having a disease vs. not having a disease.

Given the predefined group, discriminant analysis first aims at finding the best way to exhibit difference between the groups. This involves finding linear combinations of independent variables (or predictors) (Fisher, 1936; McLachlan, 2004). Then, it aims to find a classification rule for allocating new data/observations into an existing cluster.

4.2.2.1. Classification Rule

Let there be K clusters. Each cluster is modelled by a multivariate normal density and can be characterised by its mean vector μ_k and covariance matrix Σ. Their densities are given by

$$f_k(x) = (2\pi)^{-\frac{p}{2}}|\Sigma|^{-\frac{1}{2}(x-\mu_k)^T\Sigma^{-1}(x-\mu_k)}, k = 1,2,\dots,K$$

where p is the dimension of an observation. In linear discriminant analysis (LDA), $f_k(x)$ follows multivariate normal distribution with a common covariance matrix in all clusters, i.e. $\Sigma_k = \Sigma\ \forall k$. This results in a linear decision boundary (Hastie et al., 2009). A generalized version of LDA, known as quadratic discriminant analysis (QDA), is used in this study. QDA does not assume common covariance across all clusters, and is more flexible by allowing non-linear discriminant regions.

The discriminant boundary between two clusters i and j occurs where a new observation has an equal chance of falling into one of the clusters, which is determined

by the discriminant function $\delta_k(x) = \log(\pi_k f_k(x))$, where π_k is the probability of falling into cluster k.

$$\delta_i(x) = \delta_j(x)$$

$$\log(\pi_i f_i(x)) = \log\left(\pi_j f_j(x)\right)$$

$$log(\pi_i) - \frac{1}{2} log|\hat{\Sigma}_i| - \frac{1}{2}(x - \hat{\mu}_i)^T \hat{\Sigma}_i^{-1}(x - \hat{\mu}_i)$$

$$= log(\pi_j) - \frac{1}{2} log|\hat{\Sigma}_j| - \frac{1}{2}(x - \hat{\mu}_j)^T \hat{\Sigma}_j^{-1}(x - \hat{\mu}_j)$$

$$\log\left(\frac{\pi_i}{\pi_j}\right) = -\frac{1}{2}\log\left(\frac{|\hat{\Sigma}_i|}{|\hat{\Sigma}_j|}\right) - \frac{1}{2}\left[(x - \hat{\mu}_i)^T \hat{\Sigma}_i^{-1}(x - \hat{\mu}_i) - (x - \hat{\mu}_j)^T \hat{\Sigma}_j^{-1}(x - \hat{\mu}_j)\right]$$

$$2\log\left(\frac{\pi_i}{\pi_j}\right) - \log\left(\frac{|\hat{\Sigma}_i|}{|\hat{\Sigma}_j|}\right) = (x - \hat{\mu}_i)^T \hat{\Sigma}_i^{-1}(x - \hat{\mu}_i) - (x - \hat{\mu}_j)^T \hat{\Sigma}_j^{-1}(x - \hat{\mu}_j)$$

$$2\log\left(\frac{\pi_i}{\pi_j}\right) - \log\left(\frac{|\hat{\Sigma}_i|}{|\hat{\Sigma}_j|}\right)$$
$$= x^T \hat{\Sigma}_i^{-1} x - x^T \hat{\Sigma}_i^{-1} \hat{\mu}_i - \hat{\mu}_i^T \hat{\Sigma}_i^{-1} x + \hat{\mu}_i^T \hat{\Sigma}_i^{-1} \hat{\mu}_i - x^T \hat{\Sigma}_j^{-1} x + x^T \hat{\Sigma}_j^{-1} \hat{\mu}_j$$
$$- \hat{\mu}_j^T \hat{\Sigma}_j^{-1} x + \hat{\mu}_j^T \hat{\Sigma}_j^{-1} \hat{\mu}_j$$

$$2\log\left(\frac{\pi_i}{\pi_j}\right) - \log\left(\frac{|\hat{\Sigma}_i|}{|\hat{\Sigma}_j|}\right)$$
$$= x^T(\hat{\Sigma}_i^{-1} - \hat{\Sigma}_j^{-1})x - 2\hat{\mu}_i^T \hat{\Sigma}_i^{-1} x + \hat{\mu}_i^T \hat{\Sigma}_i^{-1} \hat{\mu}_i + 2\hat{\mu}_j^T \hat{\Sigma}_j^{-1} x - \hat{\mu}_j^T \hat{\Sigma}_j^{-1} \hat{\mu}_j$$

$$2\log\left(\frac{\pi_i}{\pi_j}\right) - \log\left(\frac{|\hat{\Sigma}_i|}{|\hat{\Sigma}_j|}\right)$$
$$= x^T(\hat{\Sigma}_i^{-1} - \hat{\Sigma}_j^{-1})x - 2(\hat{\mu}_i^T \hat{\Sigma}_i^{-1} - \hat{\mu}_j^T \hat{\Sigma}_j^{-1})x + \hat{\mu}_i^T \hat{\Sigma}_i^{-1} \hat{\mu}_i - \hat{\mu}_j^T \hat{\Sigma}_j^{-1} \hat{\mu}_j$$

Cluster membership of an observation is determined by the largest posterior probability of membership, i.e. a new data x is classified into class j when $\pi_j f_j(x) > \pi_i f_i(x)$. In other words, the clusters are selected by

$$k^*(x) = \arg\max_k \delta_k(x)$$

where $\delta_k(x) = \log(\pi_k) - \frac{1}{2} log|\Sigma_k| - \frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1}(x - \mu_k)$. In practice, the

estimate of π_k can be calculated from the training data, $\hat{\pi}_k = \frac{N_k}{N}$, where N_k is the number of observations in the training data classified as cluster k, and N is the total number of observations in the dataset. The estimate of π_k is defined by $\hat{\pi}_k = 1 - \hat{\pi}_i = \frac{N-N_k}{N}$.

4.2.2.2. Advantages and Disadvantages

Similar to logistic regression, discriminant analysis also has the ability to analyze categorical data. It is also able to classify data based on multiple parameters and synthesize a set of predictors using a discriminant function (Jaba et al., 2007). Moreover, it allows classification of new data/observations into an existing cluster based on the classification rules (Hubert and Van Driessen, 2004).

One of the disadvantages of discriminant analysis is that its estimates are highly influenced by outlying observations, and hence, it might be inappropriate in contaminated datasets (Hubert and Van Driessen, 2004). In addition, linear discriminant analysis has more assumptions than logistic regression, e.g. distributions are assumed to be normal with equal covariance (i.e. homoelasticity) (Marzban et al., 1997). Thus, it is generally felt that logistic regression is more robust. Nevertheless, as Hastie et al. (Hastie et al., 2009) stated that models obtained from logistic regression analysis and linear discriminant analysis give very similar results, even when an assumption is violated.

4.2.2.3. Applications

As a classification method, discriminant analysis has been applied in multiple disciplines. A recent study by Okamoto and Harasawa (Okamoto and Harasawa, 2010)

has used discriminant analysis to obtain a discriminant function as a predictive model for discrimination between elderly persons who are at higher risk of depression and normal subjects in Japan. Another interesting study by du Jardin et al. (du Jardin et al., 2009) has investigated discriminant analysis as a classification technique for sex determination for skeletal remains.

Moreover, discriminant analysis is known to be one of the most widely used statistical procedures in empirical studies of discrete economic phenomena (Lo, 1986). The majority of these applications are bankruptcy predictions (Jo et al., 1997; Gu, 2002; Cho et al., 2009;Sueyoshi and Goto, 2009).

On the other hand, discriminant analysis has also been used in face recognition and image processing (Lu et al., 2005; Dai et al., 2007; Pnevmatikakis and Polymenakos, 2009; Song et al., 2010). An interesting study by Lin et al. (Lin et al., 2010) proposed an Unsupervised Linear Discriminant Analysis (ULDA) which aims to classify and segment the three major brain tissues, i.e. gray matter, white matter and cerebro-spinal fluid, from a multi-spectral MR image of the human brain.

4.2.2.4. Illustrative Example - Discriminant Functions for PE

This section provides an example of practical application of Discriminant Analysis for prediction of Preeclampsia. Discriminant Analysis was performed on 1267 patients, with 118 PE cases. A total of 12 clinical predictors, available at 15 weeks of gestation, were included in the discriminant functions, which includes common predictors such as SEI, maternal BMI and vegetable consumption.

For each of the two groups, ‘PE’ and ‘Uncomplicated pregnancy’, the means for each predictor is obtained as group means, along with two discriminant functions $\delta_k(x) =$

$\log(\pi_k) - \frac{1}{2}log|\Sigma_k| - \frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k)$, where $k = 0, 1$ represents the 'Uncomplicated pregnancy' and 'PE' group respectively, and μ_k are the predictor means, Σ_k are the covariances, π_i is the prevalence of PE, and x_k is the recorded information for a new patient. The table of group means for each predictor is shown in Table 4.2.3.

Table 4.2.3: Group means for PE and Uncomplicated pregnancy

Predictors	Uncomplicated pregnancy	PE
Yrs of schooling	12.3	12.1
SEI	42.1	36.8
Birthweight (mother)	3334	3177
Folate supplement (μg)	555.5	515.7
sBP (15 wks)	106.2	113.1
BMI	24.9	28.5
Waking at night	2.6	2.8
Stairs per day	1.7	1.6
Fruit per day (15 wks)	1.4	1.5
Vegetables per day (15wks)	0.3	0.4
Alcohol (15 wks)	1.3	1.0
Alcohol (pre-preg)	1.5	1.1

For a new patient, information on the 12 clinical predictors must be first recorded, and her scores for falling into each group (δ_k) is then obtained by substituting the predictors, group means and prevalence of each group. Her predicted group will then be determined by the maximum score, e.g. if her score for 'Uncomplicated pregnancy' is lower than that of 'PE', she will be classified as 'PE'.

A partition plot indicating the two cluster regions defined by discriminant functions is shown in Fig. 4.2.1, with the 'Uncomplicated pregnancy' region shaded in white and the 'PE' region shaded in yellow. In a simple 2-variable case, QDA can be easily visualized and applied using the partition plot. Information of a new patient can be plotted onto the partition plot, e.g. her BMI and systolic blood pressure obtained at 15

weeks' gestation, and her predicted group will be the region her plotted point lies on the plot. For multiple variables, a more complex calculation is required using the discriminant function (δ_k).

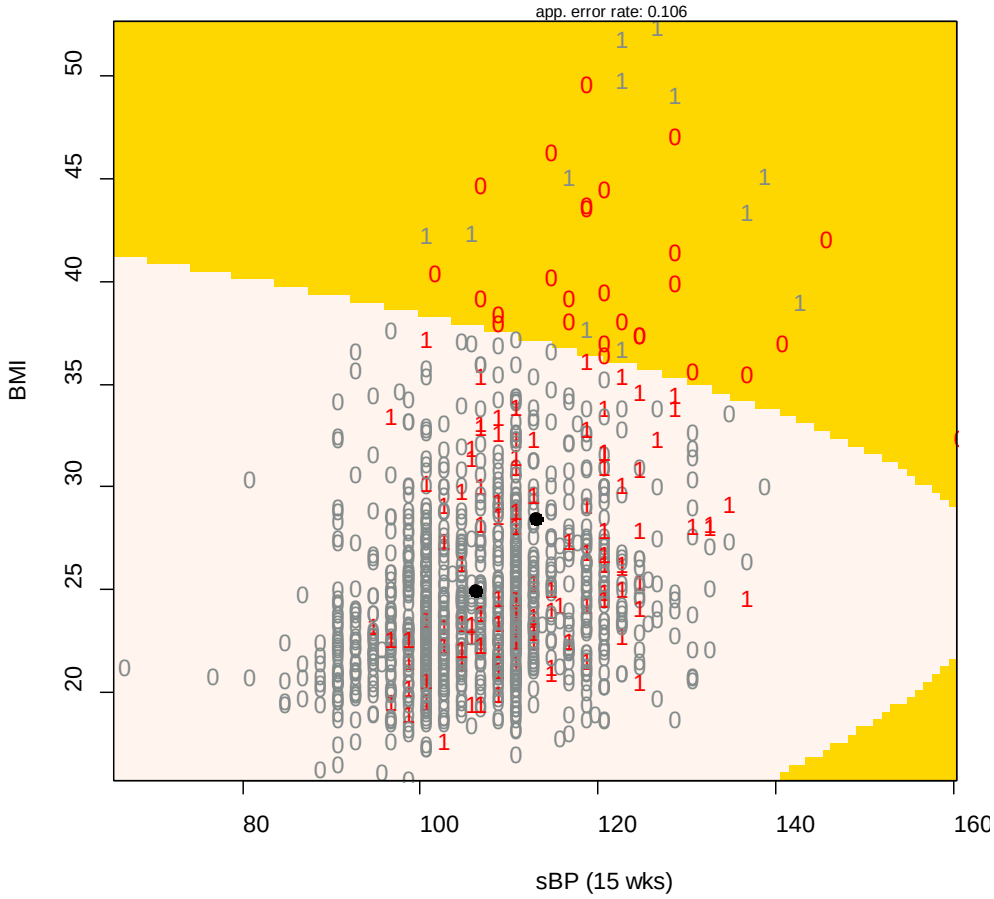

Fig. 4.2.2: Partition plot (white=Uncomplicated pregnancy, yellow=PE)

For instance, if the scores for a new patient is $\pi_0 f_0(x) = 0.458$ and $\pi_1 f_1(x) = 0.541$, which means that her score for the 'Uncomplicated pregnancy' group is 0.458, while her score for 'PE' is 0.541. Since $\pi_0 f_o(x) < \pi_1 f_1(x)$, she will be considered to be at risk for PE.

4.3. Clustering

Cluster analysis is an alternative approach to classification that aims to organize information into “clusters” (Tryon, 1939; Everitt et al., 2009). In contrast to classification methods, clustering is often referred to as ‘unsupervised learning’, in which observations are grouped based on its structure, and hence, pre-defined groups (or known outcomes) are not required.

In other words, clustering aims to sort observations into “clusters” or groups on the basis of similarities or distances (dissimilarities) in their characteristics or pattern, such that the degree of association between two observation is maximized within the same cluster, while minimized between clusters (Hill and Lewrick, 2007; Wichern and Johnson, 2007). Since clusters are determined on the basis of similarities or distances, no assumptions on group structures are necessary, and the only information required are the similarity measures or data from which similarities can be computed (Wichern and Johnson, 2007). Thus, these methods may be useful for prediction of diseases, as they can create clusters based on the similarities of certain measurements or characteristics (e.g. BMI or lifestyle) of patients.

4.3.1. K-means Clustering

K-means is a common unsupervised partitive clustering algorithm, developed by MacQueen in 1967, and then by Hartigan and Wong (MacQueen, 1967; Hartigan and Wong, 1979). It is a technique that attempts to partition a multivariate dataset into a set number of “clusters”, such that data points within a cluster are close to each other in multi-dimensional space, while data points between clusters are distinct.

Basically, K-means classifies the input data into a predefined number of clusters by

minimizing the sum of squares of distances between data and the corresponding cluster centroid. The distance between data and each cluster centroid is calculated. Then, each data point is assigned to a cluster centroid which is of minimum distance. The new cluster centroid is then calculated as the mean of all data that are assigned to that cluster. This process is iterated until convergence (see Fig. 4.3.3). A more detailed algorithm is given in Section 4.3.1.1.

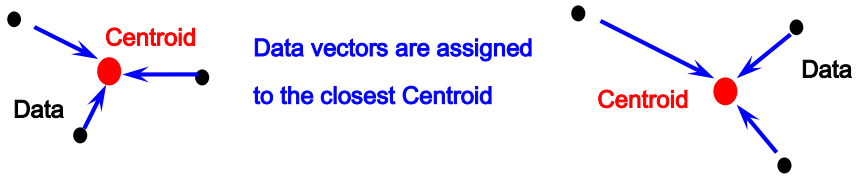

Fig. 4.3.3: K-means Clustering

4.3.1.1. K-menas Algorithm

The K-means algorithm aims at minimizing an *objective function*, in this case, a squared error function. The objective function can be expressed as

$$W = \frac{1}{2}\sum_{j=1}^{K} N_k \sum_{C(X_i)=k} \|X_i - C_j\|^2$$

where $\|.\|$ is a chosen distance measure, and $\|X_i - C_j\|$ represents the distance between a data point X_j and the cluster centre C_j. The objective function W is an indicator of the overall distance of the *n* data points from their respective cluster centers (Lattin et al., 2003).

In general, the K-means algorithm consists of three steps, as described below:

(i) The first step is to partition the data space into K initial clusters, and compute the centroid of these clusters.

(ii) In the second step, each data point is assigned to the cluster with the nearest centroid. Usually, the Euclidean distance is chosen as the distance measure. That

is, for each data point X_i, compute the distance between X_i and the centroid of each cluster C_j,

$$d(X_i, C_j) = \|X_i - C_j\|^2 = \sum_{l=1}^{p} (x_{il} - c_{jl})^2, \quad i = 1, 2, \dots, n;\ j = 1,2, \dots, K$$

where $C_j = [c_{j1}, c_{j2}, \dots, c_{jp}]^T$ is the centroid of cluster j. A data point X_i is then assigned to cluster $C(X_i) \in \{1, 2, \dots, K\}$ with the nearest centroid $C_{C(X_i)}$. That is, we determine $C(X_i) = \arg\min_j d(X_i, C_j)$ for $i = 1, 2, \dots, n$.

(iii) After all the data points have been assigned, the K centroids should be recalculated based on the new assignment of data in step two. Logically, the updated centroid of cluster j is given by the mean of the data points assigned to the cluster, that is, $C_j = \frac{1}{N_k} \sum_{i=1}^{n} \mathbb{I}(C(X_i) = j)\, X_i$, where $\mathbb{I}(.)$ is the indicator function, and $N_k = \sum_{i=1}^{n} \mathbb{I}(C(X_i) = j)$ is the number of data points assigned to cluster j.

The algorithm continues by iteratively repeating step two and three until the centroids move no more. In effect, the algorithm is minimising the within-cluster point-scatter

$$W_K = \frac{1}{2} \sum_{j=1}^{K} N_k \sum_{C(X_i)=k} \|X_i - C_j\|^2.$$

4.3.1.2. Advantages and Disadvantages

A major advantage of using k-means clustering is that it does not require previous knowledge of the outcome. In other words, unlike logistic regression, where a training dataset is needed, k-means "blindly" produce clusters solely based on the closeness of data points in multi-dimensional space.

Moreover, due to the simplicity of the k-means algorithm, it may be computationally faster than other clustering methods like hierarchical clustering for small numbers of

clusters. It also has the ability to cluster huge datasets quickly and efficiently (Arai and Barakbah, 2007). In fact, the computational time of k-means is proportional to the size of the dataset.

However, the major disadvantage is that the initial starting points are generated randomly, and hence, clustering results may vary (Khan and Ahmad, 2004). Also, as Kovesi stated that k-means method is "difficult to reach global optimum, but only in local minimum" (Kovesi et al., 2001). That is, k-means may not be an efficient clustering approach.

4.3.1.3. Applications

K-means have been mostly applied as a cluster technique to analyze multivariate data. A recent study by Bell et al. (Bell et al., 2010) used k-means to classify memory profiles in schizophrenia. Gorunescu et al. (Gorunescu et al., 2009) have shown that k-means appears to be a viable approach as an indicator for grouping patients into the five levels of Hepatic fibrosis, providing valuable knowledge of the relation between liver fibrosis and clinical data. A study by Watts and Worner (Watts and Worner, 2009) used k-means to estimate the invasive potential of insect pest species. Furthermore, McTaggart-Cowan et al. (McTaggart-Cowan et al., 2010) have used k-means to cluster Rasch results which aims to develop Rheumatoid Arthritis states for use in valuation studies. Mora-Florez et al. (Mora-Florez et al., 2009) also proposed a statistical model using k-means and mixture distributions for locating faults in power systems.

Interestingly, the k-means algorithm is useful not only as a statistical clustering method, but it can also be used as a computational method. Many developments and refinements have been done based on the algorithm, especially in areas of pattern recognition. For instance, Tsai et al. (Tsai et al., 2007) have proposed a pattern reduction algorithm

which can reduce the computation time of k-means in grouping data patterns. Moreover, Nagarajan et al. (Nagarajan et al., 2007) have used k-means to partition multi-dimensional phase-space in developing fast and efficient algorithms for retrieval of time series from a phoneme database to a user given pattern or query. Recently, Juang and Wu (Juang and Wu, 2010) have proposed a useful and interesting study on realization of tumour tracking on MRI brain image using colour-converted segmentation with k-means clustering.

4.3.1.4. Illustrative Example – Clustering of PTB

This section gives an example of practical application of K-means clustering for prediction of Preterm birth. A total of 2046 patients, including 111 Preterm birth (PTB) cases, were clustered based on 14 clinical predictors obtained at 15 weeks of gestation.

Table 4.3.1: Cluster centroids for Term birth and PTB

Predictors	Term birth	Preterm birth
Yrs of schooling	12.30	12.17
SEI	41.12	39.38
Mother's birthweight	3636.47	2814.37
Gravidity	1.31	1.33
Mths to conceive	5.51	7.33
Folate supplement (μg)	486.94	629.52
BMI	25.95	25.69
Height	166.61	163.99
MAP (15 wks)	78.86	79.44
Fruit per day (15 wks)	1.44	1.48
Vegetables per day (15 wks)	1.85	1.84
Smoking (15 wks)	0.16	0.22
Alcohol (15 wks)	1.46	1.33

For each of the 2 clusters, ‘term birth’ and ‘PTB’, the cluster centroids are obtained from the means in each group (shown in Table 4.3.1). The predicted group will be

determined based on the minimum distance between the patient's predictor and the cluster centroids. After all information required are recorded for a new patient, the Euclidean distance $\|X_i - C_j\|^2 = \sum_{l=1}^{p}(x_{il} - c_{jl})^2$, where $i = 1,2,\ldots,14$ is the number predictors and $j = 1,2$ is the number of clusters, between each of the predictors and cluster means are calculated. The cluster map (Fig. 4.3.4) shows the 95% confidence interval ellipse contour around the two cluster centers and the spread of each cluster, with the 'term birth' group shaded in blue, and 'PTB' group shaded in red.

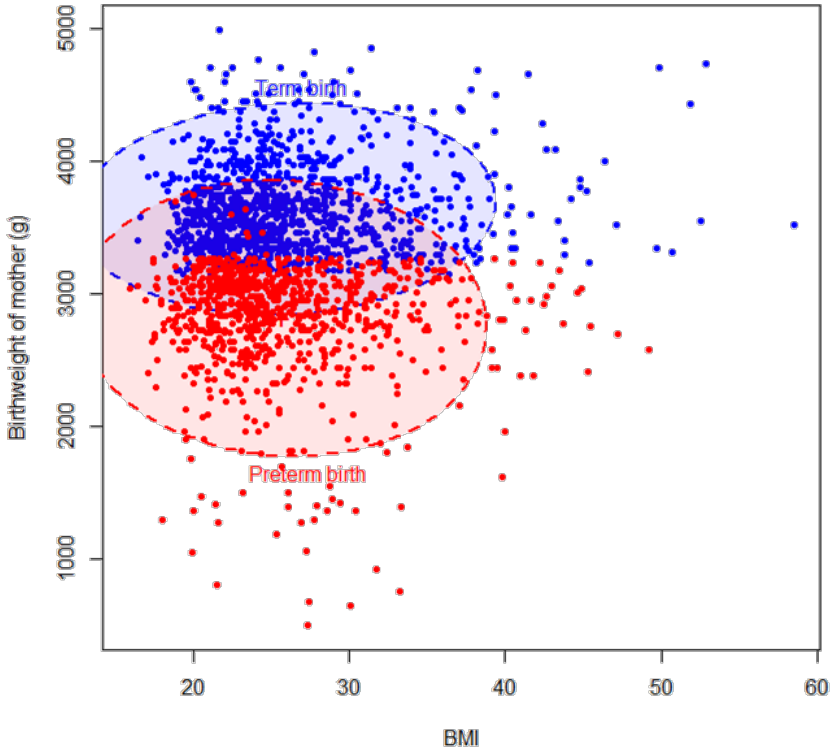

Fig. 4.3.4: K-means cluster map (blue=term birth, red=Preterm birth)

For instance, a new patient with the characteristics shown in Table 4.3.2, with her distance to each cluster centroid calculated as 554641.4 to the 'term birth' group and 36264.49 to the 'PTB' group. Since $\|x_{..} - c_{.2}\|^2 < \|x_{..} - c_{.1}\|^2$, she would be considered to be at risk of preterm birth.

Table 4.3.2: Example patient records and distance to cluster centroids

Patient records		Distance to Term birth centroid	Distance to PTB centroid
Yrs of schooling	12	$\|\|x_{..} - c_{\cdot 1}\|\|^2 = \left\|\left\| \begin{matrix} 12 \\ 39 \\ 2900 \\ 1 \\ 7.5 \\ 500 \\ 26 \\ 165 \\ 79 \\ 1 \\ 2 \\ 0 \\ 2 \end{matrix} - \begin{matrix} 12.30 \\ 41.12 \\ 3636.47 \\ 1.31 \\ 5.51 \\ 486.94 \\ 25.95 \\ 166.61 \\ 78.86 \\ 1.44 \\ 1.85 \\ 0.16 \\ 1.46 \end{matrix} \right\|\right\|^2$	$\|\|x_{..} - c_{\cdot 2}\|\|^2 = \left\|\left\| \begin{matrix} 12 \\ 39 \\ 2900 \\ 1 \\ 7.5 \\ 500 \\ 26 \\ 165 \\ 79 \\ 1 \\ 2 \\ 0 \\ 2 \end{matrix} - \begin{matrix} 12.17 \\ 39.38 \\ 2814.37 \\ 1.33 \\ 7.33 \\ 629.52 \\ 25.69 \\ 163.99 \\ 79.44 \\ 1.48 \\ 1.84 \\ 0.22 \\ 1.33 \end{matrix} \right\|\right\|^2$
SEI	39		
Mat. Birthweight	2900		
Gravidity	1		
Mths to conceive	7.5		
Folate	500		
BMI	26		
Height	165		
MAP	79		
Fruit/day	1		
Vegetables/day	2	$\therefore \sum (x_{..} - c_{\cdot 1})^2 \approx 554641.4$	$\therefore \sum (x_{..} - c_{\cdot 2})^2 \approx 36264.49$
Smoking	0		
Alcohol	2		

4.3.2. Chernoff Faces

Chernoff faces is a powerful graphical visualization tool for multi-dimensional data, introduced by Herman Chernoff in 1971 (Chernoff, 1973), where each variable is mapped to a particular face feature on a cartoon face. Its ability to visualize multiple variables simultaneously allows simple clustering of cases with similar characteristics, i.e. visual clustering. The shape, size and location of ears, eyes, nose and mouth are controlled by certain variables. Since humans have exquisite sensitivity to facial expressions, Chernoff faces took advantage of this aspect, and allow easy perception of multiple measurements in parallel.

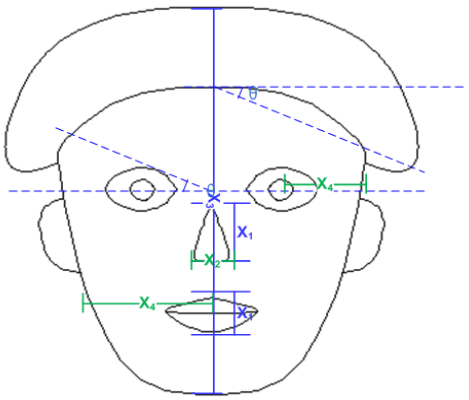

Fig. 4.3.5: Chernoff face

For a p-dimensional data, each variable $X_1, X_2, \dots, X_p$ are assigned to a feature on the schematic face and rules are designed to determine the coordinate, size and curvature of each feature. For instance, X_1 may represent the height of the nose while X_2 controls the width of the nose, and X_3 controls the width of the face (Fig. 4.3.5).

4.3.2.1. Advantages and Disadvantages

One of the main advantages of Chernoff faces is the simple multivariate visualization it provides, that allows easy clustering of individuals with similar characteristics via

human eye. This method can also be used to visualize combinations of continuous and categorical variables, which is useful for comparing combinations of clinical and SNP predictors.

However, one of the major limitations of this method is that cluster results are subjective, since the sensitivity of facial features differs for different people, and so one needs to be cautious when assigning variables to certain facial features, as different people are more sensitive on certain facial features.

Another limitation is the inability to assess the quantity of variables plotted, since all variables are plotted based on relative scales, true quantity of the variables cannot be assessed.

In addition, when Chernoff faces are plotted for large samples, it is difficult to perform visual clustering through human eye. As a result, a number of computational algorithms have been developed that allows automatic clustering of Chernoff faces, which includes K-means (Wang et al., 2007), K-NN, and V-system (Song et al., 2009; Song et al., 2010; Becker et al., 2011). However, this study will only use Chernoff faces as a simple exploratory clustering approach for small sub-groups, and as a visualization method for cluster results.

4.3.2.2. Illustrative Example - Visual Clustering of PTB

This section demonstrates the use of Chernoff faces as a clustering technique to identify Preterm birth (PTB) cases. The Chernoff faces for 100 randomly selected patients were constructed, with 8 PTB cases (highlighted with yellow) and 92 term births, with patient numbers shown on top of the faces (Fig. 4.3.6). A list of clinical predictors and SNPs with their corresponding face characteristics are shown in Table 4.3.1.

Table 4.3.1: Facial characteristics and corresponding predictors

Characteristics		**Predictors**	**Characteristics**		**Predictors**
Face	height	Age	Hair	height	Cigarettes (15 wks)
	width	Cervical length		width	anxiety index
	curvature	gravidity		style	mother born preterm
Mouth	height	alcohol (15 wks)	Nose	height	*IL-6*
	width	BMI		width	*F5*
	smiling	Marijuana (15 wks)	Ear	height	*TGFβ*
Eye	height	SEI		width	*IL1β*
	width	depression			

With all predictors plotted into faces, it is relatively easy to identify and group patients with similar characteristics. For instance, smiling faces indicates patients who are still using Marijuana at 15 weeks of gestation (e.g. patient 996 and 3196), and faces with closed-eyes indicate patients who had a low SEI and are depressed (e.g. patient 795, 1211, and 2961). Similarly, faces with a wide-open mouth indicate patients who consumed alcohol and are obese (e.g. patient 505, 1650 and 3193).

Patients who deliver PTB often have a shorter cervical length compared to uncomplicated pregnancies, which is reflected by a thinner or smaller face, indicating that most PTB cases have a shorter cervical length as well as younger in age (e.g. patient 972, 1587 and 2908). There is an exception for patient 795, which shows a relatively long face. She is, in fact, the oldest patient in this dataset, which agrees with the literature that extremes of age are also a risk factor for PTB.

Some more distinct face features that are unique to PTB cases can be seen by the absence of ears, which corresponds to a particular genotype in $TGF\beta$ and $IL1\beta$, or a high hair style, which corresponds to whether the mother herself was born preterm. These distinct features are most helpful in classifying new patients.

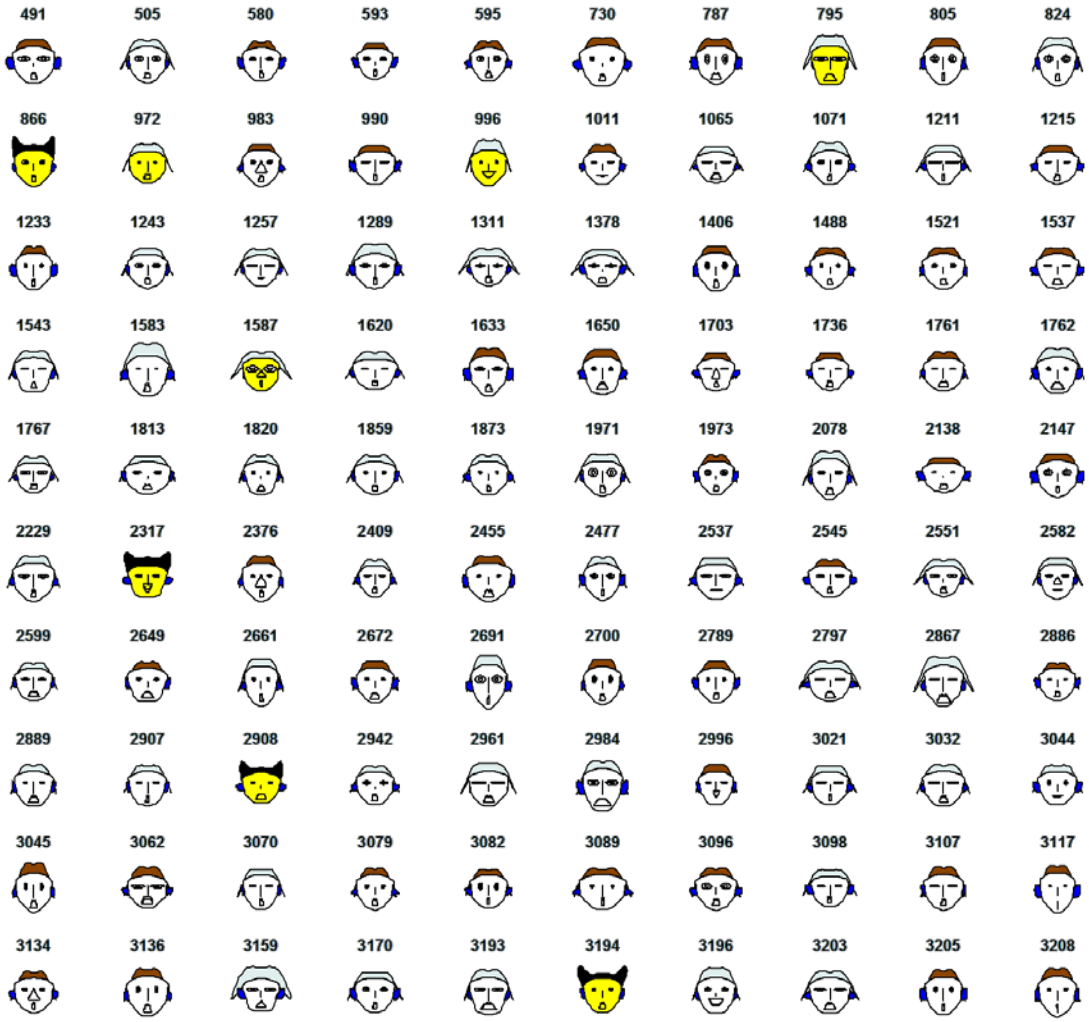

Fig. 4.3.6: Chernoff faces displaying 11 clinical and 4 genetic predictors (yellow=PTB cases)

For instance, Fig. 4.3.7 shows the Chernoff faces for 2 new patients. Clearly, a) has a high hair style, which is a distinct characteristic that is unique to PTB indicating that the mother herself was born preterm. The face also appears to be relatively thin, which indicate a shorter cervical length. For b), the open eyes indicate that the patient has a higher SEI and is not depressed, with no distinct PTB characteristics. Hence, a) would be considered to be at risk for PTB, while b) would be at low risk.

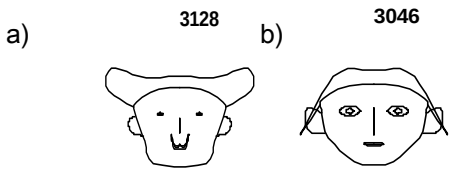

Fig. 4.3.7: Example Chernoff faces; a) PTB b) term birth

4.4. Model and Variable Selection

As discussed in Chapter 3, there are more than 1300 variables, including 100 SNPs in mother, father and baby trios in the SCOPE database. Although with a sample size of 3201, dimensionality issues such as the 'small n, large p' problem do not cause too much concern, ways to reduce the number of variables are still necessary to obtain a practically sufficient model. Hence, this section will investigate a number of model and variable selection techniques applicable to this study.

For all models, a backward-stepwise approach is used, in which predictors are eliminated from the full model (i.e. model with all predictors included) in each step. The selection process is controlled by penalty functions or regularization statistics, which aims at eliminating variables that are considered unrelated or useless in improving the fit or prediction accuracy of models to prevent model over-fitting.

4.4.1. Penalty Functions

The Akaike Information Criterion (AIC) (Akaike, 1974) is a common penalty function used in regression. It is a relative measure based on the likelihood function and the number of predictors in each step. It is given by $AIC = 2k - 2\,ln\,L(\beta)$, where$k$ is the number of predictors in the current step and$L(\beta) = \sum_{i=1}^{n}\left[y_i\beta^T x_i - log\left(1 + e^{\beta^T x_i}\right)\right]$ is the likelihood function. The model with minimum AIC is believed to be the best (i.e. optimal) model that describes uncertainty, which is often a model with balanced fit and optimal number of predictors.

The Bayesian Information Criterion (BIC) (Schwarz, 1978) penalty algorithm is another common model selection criterion similar to AIC, but instead of modelling the uncertainty, it aims at finding a true model. This is often a stricter penalization criterion,

and thus provides a simpler model than AIC. BIC is given by $BIC = k\,ln(n) - 2\,ln\,L(\beta)$, where k is the number of predictors in the current step, $L(\beta)$ is the likelihood function and$n$ is the number of data observations included in analysis.

While AIC and BIC provide informative measures on model complexity, there have been concerns on multiple testing issues, as both penalize functions are independent to the number of iterations (or models) tested. Ideally, methods such as Multiple-Step FDR (MSFDR) penalty would be most ideal (Benjamini and Gavrilov, 2009). However, although such methods have been widely applied to linear regression, an analogous algorithm for methods used in this study (e.g. logistic regression) is still under development. Hence, other sophisticated regularization methods are considered.

4.4.2. Regularization Methods

Elastic-net regularization (Zou and Hastie, 2005)is a regularization method that have gained popularity recently, due to its computational simplicity, which greatly reduces the computation time, and features that overcome conventional methods such as dimensional restriction. It has also been shown to provide a better control of FDR by balancing Type I and Type II errors (Wu et al., 2009).

In contrast to AIC and BIC, which provides model-based measures, elastic-net aims at shrinking the coefficients of each predictor to 0 (i.e. variable-based). The coefficient estimates are given by $\hat{\beta} = arg\,\min_{\beta}\|y - X\beta\|^2 + \lambda_2\|\beta\|^2 + \lambda_1\|\beta\|_1$, where λ is the tuning parameter, selected based on minimum cross-validation error for optimum prediction performance. It is worth noting that Elastic-Net is a 2-step penalty, involving both Ridge (ℓ_2: $\lambda\sum_{j=1}^{p}(\beta[j])^2 = \lambda_2\|\beta\|^2$) and Least Absolute Shrinkage and Selection Operator (LASSO) regression (ℓ_1: $\lambda\sum_{j=1}^{p}|\beta[j]| = \lambda_1\|\beta\|_1$). Hence, both variable

selection and shrinkage are performed, and are robust to correlated variables.

Fig. 4.4.8 shows the variable shrinkage pathway for different values of λ. Each predictor is represented by a line, and the distance between the coefficients indicates their correlation (i.e. lines that are closer together indicate a stronger correlation). As expected, the coefficients of predictors will eventually shrink towards zero. Since the purpose of this study is to develop prediction models, the optimal set of variables are chosen based on minimum cross-validation error.

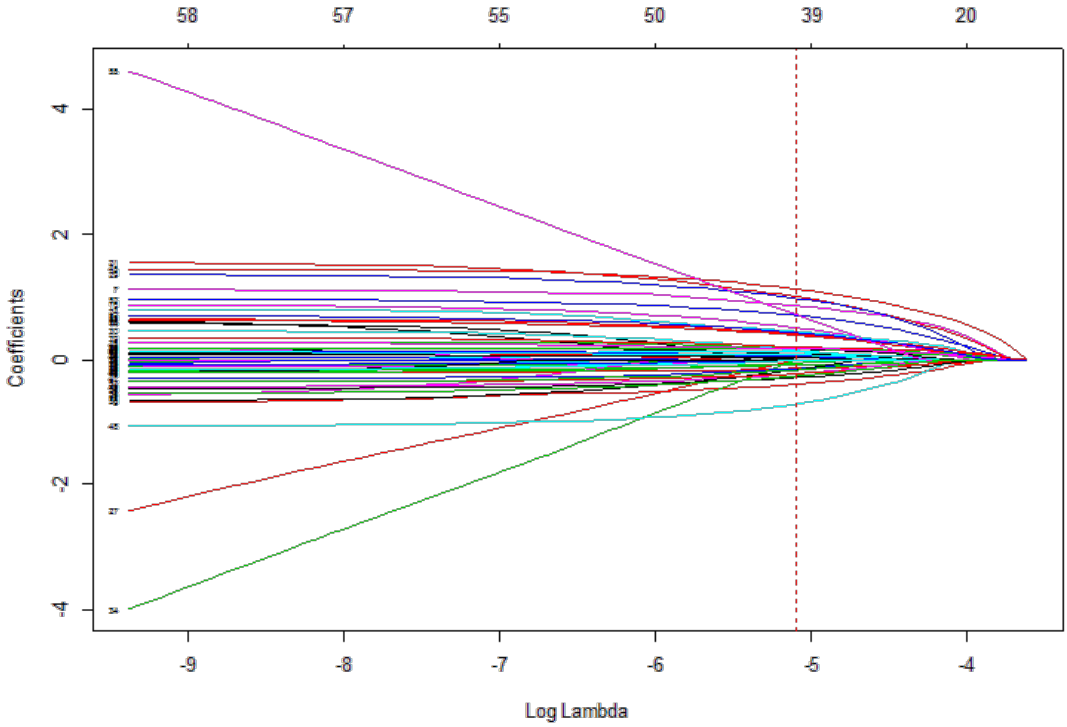

Fig. 4.4.8: Elastic-Net variable shrinkage pathway

(red dotted line indicates selected set with minimum cross-validation error)

4.5. Validation Methods

Sensitivity and specificity (Taube, 1986; Altman and Bland, 1994) are calculated as measures of goodness of classification. From the training database, where the outcomes are known, false positives (FP), false negatives (FN), true positives (TP) and true negatives (TN) can be obtained.

Table 4.5.1: Observed Frequencies in a sample of *n* subjects

	True Disease				
Predicted	**Yes (D)**		**No ($\overline{D}$)**		**Total**
Yes (+)	**a**	**TP**	**b**	**FP**	**a+b**
No (-)	**c**	**FN**	**d**	**TN**	**c+d**
Total	**a+c**		**b+d**		**n(=a+b+c+d)**

False negatives are cases of patients who are predicted to not develop the disease, but are actually diagnosed with the disease. False positives are cases of patients who are predicted to develop the disease, but who do not develop the disease. Table 4.5.1 displays the cross classification table depicting FP, FN, TP and TN.

4.5.1. Sensitivity and Specificity

Sensitivity (r) is the proportion of truly classified cases of the disease, or equivalently, patients who are predicted to develop the disease and have developed the disease as predicted. It is obtained as follows, $r = P(+|D) = \frac{P(+\cap D)}{P(D)} = \frac{a}{a+c} = \frac{\text{TP}}{\text{TP+FN}}$. Specificity ($s$) is the proportion of cases correctly predicted as "non-disease", that is, patients who are predicted to not develop the disease and did not develop as predicted. It is estimated by $s = P(-|\overline{D}) = \frac{P(-\cap\overline{D})}{P(\overline{D})} = \frac{d}{b+d} = \frac{TN}{FP+TN}$. Similarly, if α is the Type I error, i.e. probability of falsely predicted case of disease, and β is the Type II error, i.e. probability of falsely predicted "normal", then $r = 1-\alpha$ and $s = 1-\beta$.

The probability of developing the disease given that it is correctly predicted, is calculated using $P(D) = \frac{1}{a+b}$ and the probability of not developing the disease given that it is correctly predicted, is calculated using $P(\overline{D}) = \frac{1}{c+d}$. The Index of validity (I_v) represents the proportion of correctly predicted outcome, given by $I_v = \frac{a+d}{n} = \frac{\text{TP+TN}}{n}$ (Taube, 1986; Simon and Boring, 1990; Altman and Bland, 1994).

Table 4.5.2: Population model: expected frequencies (Taube, 1986)

	True	Disease	
Predicted	**Yes (D)**	**No ($\overline{D}$)**	**Total**
Yes (+)	**rP**	**(1-s)Q**	**rP+(1-s)Q**
No (-)	**(1-r)P**	**sQ**	**(1-r)P+sQ**
Total	**P**	**Q**	**1**

Suppose that, in a certain population, P represents the prevalence of disease, and $Q = 1 - P$. The expected relative frequencies in the population are given in Table 4.5.2. The situation when the test results are unrelated to the disease is characterized by $r + s = 1$. If $r + s = 1$, $P(D) = P$, $P(\overline{D}) = Q$ and $P(D) + P(\overline{D}) = 1$ (Taube, 1986).

As a measure of efficiency of a diagnostic test, Youden (Youden, 1950) suggested the index $J = r + s - 1$. Similarly, $J = 1 - \alpha - \beta$. Biggerstaff (Biggerstaff, 2000) pointed out that it is in a certain sense the best available summary measure. Hilden and Glasziou (Hilden and Glasziou, 1996) give a good geometric characterization of Youden's index as the area under the curve. In addition, Böhning et al.(Bohning et al., 2008) suggested the use of the simple sum of sensitivity and specificity ($r + s$), where $r + s = 1 - J$, and$J$ is the Youden's Index. This measure will be used as an accuracy ranking for all models in this study.

4.5.2. Predictive Values

Apart from sensitivity and specificity, the Positive Predictive Value (PPV) and Negative Predictive Value (NPV) for each model are also obtained. PPV is the proportion of true positives in predicted cases of disease, which is given by $PPV = P(D|+) = \frac{P(D \cap +)}{P(+)} = \frac{a}{a+b} = \frac{\text{TP}}{\text{TP+FN}}$. The NPV is the proportion of true negatives in predicted cases of 'normal', which is given by $NPV = P\left(\overline{D}\middle|-\right) = \frac{P\left(\overline{D} \cap -\right)}{P(-)} = \frac{d}{c+d} = \frac{\text{TN}}{\text{FP+TN}}$. Models with high sensitivity are likely to have a higher NPV, and similarly, models with high specificity are likely to have a higher PPV.

An overall ratio of true vs. false classification is also obtained. This is the proportion of truly classified cases of disease and 'normal' on all data analyzed.

$$\text{Overall} = \frac{P(+ \cap D) + P\left(- \cap \overline{D}\right)}{P\left(D \cup \overline{D}\right)}$$

One point to note is that, since PPV depends on the prevalence of disease $P(D)$, predictions on rare diseases, such as PE and PTB, will have low PPVs even when a high specificity is achieved. Hence, this study takes advantage of the high NPVs as a process of elimination to 'rule-out' the probability of PE or PTB for individuals, rather than relying on low PPVs. This will be further discussed in Chapter 5.

4.5.3. Receiver Operating Characteristic Curve

The receiver operating characteristic (ROC) curve is a graphical display of the sensitivity and false positives, i.e. 1-specificity, that is typically used to evaluate clinical utility for both diagnostic and prognostic models. It assesses how well a test or model discriminates or separates individuals into two classes, e.g. diseased and non-diseased (Griner et al., 1981; Cook, 2008).

Suppose there are two classification/predicted outcomes, e.g. disease and non-disease (Fig. 4.5.9). An overlap of the distribution of test results is likely to occur, and these are the false positives (FP) and false negatives (FN). For every threshold (or cutoff) value chosen to discriminate between the two outcomes, the sensitivity and specificity vary. The ROC curve is useful as it describes the compromises that can be made between the relative frequencies of true positive (TP), true negative (TN), false positive (FP), and false negative (FN), as the threshold (or cutoff point) is varied (Metz, 1978; Perkins and Schisterman, 2006).

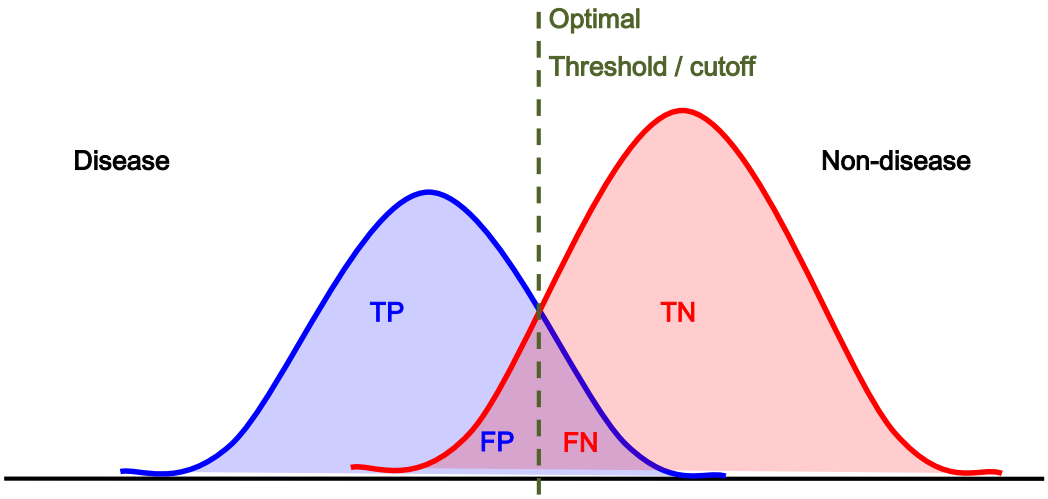

Fig. 4.5.9: Distribution of test results and optimal cutoff point

When perfect discrimination occurs, i.e. the distribution of test results are completely separated such that there is no overlap, the curve will pass through the upper left corner, indicating a sensitivity and specificity of 100%. Hence, a curve closer to the upper left corner signifies a more accurate test (Zweig and Campbell, 1993).

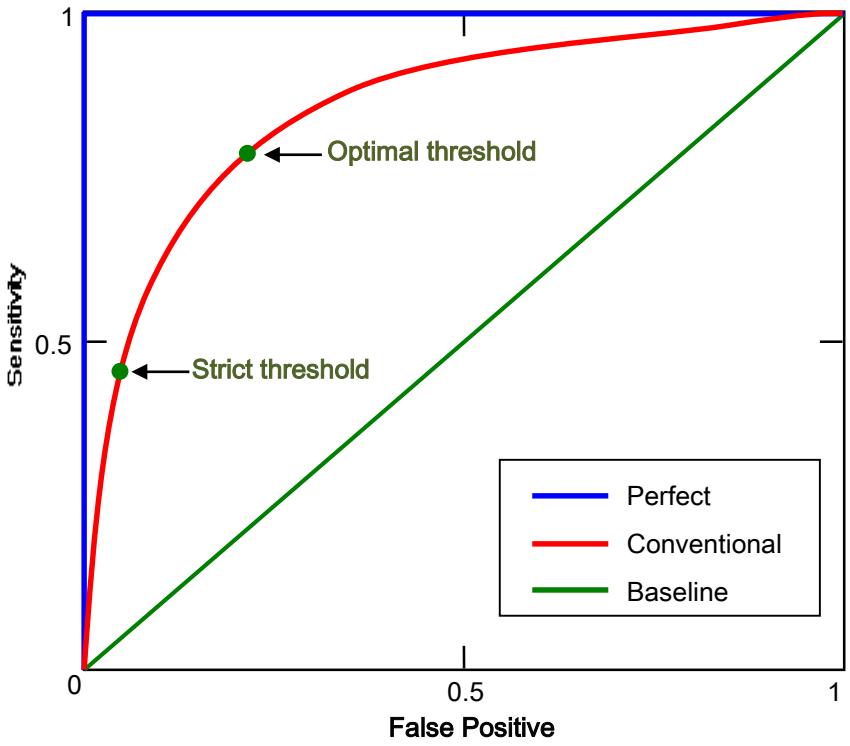

Fig. 4.5.10: Perfect, conventional and baseline ROC curves

The Area Under Curve (AUC) (Hanley and McNeil, 1982; Bradley, 1997) is the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one (Fawcett, 2006). By convention, the baseline of AUC is 0.5, that is, when the probability of being classified as disease and non-disease are equal. An AUC value of 1 indicate perfect separation of the test values, i.e. with no FP and FN (Zweig and Campbell, 1993).

4.6. Summary

The main aim of this study is to develop prediction models for two major complications of pregnancy, preeclampsia (PE) and spontaneous preterm birth (SPTB). Five statistical and data mining techniques have been proposed to identify prediction rules for PE and SPTB during early stages of pregnancy. These include classification through Logistic regression and Discriminant analysis, and clustering through K-means and Chernoff faces.

Elastic-Net regularization will be used as a model or variable selection tool for regression to obtain a practically manageable prediction model. All models obtained will then be validated and compared through sensitivity, specificity, and the area under ROC curve.

Seeing that a single model may not be satisfactory, this study also propose a tiered prediction system, which will be further discussed in Chapter 5, in which individual models are developed for specific tiers, and the estimated risk at each stage will then be integrated using Bayes' Theorem to screen women at risk by applying the ‘Process of Elimination’.

References

Agresti, A. (2002). *Categorical Data Analysis*. New Jersey, John Wiley & Sons, Inc.

Akaike, H. (1974). "A new look at the statistical model identification." *Automatic Control, Transactions on IEEE* **19**(6): 716-723.

Altman, D. G. and Bland, J. M. (1994). "Diagnostic tests. 1: Sensitivity and specificity." *Bmj* **308**(6943): 1552.

Altman, D. G. and Bland, J. M. (1994). "Diagnostic tests. 2: predictive values." *Bmj* **309**(102): 6947.

Angillieri, M. Y. E. (2010). "Application of frequency ratio and logistic regression to active rock glacier occurrence in the Andes of San Juan, Argentina." *Geomorphology* **114**(3): 396-405.

Arai, K. and Barakbah, A. R. (2007). Hierarchical K-means: an algorithm for centroids initialization for K-means. *Reports of the Faculty of Science and Engineering*, Saga University. **36:** 25-31.

Austin, P. C., Tu, J. V. and Lee, D. S. (2010). "Logistic regression had superior performance compared with regression trees for predicting in-hospital mortality in patients hospitalized with heart failure." *J Clin Epidemiol* **63**(10): 1145-1155.

Ayuso, M., Guillen, M. and Alcaniz, M. (2010). "The impact of traffic violations on the estimated cost of traffic accidents with victims." *Accident Analysis & Prevention* **42**(2): 709-717.

Becker, N., Toedt, G., Lichter, P. and Benner, A. (2011). "Elastic SCAD as a novel penalization method for SVM classification tasks in high-dimensional data." *BMC Bioinformatics* **12**(1): 138.

Bell, M. D., Johannesen, J. K., Greig, T. C. and Wexler, B. E. (2010). "Memory profiles in schizophrenia: categorization validity and stability." *Schizophr Res* **118**(1-3): 26-33.

Benjamini, Y. and Gavrilov, Y. (2009). "A simple forward selection procedure based on false discovery rate control." *The Annals of Applied Statistics* **3**(1): 179-198.

Biggerstaff, B. J. (2000). "Comparing diagnostic tests: a simple graphic using likelihood ratios." *Stat Med* **19**(5): 649-663.

Bohning, D., Bohning, W. and Holling, H. (2008). "Revisiting Youden's index as a useful measure of the misclassification error in meta-analysis of diagnostic studies." *Stat Methods Med Res* **17**(6): 543-554.

Bradley, A. P. (1997). "The use of the area under the ROC curve in the evaluation of machine learning algorithms." *Pattern Recognit* **30**(7): 1145-1159.

Brooks, J. S. (2010). "The Buddha mushroom: Conservation behavior and the development of institutions in Bhutan." *Ecological Economics* **69**(4): 779-795.

Chernoff, H. (1973). "The use of faces to represent points in k-dimensional space graphically." *Journal of American Statistical Association* **68**(342): 361-368.

Cho, S., Kim, J. and Bae, J. K. (2009). "An integrative model with subject weight based on neural network learning for bankruptcy prediction." *Expert Systems with Applications* **36**(1): 403-410.

Cook, N. R. (2008). "Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve." *Clin Chem* **54**(1): 17-23.

Dai, G., Yeung, D.-Y. and Qian, Y.-T. (2007). "Face recognition using a kernel fractional-step discriminant analysis algorithm." *Pattern Recognit* **40**(1): 229-243.

du Jardin, P., Ponsaill, J., Alunni-Perret, V. and Quatrehomme, G. (2009). "A comparison between neural network and other metric methods to determine sex from the upper femur in a modern French population." *Forensic Sci Int* **192**(1-3): 127.e121-127.e126.

Dufty, D. (2007). "What is logistic regression?". Retrieved 20 July 2010, from http://www.statgun.com/tutorials/logistic-regression.html.

Everitt, B., Landau, S. and Leese, M. (2009). *Cluster Analysis*. New York, Wiley.

Fawcett, T. (2006). "An introduction to ROC analysis." *Pattern Recognition Letters*

27(8): 861-874.

Fisher, R. A. (1936). "The use of multiple measurements in taxonomic problems." *Ann of Eugenics* **7**: 179-188.

Gorunescu, M., Gorunescu, F., Badea, R. and Lupsor, M. (2009). "Evaluation on liver fibrosis stages using the k-means clustering algorithm." *Annals of the University of Craiova - Mathematics and Computer Science Series* **36**(2): 19-24.

Griner, P. F., Mayewski, R. J., Mushlin, A. I. and Greenland, P. (1981). "Selection and interpretation of diagnostic tests and procedures. Principles and applications." *Ann Intern Med* **94**(4 Pt 2): 557-592.

Gu, Z. (2002). "Analyzing bankruptcy in the restaurant industry: A multiple discriminant model." *International Journal of Hospitality Management* **21**(1): 25-42.

Hanley, J. A. and McNeil, B. J. (1982). "The meaning and use of the area under a receiver operating characteristic (ROC) curve." *Radiology* **143**: 29-36.

Hartigan, J. A. and Wong, M. A. (1979). "A K-means clustering algorithm." *Applied Statistics* **28**: 100-108.

Hastie, T., Tabshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, Springer-Verlag.

Hilden, J. and Glasziou, P. (1996). "Regret graphs, diagnostic uncertainty and Youden's Index." *Stat Med* **15**(10): 969-986.

Hill, T. and Lewrick, P. (2007). *STATISTICS Methods and Applications*. Tulsa, StatSoft.

Hosmer, W. D. and Lemeshow, S. (2005). *Applied Logistic Regression*. Hoboken, NJ, USA, John Wiley & Sons, Inc.

Hubert, M. and Van Driessen, K. (2004). "Fast and robust discriminant analysis." *Computational Statistics & Data Analysis* **45**(2): 301-320.

Jaba, E., Jemna, D. V., Viorica, D. and Balan, C. B. (2007). "Discriminant analysis in the study of romanian regional economic development." *Analele Stiintifice ale*

Universitatii "Alexandru Ioan Cuza" din Iasi - Stiinte Economice **54**: 147-153.

Jo, H., Han, I. and Lee, H. (1997). "Bankruptcy prediction using case-based reasoning, neural networks, and discriminant analysis." *Expert Systems with Applications* **13**(2): 97-108.

Juang, L.-H. and Wu, M.-N. (2010). "MRI brain lesion image detection based on color-converted K-means clustering segmentation." *Measurement* **43**(7): 941-949.

Khan, S. S. and Ahmad, A. (2004). "Cluster center initialization algorithm for K-means clustering." *Pattern Recognition Letters* **25**(11): 1293-1302.

Kovesi, B., Boucher, J.-M. and Saoudi, S. (2001). "Stochastic K-means algorithm for vector quantization." *Pattern Recognition Letters* **22**(6-7): 603-610.

Lattin, J., Carroll, J. D. and Green, P. E. (2003). *Analyzing Multivariate Data*, Duxbury.

Lin, G.-C., Wang, W.-J., Wang, C.-M. and Sun, S.-Y. (2010). "Automated classification of multi-spectral MR images using Linear Discriminant Analysis." *Computerized Medical Imaging and Graphics* **34**(4): 251-268.

Lo, A. W. (1986). "Logit versus discriminant analysis : A specification test and application to corporate bankruptcies." *Journal of Econometrics* **31**(2): 151-178.

Lu, J., Plataniotis, K. N. and Venetsanopoulos, A. N. (2005). "Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition." *Pattern Recognition Letters* **26**(2): 181-191.

MacQueen, J. B. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of 5th Berkely Symposium on Mathematical Statistics and Probability*, University of California Press.

Marzban, C., Paik, H. and Stumpf, G. J. (1997). "Neural Networks vs. Gaussian Discriminant Analysis." *AI Applications* **10**(1): 49-58.

McLachlan, G. J. (2004). *Discriminant Analysis and Statistical Pattern Recognition*. New Jersey, Wiley-Interscience.

McTaggart-Cowan, H. M., Brazier, J. E. and Tsuchiya, A. (2010). "Clustering Rasch Results: A Novel Method for Developing Rheumatoid Arthritis States for Use in Valuation Studies." *Value Health* **13**(6): 787-795.

Metz, C. E. (1978). "Basic principles of ROC analysis." *Semin Nucl Med* **8**(4): 283-298.

Mora-Florez, J., Cormane-Angarita, J. and Ordonez-Plata, G. (2009). "k-means algorithm and mixture distributions for locating faults in power systems." *Electric Power Systems Research* **79**(5): 714-721.

Nagarajan, R., Nagarajan, A. and Milanova, M. (2007). *Information retrieval from a phoneme time series database*. Signal Processing and Information Technology, 2007 IEEE International Symposium on.

Okamoto, K. and Harasawa, Y. (2010). "Prediction of symptomatic depression by discriminant analysis in Japanese community-dwelling elderly." *Arch Gerontol Geriatr* **52**(2): 177-180.

Perkins, N. J. and Schisterman, E. F. (2006). "The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve." *Am J Epidemiol* **163**(7): 670-675.

Pnevmatikakis, A. and Polymenakos, L. (2009). "Subclass linear discriminant analysis for video-based face recognition." *Journal of Visual Communication and Image Representation* **20**(8): 543-551.

Schwarz, G. (1978). "Estimating the Dimension of a Model." *The Annals of Statistics* **6**(2): 461-464.

Simon, D. and Boring, I. J. (1990). *Sensitivity, Specificity, and Predictive Value*, Butterworths.

Song, R., Zhao, Z. and Ou, M. (2009). *A novel clustering method for Chernoff faces based on V-system*. Information and Automation, 2009. ICIA '09. International Conference on.

Song, X.-n., Zheng, Y.-j., Wu, X.-j., Yang, X.-b. and Yang, J.-y. (2010). "A complete fuzzy discriminant analysis approach for face recognition." *Applied Soft Computing*

10(1): 208-214.

Sueyoshi, T. and Goto, M. (2009). "Can R&D expenditure avoid corporate bankruptcy? Comparison between Japanese machinery and electric equipment industries using DEA-discriminant analysis." *European Journal of Operational Research* **196**(1): 289-311.

Tabachnick, B. and Fidell, L. (1996). *Using Multivariate Statistics*. New York, Harper Collins.

Tansey, R., White, M., Long, R. G. and Smith, M. (1996). "A Comparison of Loglinear Modeling and Logistic Regression in Management Research " *J Manage* **22**(2): 339-358

Taube, A. (1986). "Sensitivity, specificity and predictive values: a graphical approach." *Stat Med* **5**(6): 585-591.

Tryon, R. C. (1939). *Cluster analysis*. New York, McGraw-Hill.

Tsai, C.-W., Yang, C.-S. and Chiang, M.-C. (2007). *A Time efficient Pattern Reduction algorithm for k-means based clustering*. Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on.

Wang, J., Hong, W. and Li, X. (2007). "Novel clustering algorithm using K-means Chernoff faces." *Chinese Journal of Scientific Instrument* **28**(10): 1916-1920.

Watts, M. J. and Worner, S. P. (2009). "Estimating the risk of insect species invasion: Kohonen self-organising maps versus k-means clustering." *Ecological Modelling* **220**(6): 821-829.

Wichern, D. W. and Johnson, R. A. (2007). *Applied Multivariate Statistical Analysis*. New Jersey, Prentice Hall.

Wu, T. T., Chen, Y. F., Hastie, T., Sobel, E. and Lange, K. (2009). "Genome-wide association analysis by lasso penalized logistic regression." *Bioinformatics* **25**(6): 714-721.

Youden, W. J. (1950). "Index for rating diagnostic tests." *Cancer***3**(1): 32-35.

Zou, H. and Hastie, T. (2005). "Regularization and variable selection via the Elastic Net." *Journal of Royal Statistical Society, Series B* **67**(2): 301-320.

Zweig, M. H. and Campbell, G. (1993). "Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine." *Clin Chem* **39**(4): 561-577.

Chapter 5: Tiered Modelling Approach

5.1. Overview

In view of the fact that a single prediction model may not be satisfactory, this study proposes a multilevel modeling approach, aiming to increase the precision of identifying women at risk at each tier, and achieve prediction through "elimination". The main aim of the tiered approach is to develop a "system" or "sequence" of prediction models, where dedicated prediction models are obtained for certain stages of pregnancy or when additional information is available (Fig. 5.1.1).

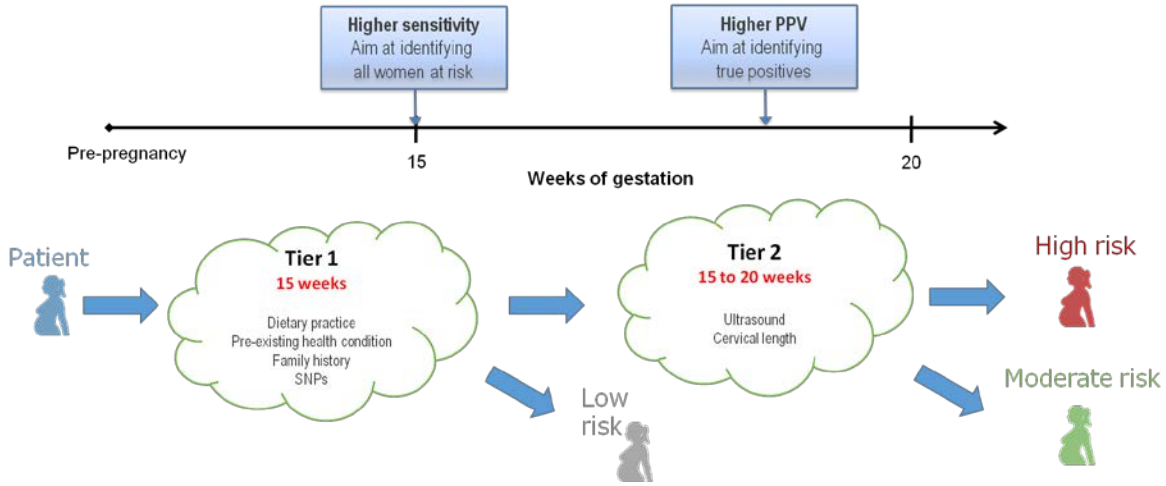

Fig. 5.1.1: Tiered prediction approach

With Tier 1 identifying women at low risk, and further classifying women at high risk in Tier 2.

Individual models developed at each tier are adjusted to accommodate the specific needs, and their predicted risks are then later integrated with the subsequent tiers to provide an overall risk classification (to be further discussed in Section 5.2).

With the first tier as initial screening, a higher sensitivity is preferred, as the main purpose of this tier is to identify all women who may be at risk. At this stage, the prediction will be based on predictors available at first antenatal visit (for SCOPE, at

15 weeks of gestation), which includes current dietary practice, pre-existing health conditions, family history, as well as clinical measurements such as blood pressure.

For the second tier prediction which can be performed at or prior to 20 weeks of gestation, a high positive predictive value (PPV), i.e. low false positive rates, is preferred to minimize the chance of unnecessary interventions. Predictors at this tier may include SNPs or details of ultrasound scan.

The individual models for tiers 1 and 2, described above, will be developed using statistical and data mining methods (as discussed in Chapter 4), with the best model selected based on penalty functions and accuracy measures, and then integrated by calculating the post-test odds using Bayes' theorem at each stage of pregnancy. The predicted risk is then further classified into 3 classes (low, moderate, and high risk). This will be further discussed in Section 5.2.

A major advantage of a tiered approach is that risk estimates or prediction can be obtained throughout pregnancy, which allows constant monitoring and update of predicted risk for individuals when new predictors are available or when conditions change, and hence, the level of care may be tailored for individual women. In addition, having the first tier with a high sensitivity at first visit will assure that the proportion of disease amongst women predicted at low risk at tier 1 is lower than those predicted at risk. This means that by 15 weeks of gestation, the first group of low-risk women can be identified and continue regular antenatal visits, while those identified at risk may go through further screening at tier 2 and may be recommended for tailored care.

5.2. Model Integration

After individual models were obtained from methods discussed in Chapter 4 for each tier, the final process of model development is to integrate risk predictions from all tiers to perform a process of elimination, which may assist in stratifying the level of care for individual patients. This can be achieved by applying the Bayes' theorem to obtain a post-test odds of tier 2 based on prior 'guess' obtained from the predicted risk of tier 1 and the likelihood of tier 2 individual model (Fig. 5.2.1).

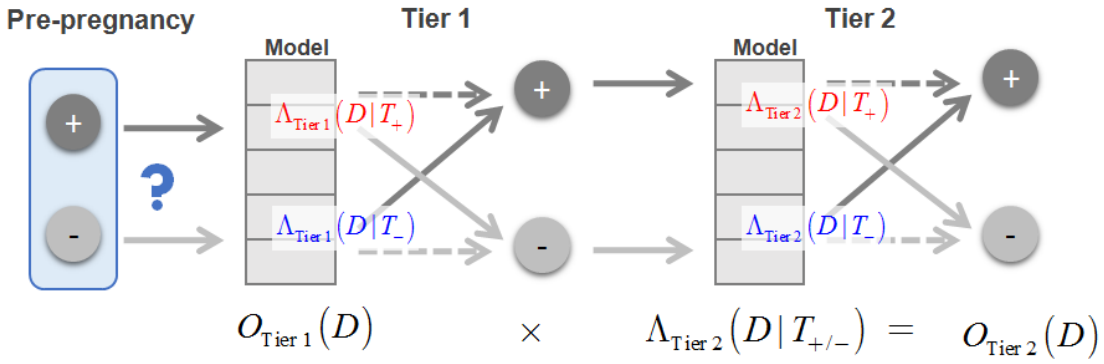

Fig. 5.2.1: Model integration overview

Bayes theorem is applied in final risk prediction, where the odds of Tier 1 is multiplied by the likelihood of a positive or negative result in Tier 2.

5.2.1. Bayes Theorem

Bayes' theorem have been widely applied in evidence-based medicine as well as clinical decision support systems for assessing risks for individual patients following a positive or negative test result (Hall, 1967; Round, 2001; Lindgaard et al., 2009; Sox et al., 2013). With the application of Bayes' theorem, an adjusted predicted probability or post-test probability of disease that incorporates with a pre-test probability can be obtained. This is a useful tool to “rule in” and “rule out” a disease for an individual. By Bayes' theorem, the post-test probability is given by:

$$P\left(D\middle|T_{+/-}\right) = \frac{P(D)P\left(T_{+/-}\middle|D\right)}{P\left(T_{+/-}\middle|D\right)P(D) + P\left(T_{+/-}\middle|\overline{D}\right)P\left(\overline{D}\right)}$$

Since $P(D)$ can be converted into odds $O(D)$ using:

$$P(D) = \frac{O(D)}{1 + O(D)}$$

Hence, the post-test probability can be expressed as:

$$P\left(D\middle|T_{+/-}\right) = \frac{\frac{O(D)P\left(T_{+/-}\middle|D\right)}{1+O(D)}}{\frac{O(D)P\left(T_{+/-}\middle|D\right)}{1+O(D)} + \left(1 - \frac{O(D)}{1+O(D)}P\left(T_{+/-}\middle|\overline{D}\right)\right)}$$

$$\frac{O\left(D\middle|T_{+/-}\right)}{1 + O\left(D\middle|T_{+/-}\right)} = \frac{O(D)P\left(T_{+/-}\middle|D\right)}{O(D)\left[P\left(T_{+/-}\middle|D\right) + P\left(T_{+/-}\middle|\overline{D}\right)\right]}$$

$$\frac{O\left(D\middle|T_{+/-}\right)}{1 + O\left(D\middle|T_{+/-}\right)} = \frac{1}{1 + \left(\frac{P\left(T_{+/-}\middle|\overline{D}\right)}{O(D)P\left(T_{+/-}\middle|D\right)}\right)}$$

$$O\left(D\middle|T_{+/-}\right) = 1 + O\left(D\middle|T_{+/-}\right) - \left[O(D|T)\frac{P\left(T_{+/-}\middle|D\right)}{O(D)P\left(T_{+/-}\middle|D\right)}\right]$$

This is also known as the odds form (Aitken and Stoney, 1991):

$$O\left(D\middle|T_{+/-}\right) = O(D)\frac{P\left(T_{+/-}\middle|D\right)}{P\left(T_{+/-}\middle|\overline{D}\right)}$$

Hence, the integrated post-test odd after Tier 2, with pre-test odds obtained from Tier 1, is given by:

$$O_{\text{Tier 2}}\left(D\middle|T_{+/-}\right) = O_{\text{Tier 1}}(D) \cdot \Lambda_{\text{Tier 2}}\left(D\middle|T_{+/-}\right)$$

where $\Lambda_{\text{Tier 2}}\left(D\middle|T_{+/-}\right) = \frac{P\left(T_{+/-}\middle|D\right)}{P\left(T_{+/-}\middle|\overline{D}\right)}$ is the likelihood of a positive or negative test result from the model based on predictors at 15 weeks of gestation. They are given by $\Lambda_{\text{Tier 2}}(D|T_+) = \frac{r}{1-s}$ and $\Lambda_{\text{Tier 2}}(D|T_-) = \frac{1-r}{s}$.

If further prediction (or tiers) is needed, the sequential odds maybe calculated based on

pre-test odds obtained from Tier 1 and the likelihood of tests for all tiers. The final post-test odd is given by:

$$O_{\text{Tier 3}}\left(D\middle|T_{+/-}, T_{+/-}\right) = O_{\text{Tier 1}}(D) \cdot \Lambda\left(D\middle|T_{+/-_{\text{Tier 1}}} \cap T_{+/-_{\text{Tier 2}}}\right) \cdot \Lambda_{\text{Tier 2}}\left(D\middle|T_{+/-}\right)$$

One point to note is that the sequential Bayes has an assumption of conditional independence, in which the result of the sequential test must be independent of the previous test. The tiered modelling approach proposed in this study have satisfied this assumption, as individual models are trained separately, i.e. they are independent tests, and thus, the sensitivity and specificity of each model will not be affected by previous models.

5.2.2. Risk Classification

After the post-test odds for tier 2 is obtained, the predicted risk of all tiers will be analyzed together and classify the risk of disease in to 3 categories: low risk, moderate risk, and high risk (Fig. 5.2.2).

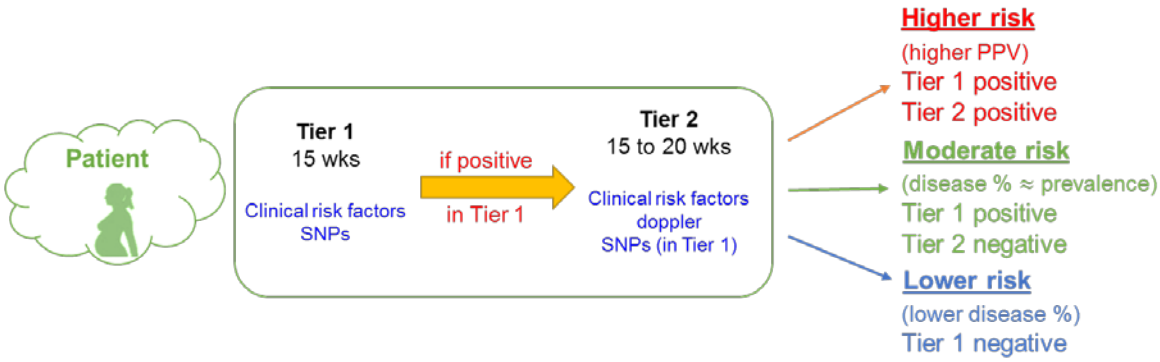

Fig. 5.2.2: Tiered model risk classification

The final risk is classified into three groups, where women with a negative result in Tier 1 considered to be at low risk, and those with positive result in Tier 1 but negative in Tier 2 considered to be at moderate risk. Women with a positive result in both Tier 1 and 2 are considered at high risk.

Women with a negative result at tier 1 will be considered as low risk, and do not need to go through further screening to tier 2. Since the sensitivity in tier 1 is high, the

likelihood of disease in women who are predicted at low risk will be relatively low. For women who are predicted at risk in tier 1, further screening through tier 2 is recommended to identify individuals who are at high risk. Since low-risk women are already "eliminated" in tier 1, the sensitivity threshold may be relaxed in tier 2 to aim for a higher positive predictive value. Therefore, individuals who may be at higher risk (i.e. those who have positive test result in both tier 1 and 2) may be further identified, amongst those who are predicted at risk.

As a result, the proportion of disease in the low-risk group (i.e. negative result in tier 1) will be lowest amongst the 3 risk groups, or at least lower than the current disease prevalence. Similarly, with a higher positive predictive value in the high-risk group, the proportion of disease will be highest, preferably more than 20% for rare diseases such as PE and SPTB. Hence, women with relatively lower risk are "eliminated" at each tier, and tailored care may be provided according to their classified predicted risk.

5.3. Process of Elimination

This section illustrates the use of tiered modelling in classifying predicted risk and identifying individuals at risk by the "process of elimination" using an example for SPTB prediction. After both individual models are obtained using penalized Logistic Regression, with Tier 1 at 15 weeks of gestation, and Tier 2 at 20 weeks' gestation.

A summary of the accuracy and predictive measures is shown in Table 5.3.1. The overall measure, shown on the last row, is the proportion of true positives and true negatives and is used as a measure to describe the efficiency of the model. The individual measures are results obtained separately from individual models, while the integrated measures of Tier 2 are the results obtained using post-test odds. Note that the

integrated odds of Tier 2 is used in the final risk classification instead of the individual model, as its properties and performance is not as biased compared to the individual model, after integrating the predicted risk from Tier 1.

Table 5.3.1: Pre-test and post-test accuracy measures

	Tier 1 (*15 weeks*)	Tier 2 (*20 weeks*)		Final
	Individual	Individual	Integrated	Classification
r	0.984	0.340	0.821	0.822
s	0.092	0.945	0.621	0.797
PPV	0.055	0.241	0.104	0.246
NPV	0.991	0.965	0.985	0.986
LR +	1.083	6.150	2.169	4.049
LR -	0.177	0.699	0.288	0.223
Overall	0.137	0.915	0.632	0.910

As discussed in Section 5.1, Tier 1 will have a high sensitivity, and in this example, a sensitivity of 98.4%.This will ensure that the majority of women who are potentially at risk would be identified by first screening at 15 weeks of gestation. By Tier 2, at 20 weeks' gestation, those who are predicted at risk will be screened based on new information available, such as transvaginal cervical length measurements. Although the overall efficiency is increased when PPV is higher in Tier 2, as expected, the performance of the integrated Tier 2 has decreased. However, the individual Tier 2 model was only intended to boost the PPV with a much more relaxed sensitivity threshold and is much more biased towards high specificity, and hence inappropriate to apply without integrating with the prior pre-test odds.

In the final classification, a sensitivity of 82.2% is achieved as a result of the high sensitivity in Tier 1, and since only those predicted at low-risk are "eliminated", the sensitivity level of the tiered model can be maintained at a higher level. Once the majority of women who may be at risk are identified, the next step is to identify those

at higher risk, amongst those who are predicted at risk. This is done by screening a model with higher PPV in Tier 2, which will identify a group of patients where the likelihood of SPTB is higher. In this example, with a PPV of 24.6%, nearly 1 in 4 predicted at high risk delivered preterm.

This tiered prediction system was tested on 1983 patients (Fig. 5.3.1). At first screening, during 15 weeks of gestation, 63% are eliminated as not at risk for SPTB, leaving only 37% (740 patients) of the 1983 patients needed to go through Tier 2 screening, and may be recommended a higher level of attention. By Tier 2 at 20 weeks of gestation, a further 29.5% are eliminated based on new predictors at this stage, with 142 identified at high risk of SPTB and may be monitored more closely. While the prediction at Tier 2 aims for a higher PPV, 1 in 5 patients identified at high risk and recommended for a higher level of care eventually delivered preterm.

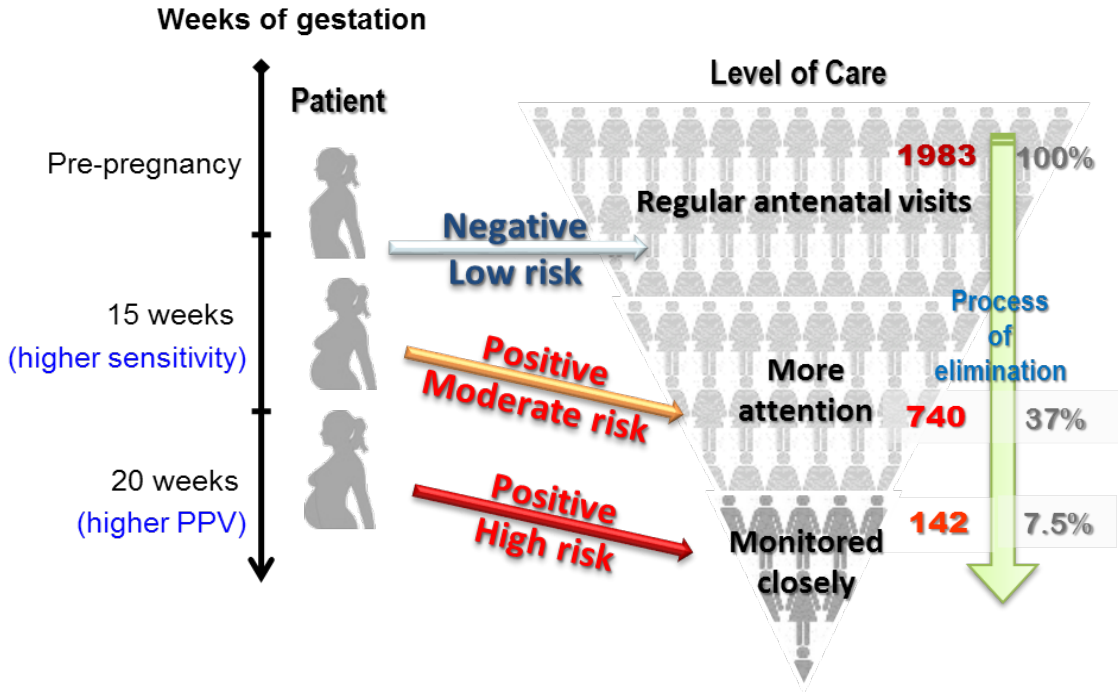

Fig. 5.3.1: Process of elimination

Women predicted to be at low risk are 'eliminated' at each Tier, and those predicted at higher risk could benefit from a higher level of care.

Interestingly, the proportion of SPTB amongst women who are predicted at low-risk is only 1.45%, with 18 SPTB cases missed out of 1243 patients predicted as low-risk in Tier 1 (Table 5.3.2). Although the proportion of SPTB in women who are predicted as moderate risk is similar to the current prevalence (5-10%), they may benefit from more frequent monitoring. A higher level of care, or intervention, may be provided for the 7.5% (142 patients) identified at high risk, in which the proportion and likelihood of SPTB is highest, and risk factors may also be addressed to reduce the severity of SPTB.

Table 5.3.2: Final risk classification

Tier 1	Tier 2	Predicted SPTB	SPTB case	% SPTB
+	+	142	35	24.65%
+	-	598	48	8.03%
-		1243	18	1.45%

Thus, through the process of elimination, 63% were successfully eliminated at Tier 1 by 15 weeks of gestation, with a sensitivity of above 80%, and a further 29.5% is eliminated, with a PPV of above 24% in those predicted as high risk. By integrating tiered models and classifying predicted risk, the efficiency of tailored care for individuals may be further enhanced, which may benefit the patient and reduce cost to community.

However, while the accuracy of those predicted as low risk and high risk is satisfactory, a major limitation is the uncertainty of those predicted as moderate risk. At this stage, patients predicted as moderate risk will rely on frequent monitoring, and further research will be needed to predict further information on this group of patients.

References

Aitken, C. G. and Stoney, D. A. (1991). *The use of statistics in Forensic Science*, Ellis Horwood.

Hall, G. H. (1967). "The clinical application of Bayes' theorem." *The Lancet* **290**(7515): 555-557.

Lindgaard, G., Pyper, C., Frize, M. and Walker, R. (2009). "Does Bayes have it? Decision Support Systems in diagnostic medicine." *International Journal of Industrial Ergonomics* **39**(3): 524-532.

Round, A. (2001). "Introduction to clinical reasoning." *Journal of Evaluation in Clinical Practice* **7**(2): 109-117.

Sox, H. C., Higgins, M. C. and Owens, D. K. (2013). *Medical Decision Making*, John Wiley & Sons, Ltd.

Chapter 6: Manuscript I – Interaction Effect of Marijuana

Maternal marijuana use has independent effects on risk for spontaneous preterm birth but not other common late pregnancy complications

Shalem Yiner-Lee Leemaqz[1], MAppStat

Gustaaf A Dekker[1], MD PhD

Louise C Kenny[2], PhD

Jenny E Myers[3], PhD

Nigel A B Simpson[4], PhD

Lesley M McCowan[5], PhD

Lucilla Poston[6], PhD

Claire T Roberts[1]*, PhD

On behalf of the SCOPE Consortium

[1]Robinson Research Institute, School of Paediatrics and Reproductive Health, University of Adelaide, Adelaide, South Australia, 5005;

[2]The Irish Centre for Fetal and Neonatal Translational Research (INFANT), University College Cork, Ireland;

[3]Maternal & Fetal Heath Research Centre, Manchester Academic Health Science Centre, University of Manchester, Central Manchester NHS Trust, Manchester, United Kingdom;

[4]Department of Women's & Children's Health, University of Leeds, Leeds, United Kingdom;

[5]Department of Obstetrics and Gynaecology, University of Auckland, Auckland, New Zealand;

[6]Division of Women's Health, King's College London, London, United Kingdom.

* Correspondence to: Claire Roberts claire.roberts@adelaide.edu.au

Word count: 3757

Abstract

Importance. Maternal marijuana use is a major contributing risk factor for SPTB, but not for PE, GHT, GDM nor SGA, , independent of both cigarette smoking status and SEI.

Objective. This study investigated the association and interaction of marijuana use three months prior to and during pregnancy, with maternal age, BMI, socioeconomic index (SEI), cigarette smoking status, and pregnancy outcomes.

Design. A prospective cohort from the Screening fOr Pregnancy Endpoints (SCOPE) study between November 2004 and February 2011 in Australia, New Zealand, Ireland and the United Kingdom.

Setting. Nulliparous women with singleton pregnancies were recruited. Women were invited to participate prior to 15 weeks' gestation when attending hospital antenatal clinics, obstetricians, general practitioners or community midwives, and were interviewed and examined by a research midwife at 15±1 and 20±1 weeks of gestation.

Participants. A total of 5989 participants from Australia, New Zealand, United Kingdom, and Ireland agreed to participate, with 5588 participants included in analysis after excluding women with late miscarriages or terminations.

Main Outcome(s) and Measure(s): Cases (278 Preeclampsia (PE) , 470 gestational hypertension (GHT), 633 small for gestational age (SGA), 236 spontaneous preterm births (SPTB),143 gestational diabetes (GDM)), were compared separately with 4114 non-cases.

Marijuana use and smoking status at 15 and 20 weeks' gestation were recorded, along with age, BMI, and socio-economic index (SEI).

Results. Continued maternal use of marijuana at 20 weeks' gestation was associated with SPTB. The effect of marijuana use on SPTB was independent of cigarette smoking status [odds adjusted for cigarette smoking only 2.28 (95% confidence interval 1.45, 3.59)] and adjusted for SEI only [2.17 (1.41, 3.34)]. When adjusted for age, smoking, and SEI in multivariate analysis, maternal marijuana had a greater magnitude of effect if still used at 20 weeks' gestation [5.18 (2.32, 11.54)].

Conclusions and Relevance. Continued use of marijuana, especially after 20 weeks' gestation, increases the risk of SPTB, and its effect is independent of smoking status. Given moves to decriminalise marijuana and its increasing use among women of reproductive age, and in pregnant women to reduce nausea, this is of major concern to public health.

Trial registration. ACTRN12607000551493.

Keywords: *Marijuana, smoking, BMI, pregnancy outcome, spontaneous preterm birth.*

Key Messages

- Marijuana increases the risk of spontaneous preterm birth independent of cigarette smoking status and socio-economic status
- Women who continue to use marijuana at 20 weeks' gestation are five times more likely to deliver preterm than those who do not
- The rate of early SPTB is higher amongst women who continue to use marijuana at 20 weeks’ gestation

Introduction

The continuing rise in marijuana consumption over recent decades has raised concerns about the impact of exposure to marijuana amongst women of reproductive age and its effects on pregnancy outcomes[1]. The recent legalization in many Western countries (e.g. USA) potentially adds to this concern. According to the National Drug Strategy Household Survey[2] in Australia, 7.7% of females aged ≥ 14 years used marijuana during 2010 (1.1% increase compared to 2007), with 32% of the female population having been exposed to marijuana at least once in their lifetime. A similar trend has also been observed in New Zealand and Europe, with 47.2% of women aged ≥ 16 years in NZ (from 2007 to 2008)[3,4], 24.6% in the United Kingdom and 17.5% in Ireland having been exposed at least once[5].

Apart from reported negative impacts on fetal growth and brain development[6-10], marijuana has been associated with adverse pregnancy outcomes, including preterm birth (PTB), small for gestational age (SGA), placental abruption and antepartum haemorrhage[11-15]. Studies have shown that exposure to marijuana during pregnancy is associated with low birthweight and increases the risk of PTB and SGA, with an odds ratio of at least 1.5 when adjusted for age, BMI and smoking[11,14-16].

The association between marijuana use and pregnancy outcomes is often confounded by other known risk factors including cigarette smoking, BMI, and SEI[17,18]. Women who use marijuana also tend to smoke cigarettes and are more likely to use other drugs and alcohol. National

statistics have shown that, amongst Australian women aged ≥ 14 years who used marijuana in 2010, 82.7% also consumed alcohol, and 68.5% were cigarette smokers, with similar patterns of prevalence in New Zealand[2,4].

However, there have been inconsistent results reported from American prospective cohort studies, in which associations with marijuana use were not found [19-21]. Hence, this study aimed to examine the association of maternal marijuana use (from pre-pregnancy and up to 20 weeks' gestation) in a multi-centre cohort with major pregnancy complications, amongst both cigarette smokers and non-smokers, controlling for well-known risk factors including age, SEI and BMI, as well as its effects on length of gestation.

Although maternal alcohol consumption is also considered as a confounder, previous results from the SCOPE study have shown that alcohol consumption during early pregnancy was not associated with adverse pregnancy outcomes[22]. Hence, the analysis in this study did not include alcohol consumption.

Methods

Data from this analysis were obtained from the Screening fOr Pregnancy Endpoints (SCOPE) study, which aims at building a clinical database and pregnancy biobank to screen candidate markers of pregnancy complications. The SCOPE study recruited nulliparous women with singleton pregnancies between November 2004 and February 2011 in Australia, New Zealand,

Ireland and the United Kingdom. Ethical approval was obtained from local ethics committees [New Zealand AKX/02/00/364, Australia REC 1712/5/2008, London, Leeds and Manchester 06/MRE01/98 and Cork ECM5 (10) 05/02/08] and all women provided written informed consent. Women were invited to participate prior to 15 weeks' gestation when attending hospital antenatal clinics, obstetricians, general practitioners or community midwives, and were interviewed and examined by a research midwife at 15±1 and 20±1 weeks of gestation.

The exclusion criteria included women who were considered to be at high risk of PE , SGA or PTB due to underlying medical conditions (e.g. chronic hypertension requiring antihypertensive medication or diabetes), previous cervical knife cone biopsy, three terminations or three miscarriages or if their pregnancy was complicated by a known major fetal anomaly or abnormal karyotype, or if they received interventions that may modify pregnancy outcome (e.g. aspirin, cervical suture).

Details of maternal age, BMI and socioeconomic index (SEI), medical and family history, along with dietary and lifestyle questionnaires with self-reported marijuana and cigarette smoking were recorded at 15 weeks' and 20 weeks' gestation and entered into an internet-accessed, password-protected centralised database with a complete audit trail (MedSciNet^{AB}, Stockholm, Sweden)[23].

The number of episodes of marijuana use over 3 months was also recorded at 15 weeks and

20 weeks of gestation. Although other drug use was also recorded, including cocaine, amphetamines, substance P, Ecstasy, opiates, and hallucinogens, with less than 0.6% of women who have taken these drugs 3 months prior to or during pregnancy in SCOPE, there are insufficient data to be included for analysis.

Marijuana and cigarette smoking status were classified into five categories (i.e. never, quit prior to pregnancy, quit prior to 15 weeks' gestation, still using at 15 weeks' gestation, and still using at 20 weeks' gestation) in univariate and multivariable analysis, with ‘non-smoking’ or ‘never used marijuana’ as the reference categories. The number of episodes of marijuana use was included as a continuous variable for dose effect estimation.

Spontaneous preterm birth (SPTB) was defined as birth at less than 37 weeks of gestation that was not a result of medical or obstetric intervention. Small for gestational age (SGA) was defined as a birthweight of less than the 10th customised centile, adjusted for maternal height, weight, parity, ethnicity, gestational age at delivery and infant sex. Preeclampsia (PE) was defined as gestational hypertension (GHT) (blood pressure of 140/90 or greater on at least 2 occasions 4 hours apart after 20 weeks' gestation) accompanied by proteinuria (300 mg/day or greater, or a spot protein creatinine ratio of 30 mg/mmol creatinine or greater). Gestational diabetes mellitus (GDM) was defined as a fasting glucose of 5.5 mmol/L or higher in a Glucose Tolerance Test, a 2 hr level of 8 mmol or higher, or a random glucose level of 11 nmol/L or higher. Universal screening was not employed for GDM in the UK and Ireland.

Statistical analysis

A total of 5588 participants were included in the analysis, with 1155 participants recruited from Australia, 2014 from New Zealand, 1765 from Ireland, and 654 from the United Kingdom. Within the 1514 pregnancies with complications, 278 had PE, 633 had SGA, 236 had SPTB, 470 had GHT, and 143 had GDM (Figure 1). Details on age, BMI, SEI, as well as marijuana use and cigarette smoking status were complete for all participants.

Marijuana and cigarette smoking status were compared between non-cases and each of the outcomes separately using Fisher's exact test. Although women may have had more than one pregnancy complication, each outcome was analysed separately compared with non-cases. Continuous factors, including maternal age, BMI and SEI were compared using Student's *t* test.

To investigate the effects of marijuana use between smokers and non-smokers, analysis of marijuana use stratified by cigarette smoking status for each outcome was performed. Breslow-Day test was used to assess the homogeneity of the odds of marijuana use between cigarette smokers and non-smokers, along with an adjusted common odds estimated from Mantel-Haenszel test[24,25].

Marijuana and cigarette smoking status were then analysed with mixed effects logistic regression to determine the association with pregnancy outcomes, adjusting for maternal age,

BMI and SEI, and with population differences as a random effect. Interaction tests were also performed by comparing logistic regression models that included interaction terms. A linear mixed model was also fitted for length of gestation, with quadratic terms for the number of marijuana used over 3 months at 15 and 20 weeks of gestation, age, and BMI, to investigate the dose effect of marijuana and cigarette smoking status on the length of gestation adjusted for other factors in the model. The estimated power of this analysis, involving logistic regression with interaction terms, is 0.99 [26]. All statistical analyses were performed using R version 3.2.0.

Results

Of the 5588 participants, the overall proportion of women reporting the use of marijuana before or during pregnancy was 5.6%, with the participating centre in Australia having the highest rate of women using marijuana (11.6%), followed by New Zealand (4.5%), Ireland (3.8%), and United Kingdom (3.7%). Compared to marijuana use, the proportion of cigarette smokers was higher, with an overall 26.4% of women who smoked cigarettes before or during pregnancy. Amongst Australian participants, 40.8% were cigarette smokers at conception with 29.7% of Irish, 29.5% of UK and 14.2% of NZ participants. Country specific demographics are shown in Table 1.

The overall characteristics comparing each of the pregnancy outcomes to non-cases are shown in Table 2. There were significant differences in the average BMI and SEI between non-

cases and all outcomes analysed, where BMI was higher in women who developed either PE (27.8 ± 0.38 vs 24.8 ± 0.07 in non-cases; P<0.001), GHT (27.9 ± 0.27; P<0.001), GDM (29.1 ± 0.52; P<0.001), SGA (P<0.001) or SPTB (25.9 ± 0.22; P=0.035). Similarly, SEI was lower on average in women with complicated pregnancies including PE (38 ± 0.93 vs 42.5 ± 0.26 in non-cases; P<0.001), GHT (39.7 ± 0.76; P=0.001), GDM (38.9 ± 1.36; P=0.011), and SGA (40.1 ± 0.64; P=0.001). Women who developed PE were also slightly younger on average (27.7 ± 0.34 vs 28.7 ± 0.09 in non-cases; P=0.002), while patients who developed GDM were older (30 ± 0.44; P=0.008).

Marijuana use and cigarette smoking at 20 weeks of gestation were both associated with SGA (18.6% smoking vs only 8.9% in non-cases; P<0.001, and 1.9% marijuana use vs 0.7% in non-cases; P<0.005) and SPTB (16.1% smoking vs 8.9% in non-cases; P=0.001, and 4.7% marijuana use vs 0.7% in non-cases ; P<0.001). For both of these outcomes, there was a higher proportion of women who continued to smoke cigarettes or use marijuana at 20 weeks' gestation. In women who delivered a SGA infant, 18.6% continued to smoke cigarettes (compared to 8.9% in non-cases) and 1.9% continued to use marijuana (compared to 0.7% in non-cases), while in women who delivered preterm, 16.1% continued to smoke cigarettes and 4.7% continued to use marijuana at 20 weeks’ gestation.

Furthermore, in the whole cohort, the average gestational age at delivery was lower in women who continued to use marijuana at 20 weeks’ gestation compared to non-users (37.4 ± 0.7

weeks vs 39.6 ± 0.3 weeks in non-users; P<0.001), with 15.1% delivering at less than 32 weeks of gestation (Table 3). Similarly, amongst women with SPTB, those who continued to use marijuana at 20 weeks' gestation had a significantly shorter gestation on average of 29.6±1.6 weeks, compared to 34.1±0.3 weeks in those who did not use marijuana (P=0.005) (Table 4). The proportion of very early SPTB was also higher, with 36.4% having delivered at less than 28 weeks of gestation and 63.6% at less than 32 weeks in women who continued to use marijuana at 20 weeks' gestation, compared to 4.7% and 15.8% amongst non-users.

When assessing the proportion of SPTB amongst women who used marijuana, 11.0% of women who used marijuana in the 3 months prior to or during pregnancy delivered preterm, compared to 5.1% in non-cases (P<0.001). In particular, women who continued to use marijuana at 20 weeks' gestation were at a markedly higher risk of SPTB (adjusted OR 5.18; CI 2.32 to 11.54; P<0.001) than those who did not use marijuana (Table 5).

Interaction between Maternal Marijuana use and Cigarette Smoking

When comparing any marijuana use, three months prior to or during pregnancy, between cigarette smokers and non-smokers, there was a significant independent association between any marijuana use and SPTB (P=0.001). Breslow-Day test showed no evidence of heterogeneity in the association of marijuana use and pregnancy outcomes between smokers and non-smokers (P=0.238), which indicates that the association between marijuana and

SPTB was consistent regardless of cigarette smoking status.

While the association between marijuana use and SPTB was independent of smoking status, the Mantel-Haenszel test (Table 6) further indicated that the overall association was also significant (P<0.001), with an adjusted common odds of 2.28 (95% CI 1.45 to 3.59). That is, the odds of SPTB for any marijuana use three months prior to or during pregnancy was more than doubled for both cigarette smokers and non-smokers.

Regarding the interaction effect of marijuana in women who ceased cigarette smoking during pregnancy, results from Breslow-Day test on the homogeneity of the odds of any marijuana use (three months prior to or during pregnancy), between women who continued cigarette smoking before 20 weeks' gestation and those who stopped smoking, showed no evidence of heterogeneity (P=0.541), with a Mantel-Haenszel adjusted odds of 1.97 (CI 1.26 to 3.09; P=0.004). This indicated that the effect of marijuana use was not only independent of any cigarette smoking three months prior to or during pregnancy (as reported above), but was also consistent, with nearly doubled odds, irrespective of whether cigarette smoking ceased prior to 20 weeks' gestation.

Results from Logistic regression with an interaction term between marijuana use and cigarette smoking status also showed no significant interaction effects on SPTB (P=0.723).

Interaction between Maternal Marijuana use and Low Socio-economic Status

Interaction between marijuana use and socio-economic status was also tested, and no significant interaction effect was seen for all pregnancy complications analysed, when added as an interaction term in multivariable models. When comparing low socio-economic status, in the lower quartile (SEI <28), with any marijuana use, Breslow-Day test also showed no evidence of heterogeneity (P=0.656), indicating that the marijuana association with SPTB was also independent of socio-economic status (adjusted odds 2.17; 95% CI 1.41 to 3.34; P=0.001).

Estimated Risk

In logistic regression models controlling for maternal age, SEI and smoking (Table 5), continued use of marijuana at 20 weeks' gestation was a significant risk factor for SPTB (OR 5.13; CI 2.30 to 11.43; P<0.001), but not for any other outcomes analysed. Similarly, as expected, continuing to smoke cigarettes at 20 weeks' gestation was associated with SGA, with an adjusted odds of 3.46 (CI 1.31 to 9.10; P=0.012).

BMI was a significant risk factor for most outcomes (P<0.001) except SPTB (P=0.062). By contrast, age was not a significant factor for most pregnancy outcomes assessed except for GDM (OR 1.08; CI 1.04 to 1.12; P<0.001) and SGA (OR 1.02; CI 1.00 to 1.04; P=0.010). Consistent with previous studies, higher SEI was a protective factor for PE (P=0.023), with an estimated 1-2% decrease in risk for every unit increase in SEI.

Effect on Length of Gestation

The results from linear mixed modelling showed that marijuana use in first (P=0.000) or second trimester (P=0.002) had significant effects on length of gestation, when adjusted for age, BMI, SEI, and cigarette smoking status. The predicted length of gestation (Figure 2) was lower for women who continued to use marijuana at 20 weeks of gestation for both smokers and non-smokers, with an estimated gestation of less than 37 weeks when more than 100 episodes of marijuana use within the last 12 weeks before 20 weeks' gestation (i.e. more than once per day for the preceding 3 months).

It is interesting to note that there was a slight decrease in the predicted length of gestation for smokers compared to non-smokers (P=0.003 at 15 weeks' gestation, and P=0.020 at 20 weeks' gestation). However, the difference between marijuana use at 15 and 20 weeks of gestation appears to be greater than the effect of cigarette smoking status.

Discussion

Marijuana use is increasing in women of reproductive age and its continued use in pregnancy has been of concern for some time[6]. In addition, there is evidence to suggest that some pregnant women are using marijuana to reduce nausea in early pregnancy[27]. In this large

prospective cohort of nulliparous women we have demonstrated that continued maternal use of marijuana at 20 weeks' gestation is a major contributing risk factor for SPTB. Univariate analysis showed a significant association of marijuana use at 20 weeks' gestation with SGA and SPTB, but when adjusted for other factors, in particular cigarette smoking, marijuana use still represented a significant independent risk factor for SPTB. Furthermore, if marijuana use was continued at 20 weeks' gestation, women were over five times more likely to deliver preterm than nonusers. Of the women who delivered preterm and who also continued to use marijuana at 20 weeks' gestation nearly 64% delivered at less than 32 weeks' gestation. Our data do not have sufficient power to determine a gestational age prior to 20 weeks when it is safe to cease marijuana use. Hence, at this stage we cannot comment on its safety in early pregnancy but despite this lack of evidence, it would be prudent to abstain from marijuana use during pregnancy.

Based on the current findings and some earlier reports[11,16,18,28,29], it is likely that maternal marijuana use is an independent risk factor for SPTB. It has been shown that the active compound of marijuana (δ9-tetrahydrocannabinol) and its metabolites are able to cross the placental barrier and thereby have the potential to directly affect perinatal outcomes[30,31]. Whereas the results from this study are in agreement with other studies, it needs to be noted that a few American prospective cohort studies did not find an association between marijuana use and SPTB[19-21]. Nevertheless, the data from this study are from a large prospective cohort,

and all data were obtained during face-to-face contact between dedicated research midwives and patients.

While African American ethnicity has been associated with an increased risk of SPTB [32,33], it has also been commonly associated with lower socio-economic status. The relationship of low SEI with pregnancy complications was apparent in this study, where SEI was significantly negatively associated with PE, GHT, GDM, SGA, and SPTB. When adjusted for age, BMI, cigarette smoking, and marijuana use, higher SEI was a protective factor, where per unit increase in SEI had a 1 to 2% decrease in the risk of PE. Similar trends were also seen in previously published SCOPE data [34,35]. However, the results from the current study showed no significant interaction effects between marijuana use and SEI, suggesting that the association between marijuana use and SPTB was also independent of socio-economic status.

Maternal cigarette smoking is typically considered to be a risk factor for SPTB and SGA [36-41]. Indeed, maternal cigarette smoking at 20 weeks' gestation was significant for SPTB and SGA in univariate tests, but no longer significant for SPTB when adjusted for other factors. Similar results have been found previously in a study by Dekker et al. [42], which incorporated multiple novel risk factors for SPTB. In the current study an association was seen between smoking and SPTB (in univariate analysis), but cigarette smoking was not found to be an independent risk factor for SPTB after adjustment for marijuana use. Nevertheless, continued cigarette smoking is a significant risk factor for multiple pregnancy complications including stillbirth and

SGA and women should be encouraged to quit.

The association between smoking and marijuana is often considered as an interaction effect for pregnancy complications, as the majority of women who use marijuana also smoke cigarettes [2,4,43]. In fact, amongst women who used marijuana in the SCOPE cohort, 74% also smoked cigarettes. With a high concurrence rate, the independent effect of marijuana on pregnancy outcomes has generally been unrecognised and just considered to be subsidiary, partly due to the low availability of data on marijuana use compared to cigarette smoking for statistical analysis [28,43]. However, our data from the SCOPE cohort, with 316 participants (5.62%) who were marijuana users, demonstrate that the association of marijuana use with SPTB is consistent across cigarette smokers and non-smokers.

The consistent effect of marijuana use is also apparent when analysing the effect of number of episodes of marijuana use during pregnancy on the length of gestation. While there was a slight decrease in the predicted length of gestation amongst smokers, the trend for smokers and non-smokers was similar. In contrast, the predicted length of gestation for women who continued to use marijuana at 20 weeks' gestation was significantly decreased compared to those who ceased earlier in gestation, regardless of smoking status. This is consistent with similar studies which showed that marijuana use is associated with a decreased length of gestation [44,45].

Furthermore, apart from a cigarette smoking-marijuana interaction, it is also well recognised that cigarette smoking and illicit drug use are associated with low socio-economic status [43,46-48], along with a complex inter-relationship with obesity, where smoking cessation may also lead to obesity [43,49-51]. As described in many studies, the prevalence of cigarette smoking and obesity is higher amongst those who are socio-economically disadvantaged, and the incidence of SPTB is higher amongst women with lower income and lower educational status, with a previously published estimated odds ratio of 2.73 for mothers who had less than five years of education [32,52], which may indicate associations with other lifestyle risk factors.

Furthermore, if there was no maternal marijuana exposure, with an estimated population attributable risk of 0.003 for marijuana use, the incidence of SPTB would be expected to decrease by 3 cases per 1000 pregnant women. With a rate of SPTB of 4.2% in this study, this represents an estimated 6.2% reduction in the incidence of SPTB in the population, i.e. about 3 out of 50 SPTB cases would be attributed to marijuana use.

Strengths and Limitations

A major strength of this study was its large international multicentre prospective cohort with excellent follow-up and complete data available for this analysis. Women were recruited from a clearly defined population of nulliparous women, with meticulous data monitoring protocols to reduce data entry or transcription errors and ensure the quality of data.

Conclusion

In this large prospective cohort, maternal marijuana use had a major contribution to SPTB and this association was consistent for both cigarette smokers and non-smokers, with doubled odds in women who used marijuana three months prior to or during pregnancy. For women who use marijuana during pregnancy, it should be emphasised that stopping early in pregnancy should be encouraged since continued use of marijuana at 20 weeks of gestation was associated with a five-fold increased risk of SPTB in this study following adjustment for other confounders, including maternal age, BMI, SEI, and cigarette smoking. There would be an estimated 6.2% reduction in the incidence of SPTB if women were not exposed to marijuana during pregnancy.

Preterm birth is increasing in developed nations, with attendant increases in adverse infant outcomes, as well as psychological and social impacts, and is of great concern to public health. The increasing exposure to marijuana in women of reproductive age and its contribution to the risk for preterm birth make it a modifiable target for intervention. In nations where authorities are considering decriminalisation of marijuana, the risks to pregnant women and their babies need much greater consideration.

Acknowledgements

The authors wish to thank all of the SCOPE participants and the SCOPE research midwives

in each centre. The SCOPE database is provided and maintained by MedSciNet AB.

Funding

The Australian SCOPE study was supported by the Premier's Science and Research Fund, South Australian Government. The New Zealand SCOPE study was funded by the New Enterprise Research Fund, Foundation for Research Science and Technology; Health Research Council [04/198]; Evelyn Bond Fund, Auckland District Health Board Charitable Trust. The Irish SCOPE study was funded by the Health Research Board of Ireland [CSA/2007/2]. The UK SCOPE study was funded by National Health Service NEAT Grant [Neat Grant FSD025], Biotechnology and Biological Sciences Research council [GT084] and University of Manchester Proof of Concept Funding (University of Manchester); Guy's and St. Thomas' Charity (King's College London) and Tommy's charity (King's College London and University of Manchester); and Cerebra UK (University of Leeds). The funders had no role in study design, data collection, and analysis, decision to publish or preparation of the manuscript.

References

1. Narkowicz S, Plotka J, Polkowska Z, Biziuk M, Namiesnik J. Prenatal exposure to substance of abuse: a worldwide problem. *Environ Int.* 2013;54:141-163.
2. Australian Institute of Health and Welfare. 2010 National Drug Strategy Household Survey report. Vol 25. Canberra: AIHW; 2011.
3. Ministry of Health. New Zealand Drug Statistics. Wellington: New Zealand Health Information Service; 2001.
4. Ministry of Health. Drug Use in New Zealand: Key Results of the 2007/08 New Zealand Alcohol and Drug Use Survey. Wellington: Ministry of Health; 2010.
5. European Monitoring Centre for Drugs and Drug Addiction. European Drug Report: Trends and developments. Spain: Publications Office of the European Union; 2013.
6. Day NL, Richardson GA, Geva D, Robles N. Alcohol, marijuana, and tobacco: effects of prenatal exposure on offspring growth and morphology at age six. *Alcoholism, clinical and experimental research.* 1994;18(4):786-794.
7. Eyler FD, Behnke M, Conlon M, Woods NS, Wobie K. Birth outcome from a prospective, matched study of prenatal crack/cocaine use: I. Interactive and dose effects on health and growth. *Pediatrics.* 1998;101(2):229-237.
8. Wang X, Dow-Edwards D, Anderson V, Minkoff H, Hurd YL. Discrete opioid gene expression impairment in the human fetal brain associated with maternal marijuana use. *Pharmacogenomics J.* 2006;6(4):255-264.
9. Hingson R, Alpert JJ, Day N, et al. Effects of maternal drinking and marijuana use on fetal growth and development. *Pediatrics.* 1982;70(4):539-546.
10. Faden VB, Graubard BI. Maternal substance use during pregnancy and developmental outcome at age three. *J Subst Abuse.* 2000;12(4):329-340.
11. Black M, Bhattacharya S, Fairley T, Campbell DM, Shetty A. Outcomes of pregnancy in women using illegal drugs and in women who smoke cigarettes. *Acta obstetricia et gynecologica Scandinavica.* 2013;92(1):47-52.
12. Gibson GT, Baghurst PA, Colley DP. Maternal alcohol, tobacco and cannabis consumption and the outcome of pregnancy. *The Australian & New Zealand journal of obstetrics & gynaecology.* 1983;23(1):15-19.
13. Linn S, Schoenbaum SC, Monson RR, Rosner R, Stubblefield PC, Ryan KJ. The association of marijuana use with outcome of pregnancy. *American journal of public health.* 1983;73(10):1161-1164.
14. Kennare R, Heard A, Chan A. Substance use during pregnancy: risk factors and obstetric and perinatal outcomes in South Australia. *The Australian & New Zealand journal of obstetrics & gynaecology.* 2005;45(3):220-225.
15. Hatch EE, Bracken MB. Effect of marijuana use in pregnancy on fetal growth. *American journal of epidemiology.* 1986;124(6):986-993.

16. Hayatbakhsh MR, Flenady VJ, Gibbons KS, et al. Birth outcomes associated with cannabis use before and during pregnancy. *Pediatric research.* 2012;71(2):215-219.
17. Brown HL, Graves CR. Smoking and marijuana use in pregnancy. *Clinical obstetrics and gynecology.* 2013;56(1):107-113.
18. Kozer E, Koren G. Effects of prenatal exposure to marijuana. *Canadian family physician Medecin de famille canadien.* 2001;47:263-264.
19. van Gelder MM, Reefhuis J, Caton AR, Werler MM, Druschel CM, Roeleveld N. Characteristics of pregnant illicit drug users and associations between cannabis use and perinatal outcome in a population-based study. *Drug and alcohol dependence.* 2010;109(1-3):243-247.
20. Bada HS, Das A, Bauer CR, et al. Low birth weight and preterm births: etiologic fraction attributable to prenatal drug exposure. *Journal of perinatology : official journal of the California Perinatal Association.* 2005;25(10):631-637.
21. Shiono PH, Klebanoff MA, Nugent RP, et al. The impact of cocaine and marijuana use on low birth weight and preterm birth: A multicenter study. *American journal of obstetrics and gynecology.* 1995;172(1, Part 1):19-27.
22. McCarthy FP, O'Keeffe LM, Khashan AS, et al. Association between maternal alcohol consumption in early pregnancy and pregnancy outcomes. *Obstetrics and gynecology.* 2013;122(4):830-837.
23. McCowan L, North R, Taylor R. ACTRN12607000551493. 2007; www.anzctr.org.autrialSearch.aspx.
24. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute.* 1959;22(4):719-748.
25. Liu IM. Breslow–Day Test. *Encyclopedia of Biostatistics*: John Wiley & Sons, Ltd; 2005.
26. Demidenko E. Sample size and optimal design for logistic regression with binary interaction. *Statistics in medicine.* 2008;27(1):36-46.
27. Westfall RE, Janssen PA, Lucas P, Capler R. Survey of medicinal cannabis use among childbearing women: patterns of its use in pregnancy and retroactive self-assessment of its efficacy against 'morning sickness'. *Complement Ther Clin Pract.* 2006;12(1):27-33.
28. Kesmodel U, Kesmodel PS, Larsen A, Secher NJ. Use of alcohol and illicit drugs among pregnant Danish women, 1998. *Scand J Public Health.* 2003;31(1):5-11.
29. El Marroun H, Tiemeier H, Steegers EA, et al. Intrauterine cannabis exposure affects fetal growth trajectories: the Generation R Study. *Journal of the American Academy of Child and Adolescent Psychiatry.* 2009;48(12):1173-1181.
30. Huizink AC. Prenatal cannabis exposure and infant outcomes: overview of studies. *Progress in neuro-psychopharmacology & biological psychiatry.* 2014;52:45-52.
31. Fergusson DM, Horwood LJ, Northstone K, Team AS. Maternal use of cannabis and

pregnancy outcome. *BJOG : an international journal of obstetrics and gynaecology.* 2002;109(1):21-27.

32. Murphy DJ. Epidemiology and environmental factors in preterm labour. *Best Practice & Research Clinical Obstetrics & Gynaecology.* 2007;21(5):773-789.
33. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. *Lancet.* 2008;371(9606):75-84.
34. Myers JE, Kenny LC, McCowan LM, et al. Angiogenic factors combined with clinical risk factors to predict preterm pre-eclampsia in nulliparous women: a predictive test accuracy study. *BJOG : an international journal of obstetrics and gynaecology.* 2013;120(10):1215-1223.
35. North RA, McCowan LME, Dekker GA, et al. Clinical risk prediction for pre-eclampsia in nulliparous women: development of model in international prospective cohort. *BMJ (Clinical research ed.).* 2011;342:d1875.
36. Braillon A, Bewley S. The enigma of spontaneous preterm birth. *The New England journal of medicine.* 2010;362(21):2032; author reply 2033-2034.
37. Ananth CV, Savitz DA, Luther ER. Maternal cigarette smoking as a risk factor for placental abruption, placenta previa, and uterine bleeding in pregnancy. *American journal of epidemiology.* 1996;144(9):881-889.
38. Tsai HJ, Liu X, Mestan K, et al. Maternal cigarette smoking, metabolic gene polymorphisms, and preterm delivery: new insights on GxE interactions and pathogenic pathways. *Human genetics.* 2008;123(4):359-369.
39. Zhang J, Klebanoff MA, Levine RJ, Puri M, Moyer P. The puzzling association between smoking and hypertension during pregnancy. *American journal of obstetrics and gynecology.* 1999;181(6):1407-1413.
40. Conde-Agudelo A, Althabe F, Belizan JM, Kafury-Goeta AC. Cigarette smoking during pregnancy and risk of preeclampsia: a systematic review. *American journal of obstetrics and gynecology.* 1999;181(4):1026-1035.
41. McCowan LM, Thompson JM, Taylor RS, et al. Clinical prediction in early pregnancy of infants small for gestational age by customised birthweight centiles: findings from a healthy nulliparous cohort. *PloS one.* 2013;8(8):e70917.
42. Dekker GA, Lee SY, North RA, McCowan LM, Simpson NA, Roberts CT. Risk factors for preterm birth in an international prospective cohort of nulliparous women. *PloS one.* 2012;7(7):e39154.
43. Passey ME, Sanson-Fisher RW, D'Este CA, Stirling JM. Tobacco, alcohol and cannabis use during pregnancy: clustering of risks. *Drug and alcohol dependence.* 2014;134:44-50.
44. Fried PA, Watkinson B, Willan A. Marijuana use during pregnancy and decreased length of gestation. *American journal of obstetrics and gynecology.* 1984;150(1):23-27.

45. Janisse JJ, Bailey BA, Ager J, Sokol RJ. Alcohol, tobacco, cocaine, and marijuana use: relative contributions to preterm delivery and fetal growth restriction. *Subst Abus.* 2014;35(1):60-67.
46. Tehranifar P, Liao Y, Ferris JS, Terry MB. Life course socioeconomic conditions, passive tobacco exposures and cigarette smoking in a multiethnic birth cohort of U.S. women. *Cancer causes & control : CCC.* 2009;20(6):867-876.
47. Graham H, Hawkins SS, Law C. Lifecourse influences on women's smoking before, during and after pregnancy. *Soc Sci Med.* 2010;70(4):582-587.
48. Kendzor DE, Businelle MS, Cofta-Woerpel LM, et al. Mechanisms linking socioeconomic disadvantage and BMI in smokers. *American journal of health behavior.* 2013;37(5):587-598.
49. Koster A, Leitzmann MF, Schatzkin A, et al. The combined relations of adiposity and smoking on mortality. *The American journal of clinical nutrition.* 2008;88(5):1206-1212.
50. Salonen MK, Kajantie E, Osmond C, et al. Role of socioeconomic indicators on development of obesity from a life course perspective. *J Environ Public Health.* 2009;2009:625168.
51. Chiolero A, Faeh D, Paccaud F, Cornuz J. Consequences of smoking for body weight, body fat distribution, and insulin resistance. *The American journal of clinical nutrition.* 2008;87(4):801-809.
52. Begum F, Buckshe K, Pande JN. Risk factors associated with preterm labour. *Bangladesh Medical Research Council bulletin.* 2003;29(2):59-66.

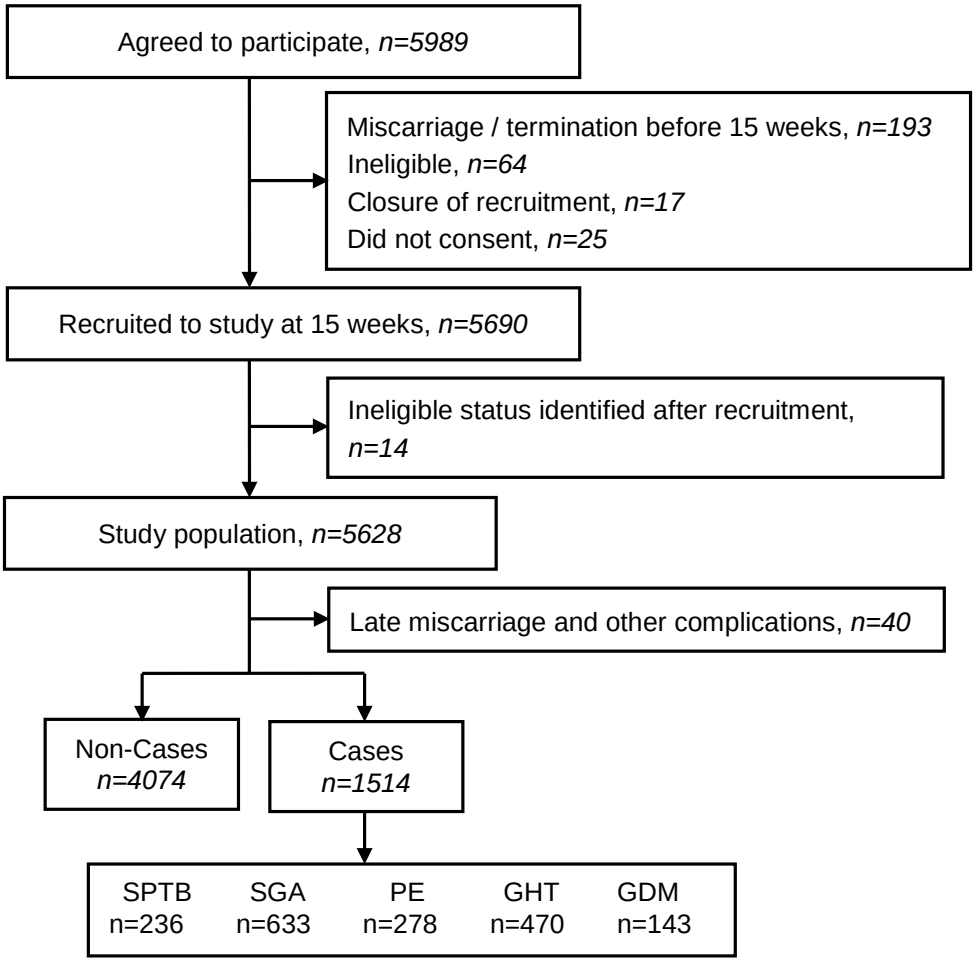

Figure 1 Participants recruited and study population

Table 1: Country specific demographics

Variable	Category	Overall (n=5588)	Australia (n=1155)	New Zealand (n=2014)	Ireland (n=1765)	United Kingdom (n=654)
		Mean ± SEM N (%)	Mean ± SEM N (%)	Mean ± SEM N (%)	Mean ± SEM N (%)	Mean ± SEM N (%)
Age		28.6 ± 0.1	23.8 ± 0.2	30.4 ± 0.1	29.9 ± 0.1	28.5 ± 0.2
SEI		41.8 ± 0.2	27.8 ± 0.3	47.9 ± 0.3	42.7 ± 0.4	45.4 ± 0.7
BMI		25.3 ± 0.1	27.0 ± 0.2	24.8 ± 0.1	24.9 ± 0.1	25.0 ± 0.2
Cigarette smoking	Yes ‡	1473 (26.36)	471 (40.78)	285 (14.15)	524 (29.69)	193 (29.51)
	Quit (pre-preg)	113 (2.02)	17 (1.46)	40 (1.99)	36 (2.04)	20 (3.06)
	Quit (<15 wks)	699 (12.51)	157 (13.66)	154 (7.65)	294 (16.66)	94 (14.37)
	Quit (<20 wks)	94 (1.68)	41 (3.61)	17 (0.84)	24 (1.36)	12 (1.83)
	Yes (at 20 Wks)	567 (10.15)	256 (21.99)	74 (3.67)	170 (9.63)	67 (10.24)
Marijuana	Yes ‡	315 (5.64)	134 (11.60)	90 (4.47)	67 (3.80)	24 (3.67)
	Quit (pre-preg)	95 (1.70)	12 (1.04)	45 (2.23)	26 (1.47)	12 (1.83)
	Quit (<15 wks)	145 (2.59)	70 (6.06)	32 (1.59)	35 (1.98)	8 (1.22)
	Quit (<20 wks)	22 (0.39)	14 (1.21)	4 (0.20)	3 (0.17)	1 (0.15)
	Yes (at 20 Wks)	53 (0.95)	38 (3.29)	9 (0.45)	3 (0.17)	3 (0.46)

‡ *Yes = smoked cigarette / used marijuana at least once*

Table 2: Overall demographics

Variable	Category	Non-Cases (n=4074)	SPTB (n=236)	P	SGA (n=633)	P	PE (n=278)	P	GHT (n=470)	P	GDM (n=143)	P
		Mean ± SEM N (%)	Mean ± SEM N (%)		Mean ± SEM N (%)		Mean ± SEM N (%)		Mean ± SEM N (%)		Mean ± SEM N (%)	
Age		28.7 ± 0.09	28.3 ± 0.39	0.217	28.6 ± 0.23	0.519	27.7 ± 0.34	***0.002***	28.8 ± 0.25	0.712	30 ± 0.44	***0.008***
SEI		42.5 ± 0.26	40.4 ± 1.08	0.059	40.1 ± 0.64	***< 0.001***	38 ± 0.93	***< 0.001***	39.7 ± 0.76	***< 0.001***	38.9 ± 1.36	***0.011***
BMI		24.8 ± 0.07	25.4 ± 0.35	***0.028***	25.9 ± 0.22	***< 0.001***	27.8 ± 0.38	***< 0.001***	27.9 ± 0.27	***< 0.001***	29.1 ± 0.52	***< 0.001***
BMI (category)	< 20	310 (7.61)	24 (10.17)	0.054	50 (7.9)	0.138	10 (3.6)	0.362	13 (2.77)	0.116	4 (2.8)	0.509
	21 – 25	2187 (53.68)	108 (45.76)	Ref	276 (43.6)	Ref	96 (34.53)	Ref	146 (31.06)	Ref	40 (27.97)	Ref
	26 – 30	1093 (26.83)	68 (28.81)	0.147	188 (29.7)	***0.002***	95 (34.17)	***< 0.001***	181 (38.51)	***< 0.001***	40 (27.97)	***0.002***
	> 30	484 (11.88)	36 (15.25)	***0.040***	119 (18.8)	***< 0.001***	77 (27.7)	***< 0.001***	130 (27.66)	***< 0.001***	59 (41.26)	***< 0.001***
Smoking	Yes	1024 (25.14)	69 (29.24)	-	213 (33.65)	-	70 (25.18)	-	138 (29.36)	-	34 (23.78)	-
	Quit (pre-preg)	85 (2.09)	3 (1.27)	0.459	7 (1.11)	0.195	4 (1.44)	0.473	16 (3.4)	***0.049***	4 (2.8)	0.597
	Quit (<15 wks)	513 (12.59)	23 (9.75)	0.380	74 (11.69)	0.731	36 (12.95)	0.878	69 (14.68)	0.133	13 (9.09)	0.248
	Quit (<20 wks)	64 (1.57)	5 (2.12)	0.451	14 (2.21)	0.122	6 (2.16)	0.462	9 (1.91)	0.478	2 (1.4)	0.853
	Yes (at 20 Wks)	362 (8.89)	38 (16.1)	***< 0.001***	118 (18.64)	***< 0.001***	24 (8.63)	0.899	44 (9.36)	0.516	15 (10.49)	0.599
Marijuana	Yes	217 (5.33)	27 (11.44)	-	45 (7.11)	-	10 (3.60)	-	21 (4.47)	-	8 (5.59)	-
	Quit (pre-preg)	71 (1.74)	7 (2.97)	0.137	10 (1.58)	0.816	0 (0.0)	0.961	5 (1.06)	0.280	3 (2.1)	0.752
	Quit (<15 wks)	102 (2.5)	7 (2.97)	0.552	18 (2.84)	0.573	8 (2.88)	0.745	14 (2.98)	0.569	3 (2.1)	0.769
	Quit (<20 wks)	14 (0.34)	2 (0.85)	0.202	5 (0.79)	0.104	1 (0.36)	0.979	1 (0.21)	0.637	1 (0.7)	0.492
	Yes (at 20 Wks)	30 (0.74)	11 (4.66)	***< 0.001***	12 (1.9)	***0.005***	1 (0.36)	0.471	1 (0.21)	0.219	1 (0.7)	0.962

Table 3: Gestational age (wks) at delivery by marijuana use in whole cohort ǂ

Marijuana	n	mean ± SEM	P	<28wks (n=56)	<32wks (n=93)	<37wks (n=401)
No	5312	39.6 ± 0.3	Reference	50 (0.94%)	80 (1.51%)	367 (6.91%)
Quit (pre-preg)	96	39.5 ± 0.3	0.507	2 (2.08%)	2 (2.08%)	10 (10.42%)
Quit (<15 wks)	145	39.6 ± 0.2	0.854	0 (0%)	3 (2.07%)	10 (6.90%)
Quit (<20 wks)	22	39.4 ± 0.5	0.658	0 (0%)	0 (0%)	2 (9.09%)
Yes (at 20 wks)	53	37.4 ± 0.7	***< 0.001***	4 (7.55%)	8 (15.09%)	12 (22.64%)

ǂ *includes 165 Iatrogenic PTBs and 236 spontaneous PTBs*

Table 4: Gestational age (wks) at delivery by marijuana use within SPTB cases

Marijuana	n	mean ± SEM	P	<28wks (n=56)	<32wks (n=93)	<37wks (n=401)
No	209	34.1 ± 0.3	Reference	16 (4.66%)	33 (15.79%)	209 (100%)
Quit (pre-preg)	7	33.8 ± 1.6	0.934	1 (14.29%)	1 (14.29%)	7 (100%)
Quit (<15 wks)	7	33.8 ± 1.2	0.649	0 (0%)	2 (28.57%)	7 (100%)
Quit (<20 wks)	2	33.4 ± 1.0	0.247	0 (0%)	0 (0%)	2 (100%)
Yes (at 20 Wks)	11	29.6 ± 1.6	0.005	4 (36.36%)	7 (63.64%)	11 (100%)

Table 5: Logistic regression model specifications for SPTB, SGA, PE, GHT, and GDM

		SPTB		SGA		PE		GHT		GDM	
Variable	**Category**	**Adj Odds (95% CI)**	**P**	**Adj Odds (95% CI)**	**P**	**Adj Odds (95% CI)**	**P**	**Adj Odds (95% CI)**	**P**	**Adj Odds (95% CI)**	**P**
Age		1.01 (0.98 - 1.04)	0.395	1.02 (1.00 - 1.04)	***0.010***	0.98 (0.95 - 1.01)	0.122	1.01 (0.99 - 1.04)	0.222	1.08 (1.04 - 1.12)	***< 0.001***
SEI		1.00 (0.99 - 1.01)	0.535	0.99 (0.99 - 1.00)	0.054	0.99 (0.98 - 1.00)	***0.023***	0.99 (0.99 - 1.00)	0.171	0.99 (0.97 – 1.00)	0.061
BMI		1.03 (1.00 - 1.06)	0.058	1.04 (1.03 - 1.06)	***0.000***	1.10 (1.08 - 1.13)	***< 0.001***	1.12 (1.10 - 1.14)	***< 0.001***	1.13 (1.10 - 1.17)	***< 0.001***
Smoking	No	Reference		Reference		Reference		Reference		Reference	
	Quit (pre-preg)	0.73 (0.15 - 3.53)	0.698	0.90 (0.27 - 3.03)	0.882	0.33 (0.06 - 1.92)	0.218	7.31 (1.47 - 36.29)	***0.015***	4.70 (0.44 - 49.92)	0.199
	Quit (<15 wks)	0.93 (0.28 - 3.05)	0.906	1.55 (0.58 - 4.12)	0.362	0.40 (0.09 - 1.71)	0.216	4.39 (0.95 - 20.25)	0.058	2.34 (0.25 - 21.84)	0.455
	Quit (<20 wks)	1.50 (0.35 - 6.42)	0.587	2.25 (0.74 - 6.85)	0.134	0.50 (0.10 - 2.52)	0.397	5.32 (1.00 - 28.31)	0.050	3.51 (0.26 - 47.25)	0.344
	Yes (at 20 Wks)	1.72 (0.53 - 5.52)	0.366	3.46 (1.31 - 9.10)	***0.012***	0.31 (0.07 - 1.33)	0.115	3.70 (0.80 - 17.19)	0.095	3.36 (0.36 - 30.89)	0.285
Marijuana	No	Reference		Reference		Reference		Reference		Reference	
	Quit (pre-preg)	2.16 (0.82 - 5.73)	0.120	1.08 (0.50 - 2.32)	0.839	-	0.995	0.93 (0.34 - 2.50)	0.880	2.45 (0.64 - 9.36)	0.189
	Quit (<15 wks)	1.21 (0.50 - 2.90)	0.672	0.96 (0.55 - 1.68)	0.897	0.71 (0.26 - 1.96)	0.514	1.51 (0.80 - 2.86)	0.205	1.39 (0.39 - 4.99)	0.610
	Quit (<20 wks)	2.14 (0.46 - 9.96)	0.334	1.68 (0.57 - 4.93)	0.343	0.89 (0.11 - 7.20)	0.909	0.66 (0.08 - 5.48)	0.697	1.79 (0.17 - 18.48)	0.623
	Yes (at 20 Wks)	5.13 (2.30 - 11.43)	***< 0.001***	1.88 (0.92 - 3.85)	0.083	0.44 (0.06 - 3.36)	0.427	0.40 (0.05 - 3.04)	0.377	1.49 (0.19 - 11.96)	0.706
Smoking X Marijuana	Interaction term	1.25 (0.38 - 4.08)	0.709	1.58 (0.60 - 4.20)	0.356	0.43 (0.10 - 1.86)	0.261	4.41 (0.95 - 20.39)	0.058	4.00 (0.43 - 36.92)	0.221

Table 6: Risk of pregnancy complications for any marijuana use (3 months prior to or during pregnancy) adjusted for cigarette smoking status

Outcomes	Marijuana Odds (95% CI)	P-value ‡	Odds (95% CI) adjusted for any Smoking *	P-value	Odds (95% CI) adjusted for Smoking at 20 wks **	P-value
SPTB	2.31 (1.45 - 3.55)	***< 0.001***	2.28 (1.49- 3.60)	***< 0.001***	1.97 (1.26 - 3.09)	***0.004***
SGA	1.37 (0.96 - 1.92)	0.064	1.13 (0.80 - 1.60)	0.555	1.04 (0.73 - 1.47)	0.917
PE	0.67 (0.312 - 1.27)	0.216	0.66 (0.34 - 1.27)	0.259	0.66 (0.34 - 1.28)	0.272
GHT	0.74 (0.458 - 1.19)	0.443	0.25 (0.13 - 3.54)	0.671	0.81 (0.51 - 1.30)	0.454
GDM	1.06 (0.442 - 2.19)	0.877	1.11 (0.52 - 2.38)	0.949	1.01 (0.48 - 2.10)	0.986

*‡ overall p-value comparing marijuana and corresponding outcome; * Mantel-Haenszel adjusted odds adjusted for any cigarette smoking (3 months prior to or during pregnancy); ** Mantel-Haenszel adjusted odds adjusted for ceased cigarette smoking at 15 weeks' gestation*

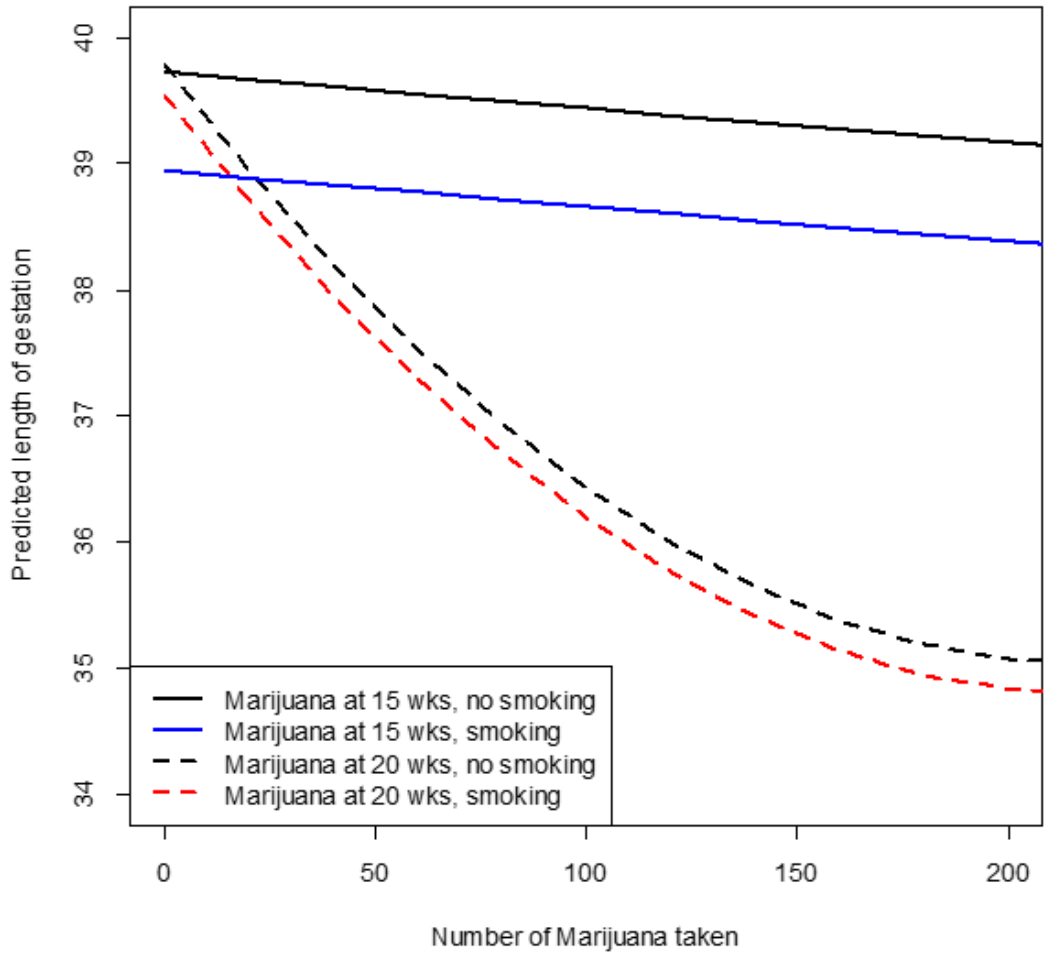

Figure 2 Predicted length of gestation and number of episodes of marijuana taken over 3 months (adjusted for age, BMI, SEI, and cigarette smoking)

Chapter 7: Manuscript II – Preeclampsia Model

Preeclampsia Prediction at 15 weeks of gestation: A Tiered Modelling Approach

Abstract

Background. For years, it has been a challenge to identify nulliparous women at risk of Preeclampsia (PE), one of the leading causes of maternal and perinatal morbidity and mortality. This would be especially useful in early pregnancy when modifiable factors can be addressed to reduce the risk or severity of outcome. Despite an increasing number of clinical and statistical prediction models being developed, which have been shown to outperform traditional maternal history or Doppler ultrasound approaches, it is still difficult to make accurate predictions based on a single model. Hence, this paper proposes a tiered modelling approach for prediction at 15 weeks' gestation.

Methods. A total of 2977 participants from the Australian and New Zealand cohorts of the Screening for Pregnancy Endpoints (SCOPE) study were included in the analysis, with 167 PE cases and 2810 women with no PE. Two models based on predictors available at 15 weeks of gestation were developed with clinical predictors in Tier 1, and adding SNP predictors in Tier 2. Post-test probabilities are then calculated based on the Likelihood of each model using Bayes' theorem, and the final risk is classified into 3 levels.

Results. The prediction of truly identified cases has improved using the tiered modelling approach, with a sensitivity of 91% in Tier 1 and PPV of 22.94% in Tier 2.1032 women were classified as low risk of PE at 15 weeks' gestation, with 15 cases (1.45%) missed. Amongst the 327 women further predicted as high risk in Tier 2, 75 (22.94%) developed PE.

Conclusion. Through tiered modelling, the accuracy and precision of prediction is further enhanced and tailored for individual women. This model also provides a risk prediction that does not depend on 2nd trimester predictors, e.g. uterine artery Doppler, and could be used to identify women at risk for PE who could then have tailored antenatal care. Modifiable predictors at 15 weeks of gestation may also be addressed to reduce the risk or severity of PE. Identification of women at a high risk is essential to implement existing and novel interventions.

Keywords: *Preeclampsia, prediction, Bayes' theorem, 3-tiered model*

Introduction

Preeclampsia (PE), a hypertensive disorder of pregnancy, is one of the major causes of maternal and perinatal morbidity and mortality and affects around 3-5% pregnancies worldwide [1, 2]. With an increased risk of severe complications due to delays in diagnosis, screening or prediction tools prior to symptoms are essential for assessment of interventions and tailored antenatal care.

The complexity in developing methods of prediction for preeclampsia is largely due to its low prevalence, unknown aetiology and absence of a 'gold standard' [3]. Current approaches based on maternal history uterine artery Doppler ultrasound studies during 2nd trimester have an estimated sensitivity of only 40% and 60%, respectively[4, 5]. Despite an increasing number of recent clinical and statistical prediction models[6-9], which have been shown to outperform traditional approaches, the majority of models only provide risk estimation during late second trimester, and preventative treatment is often delayed.

Since an early prediction of risk is desired but a single model may not be satisfactory, a multi-model or tiered approach is considered with individual models tailored for each tier. This paper will discuss the application and effectiveness of a tiered approach, integrated by Bayes' theorem, on the early prediction of PE.

Methods

The models are developed based on the Australian and New Zealand cohorts of the SCOPE (Screening fOr Pregnancy Endpoints) study. This study recruited nulliparous women with singleton pregnancies between November 2004 and August 2008, with ethical approval from local ethics committees [New Zealand AKX/02/00/364, Australia REC 1712/5/2008] and all women provided written informed consent[10].

Women prior to 15 weeks' gestation attending hospital antenatal clinics, obstetricians, general practitioners or community midwives were invited to participate, and were later interviewed and examined by a research midwife at 15±1 and 20±1 weeks of gestation.

Participants who were considered to be at high risk of PE, SGA or PTB due to underlying medical conditions (e.g. chronic hypertension requiring antihypertensive medication or diabetes), previous cervical knife cone biopsy, 3 terminations or 3 miscarriages, current ruptured membranes, or their pregnancy was complicated by a known major fetal anomaly or abnormal karyotype, or if they received interventions that may modify pregnancy outcome (e.g. aspirin, cervical suture), were excluded.

Details of maternal history, dietary practices and clinical measurements at 3 time points

(pre-pregnancy, 15 weeks and 20 weeks of gestation), were recorded into an internet accessed, central database with a complete audit trial (Medscinet[AB], Stockholm, Sweden)[10]. Blood samples were obtained from women at 15 weeks' gestation, from partners at some time during the women's pregnancy and from cord blood at birth. DNA was extracted and genotype data were obtained for 100 candidate single nucleotide polymorphisms (SNPs) using the Sequenom Mass Array Platform as previously described [11].

Preeclampsia (PE) was defined as gestational hypertension (GHT) (blood pressure of ≥140/90 on at least 2 occasions 4 hours apart after 20 weeks' gestation) accompanied by proteinuria (300 mg/day or greater, or a spot protein creatinine ratio of 30 mg/mmol creatinine or greater).

Tiered Modelling

A total of 2977 participants were included in the analysis, with 167 cases of PE and 2810 non-PE (Figure 1). Two models (Tiers 1 and 2) were developed separately using penalized Logistic regression, with variable selection from Elastic Net regularization [12, 13], based on predictors collected at 15 weeks of gestation. To assess the classification of the models, sensitivity and specificity were calculated, and models were selected based on specific levels of sensitivity and specificity.

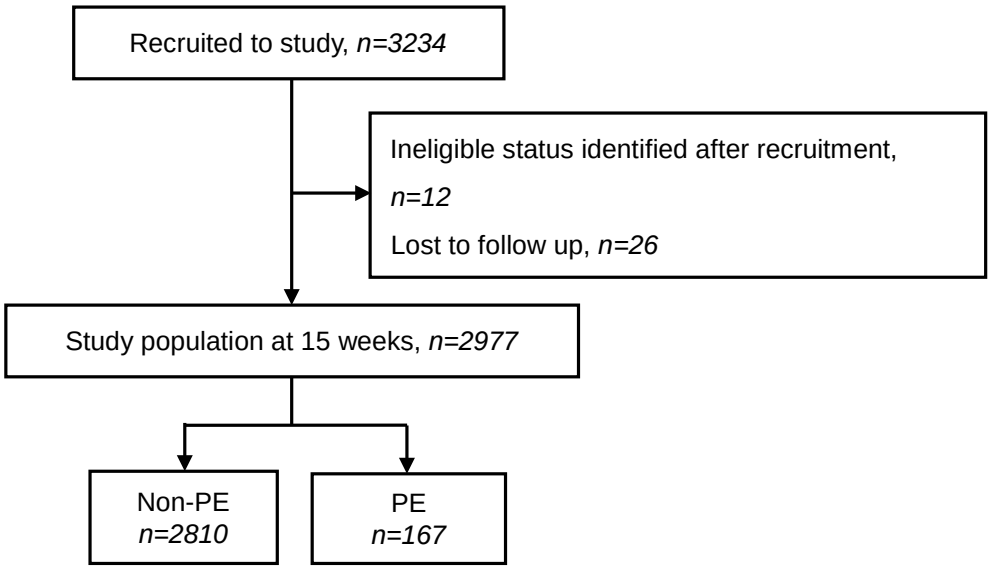

Figure 1. Participants recruited and study population

Tier 1 was developed using clinical predictors at 15 weeks' from a model previously published by SCOPE [6], but the prediction probability threshold was set lower for higher sensitivity, since Tier 1 will serve as an initial screening to identify all patients with potential risk for PE. The subsequent prediction in Tier 2, also at 15 weeks' gestation, but aimed at identifying a high risk group with a higher positive predictive value, SNP predictors were added in combination with clinical predictors obtained in Tier 1.

To obtain a probability estimate that integrates prior predicted risk and the likelihood of current prediction, Bayes' theorem was applied [14-16], where the post-test odds of Tier 2 is determined by the odds from Tier 1 and likelihood of Tier 2 test result:

$$O_{\text{Tier 2}}\left(D\middle|T_{+/-}\right) = O_{\text{Tier 1}}(D) \cdot \Lambda_{\text{Tier 2}}\left(D\middle|T_{+/-}\right) \quad (1)$$

where $\Lambda_{\text{Tier 2}}\left(D\middle|T_{+/-}\right) = \frac{P\left(T_{+/-}\middle|D\right)}{P\left(T_{+/-}\middle|\overline{D}\right)}$ is the likelihood of PE given a positive or negative test result in Tier 2, calculated from the sensitivity and specificity, where

$\Lambda_{\text{Tier 2}}(D|T_+) = \frac{r}{1-s}$ and $\Lambda_{\text{Tier 2}}(D|T_-) = \frac{1-r}{s}$.

Risk Classification

With tiered modelling, the risk of PE can be classified into 3 levels (low, moderate and high) according to the result of each tier (Figure 2). Since the sensitivity is relatively high at initial screening in Tier 1, the likelihood of PE in patients predicted as low risk at this tier will be low. Hence, a first group of patients considered at low-risk can be 'eliminated' from Tier 2 screening.

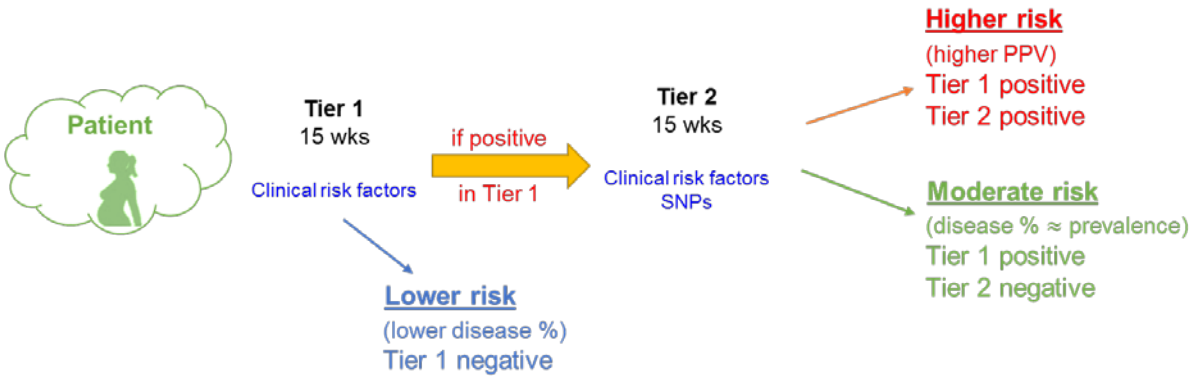

Figure 2. Tiered modelling approach.

Once patients who are potentially at risk are identified, the next goal is to further predict patients who are at high risk. This is the purpose of Tier 2 having a higher positive predictive value, preferably higher than 20%. As a result, the proportion of PE will increase for the high risk group, with patients predicted at lower risk 'eliminated' at each tier. Tailored antenatal care may be provided according to women's classified predicted risk.

Results

Two models were trained on 90% of the study population at 15 weeks' gestation, with one that includes clinical predictors only, and then adding SNP predictors in the second model (Table 1). Maternal BMI is a significant risk factor for both tiers, with an odds of 1.06 (95% CI 1.02 to 1.09). Mean arterial pressure (MAP) measured at 15 weeks' gestation also increased the risk of PE (OR 2.03; 95% CI 1.04 to 1.08; P=0.000). Having a family history of diseases associated with hypertension, such as family history of PE (OR 2.03; 95% CI 1.34 to 3.08; P=0.001), also increased the risk of PE.

Other factors, including vaginal bleeding for 5 days or more before 15 weeks of gestation (OR 2.27; 95% CI 1.21 to 4.25; P=0.010) increased the risk of PE, while having had a miscarriage at less than 10 weeks' with the same partner (OR 0.42; 95% CI 0.19 to 0.92; P=0.030), and increased number of months to conceive (OR 0.42; 95% CI 0.22 to 0.80; P=0.009) are protective factors.

A total of 13 SNP predictors were included in Tier 2, with 6 maternal and 7 paternal SNPs. Genes that have previously been associated with PE, including IL10, AGTR1 and MTHFR were included in the final models.

Table 1. Predictors for PE in Tier 1 and 2 at 15 weeks of gestation

	Tier 1		*Tier 2*	
Predictors	**Odds (95% CI)**	**P**	**Odds (95% CI)**	**P**
Age (maternal)	0.962 (0.933 - 0.992)	***0.0141***	-	-
MAP (at 15 wks)	1.062 (1.040 - 1.084)	***0.0000***	1.073 (1.041 - 1.106)	***0.0000***
BMI (maternal)	1.059 (1.031 - 1.087)	***0.0000***	1.058 (1.018 - 1.099)	***0.0040***
FH (PE)[‡]	2.030 (1.338 - 3.080)	***0.0009***	2.889 (1.565 - 5.332)	***0.0007***
FH (CH)[*]	1.143 (0.820 - 1.594)	0.4312	-	-
Participant's birthweight	1.000 (0.999 - 1.000)	***0.0105***	1.000 (0.999 - 1.000)	0.0508
Vaginal bleeding ≥5days	2.272 (1.214 - 4.250)	***0.0102***	-	-
Miscarriage ≤10wks	0.422 (0.193 - 0.922)	***0.0304***	0.365 (0.109 - 1.224)	0.1027
≥12mths to conceive	0.418 (0.218 - 0.802)	***0.0088***	0.377 (0.150 - 0.951)	***0.0387***
Fruit (≤1-2x per week)	1.336 (0.859 - 2.080)	0.1989	-	-
Alcohol consumption (1st trim)	1.002 (0.991 - 1.014)	0.6766	0.944 (0.886 - 1.007)	0.0803
Cigarettes per day (at 15 wks)	0.951 (0.893 - 1.013)	0.1183	-	-
AGTR1 (maternal)[SNP]	-	-	0.243 (0.055 - 1.068)	0.0611
IL10 (maternal)[SNP]	-	-	0.513 (0.312 - 0.846)	***0.0089***
MTHFR (maternal)[SNP]	-	-	3.424 (1.730 - 6.776)	***0.0004***
PGF (maternal)[SNP]	-	-	2.151 (1.032 - 4.486)	***0.0411***
PLG (maternal)[SNP]	-	-	1.745 (1.078 - 2.825)	***0.0235***
INSR (maternal)[SNP]	-	-	0.556 (0.236 - 1.307)	0.1782
NOS2A (paternal)[SNP]	-	-	0.578 (0.342 - 0.978)	***0.0411***
TP53 (paternal)[SNP]	-	-	1.625 (0.992 - 2.662)	0.0536
MTHFR (paternal)[SNP]	-	-	1.763 (0.881 - 3.527)	0.1092
INS (paternal)[SNP]	-	-	2.699 (1.157 - 6.298)	***0.0217***
TGFB (paternal)[SNP]	-	-	1.906 (0.706 - 5.142)	0.2029
PGF (paternal)[SNP]	-	-	0.521 (0.297 - 0.914)	***0.0231***
MMP2 (paternal)[SNP]	-	-	2.355 (1.076 - 5.157)	***0.0322***

[‡]Family history of Preeclampsia; [*]Family history of Chronic Hypertension

Of the 2977 patients analysed (Table 2), at initial screening (Tier 1) at 15 weeks of gestation, 35% were eliminated as at low risk for PE, and leaving 1945 patients (65%) needed to go through Tier 2 screening, whom may be recommended a higher level of attention. By Tier 2 (also at 15 weeks of gestation), a further 54% are eliminated based

on new SNP predictors at this stage, with 327 identified at high risk of PE and may be monitored more closely. Amongst the 327 patients predicted at high risk, whom may be recommended for a higher level of care and/or initiate preventative treatment, 75 patients (PPV 22.94%) eventually had PE.

Table 2. Final risk classification for preeclampsia.

Tier 1	Tier 2	Classification	Predicted	Observed	PE %
-		Low risk	1032	15	1.45%
+	-	Moderate risk	1618	77	4.76%
+	+	High risk	327	75	22.94%
		Overall		167	5.6%

With 152 out of 167 PE cases successfully predicted at risk in Tier 1, a sensitivity of 91% was achieved. The percentage of PE in the low-risk group 'eliminated' at Tier 1 should be relatively low, with only 15 PE cases (1.45%) missed, resulting in a negative predictive value of 98.6%.

It is important to note that amongst the PE cases missed in women predicted to be at low risk, none delivered before 34 weeks of gestation, and only 3 cases (20%) delivered between 34 to 37 weeks (Table 3). In addition, although the average gestational ages of the 3 risk groups are similar, the average birthweight of the high-risk group (3284 ± 13) is lower than those predicted at low (3394g ± 10; P=0.01).

Table 3a. Birth characteristics for overall cohort

		Birthweight (g)	Gestational age (wks)			
Risk	**N**	**Mean ± SEM**	**Mean ± SEM**	**<28 wks (n=25)**	**<32wks (n=68)**	**<37 wks (n=231)**
Low	1032	3394 ± 10	39.6 ± 0.04	5 (0.49%)	7 (0.68%)	62 (6.01%)
Moderate	1618	3414 ± 11	39.5 ± 0.04	15 (0.93%)	22 (1.36%)	126 (7.79%)
High	327	3284 ± 13	39.0 ± 0.05	5 (1.53%)	39 (3.06%)	43 (13.15%)

Table 3b. Birth characteristics within PE cases

		Birthweight (g)	Gestational age (wks)			
Risk	**N**	**Mean ± SEM**	**Mean ± SEM**	**<28 wks (n=1)**	**<32wks (n=6)**	**<37 wks (n=46)**
Low	15	2814 ± 44	38.0 ± 0.12	0 (0%)	0 (0%)	3 (20%)
Moderate	77	3090 ± 58	38.2 ± 0.19	0 (0%)	0 (0%)	21 (27.28%)
High	75	2995 ± 67	37.5 ± 0.23	1 (4.92%)	6 (13.12%)	22 (29.33%)

Discussion

Our data show that prediction for PE using tiered models is enhanced, with individual models developed for specific purposes. Update of predicted risk for individuals is possible when new predictors are available or when conditions change, and hence, the level of care may be tailored for individual women (Figure 3).

The majority of clinical predictors in the models are well recognized factors, which include obesity, obstetric history and family history [3, 17-19], consistent with previously published data [6]. Other studies have reported that the risk of PE increases with higher BMI, where risk in women who have a BMI of 26 is doubled compared to those who have a BMI of 21, and increases further with severe obesity [20-22]. Having a low maternal

birth weight and preterm birth have also been found to increase risk of PE[23].

It has been shown that preeclampsia occurs in 26% of the daughters and 16% of the granddaughters of women who had preeclampsia [24], with subsequent studies estimating that the incidence of preeclampsia is nearly tripled amongst women with a family history of preeclampsia [25-27], and that having a family history of hypertension is also associated with PE, with an estimated odds ratio of 1.7 [28].

Most genetic factors that are in the models are candidate genes that are relevant to the physiological pathways for PE[29]. These includes genes that are linked with inflammation, such as Interleukin-10 (*IL10*)[30-33] and Transforming Growth Factor β (*TGFβ*)[29, 34, 35], Methylenetetrahydrofolatereductase (*MTHFR*) [36-39], and Angiotensin type 1 receptors (*AGTR1*) [40, 41]. Although the associations between PE and other genes such as nitric oxide synthase (NOS) [42-44] and matrix metallopeptidase 2 (MMP2)[45, 46]have been studied, PE is often considered as a maternal disease and its association with paternal SNPs is a less discussed topic [47].

Nevertheless, studies have shown that men who were born to a PE pregnancy are more likely to parent a PE pregnancy [25, 48, 49]. Also, previous results from SCOPE study have identified paternal SNPs associated with PE [11] and with SGA [50]. In addition, imprinted genes expressed by the paternal allele in the placenta or fetus could confer

risk for pregnancy complications [49].

Application of Tiered Prediction Model

The tiered prediction model utilizes the process of elimination to screen patients at risk for PE at each tier, and to classify the level of risk (Figure 3). With a high sensitivity of 91% and PPV of 98.6%, the majority of PE cases should be identified at Tier 1. Women predicted at low risk can be 'eliminated' from Tier 2 screening and continue regular antenatal visits, while those predicted at risk may benefit from more frequent monitoring.

A subsequent prediction, also available at 15 weeks' gestation, is then used to identify women at higher risk, but with maternal and paternal SNP predictors in combination with clinical predictors already obtained in Tier 1. The likelihood of PE in the predicted high-risk group should be highest (PPV above 20%) amongst all 3 classified risk groups.

Patients with a negative test result in Tier 2 are considered at moderate risk, with 4.76% PE cases. The current antenatal care system used for nulliparous pregnant women may still be beneficial, as there is still a potential risk.

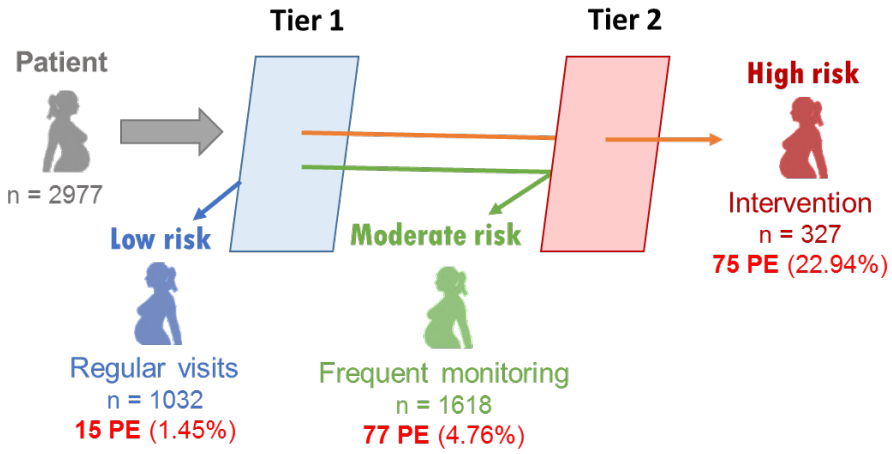

Figure 3. Process of elimination using tiered modelling

At 15 weeks' gestation, modifiable risk factors such as fruit consumption may still be addressed. Other studies have shown that higher fruit and vegetable intake during the first 4 to 5 months of pregnancy reduces the risk of preeclampsia [51, 52]. A model to identify women likely to have an uncomplicated pregnancy in SCOPE women found that consumption of fruit and vegetables protected women from developing pregnancy complications [53]. More importantly, women who are predicted to be at high risk may benefit from early preventative treatments. Recent studies have reported that low dose aspirin administered in high-risk women before 16 weeks of gestation have significant reduction in the risk of PE [54-56]. In addition, calcium supplementation before 34 weeks of gestation has also been shown to reduce the risk of PE in women with low calcium intakes [57, 58].

In this analysis of 2432 participants, 1032 women (34.7%) were 'eliminated' as at low

risk at Tier 1 at 15 weeks' gestation, with only 1.45% of PE cases missed. A further 54.3% were predicted to be at moderate risk at Tier 2 using additional SNP predictors, leaving 327 women (11%) identified at high risk, in which 75 patients (22.94%) had PE.

The estimated number needed to treat in the high-risk group at Tier 2 is 5.2 to prevent one PE case, with an absolute risk reduction of 19.3%, and at least 144 women needed to screen [59] to prevent a PE case.

Strengths and Limitations

The ability to classify risk from different tiers, and 'eliminating' patients considered as low risk from the screening process to minimize the chance of unnecessary interventions is a major strength of the tiered model. By classifying women at different levels of risk, antenatal care may be tailored for individuals, and women who are at risk may benefit from a higher level of care.

Moreover, with predictors that can be obtained at 15 weeks' gestation, this model provides an advantage for earlier risk prediction that does not depend on known second trimester factors such as abnormal uterine artery Doppler studies. In addition, since this analysis is from a cohort of nulliparous women, the tiered model is also independent of previous history of PE, in contrast to existing models [7, 8, 60] where a previous history of PE is included as a predictor. More importantly, at 15 weeks of

gestation, modifiable risk factors may be addressed and preventative strategies may still be applied after prediction to reduce the risk or severity of PE.

Although the tiered model may not outperform other prediction models recently published, with some models achieving a sensitivity or specificity of above 90% [61-64], it is important to note that most of these prediction models are for early-onset preeclampsia, which is quite rare, and the models do not perform well for the more common preeclampsia at term. Our tiered model predicts all PE. It is also worth noting that the results from the tiered model showed that all women with early-onset preeclampsia, delivered before 34 weeks' gestation, were successfully identified to be at risk, and more importantly, of the 1.45% PE cases missed, none had early-onset PE.

Interestingly, the tiered model with a positive predictive value of 22.94% at 15 weeks' is similar to the estimated PPV of 22.6% using Doppler assessment performed at 23 weeks of gestation [65]. However, 23 weeks' would be too late for currently known preventative strategies such as low dose aspirin which should be commenced at 16 weeks' [54].

The uncertainty for patients identified as at moderate risk is a limitation of the tiered model, where the proportion of PE is similar to the current prevalence. At this stage,

patients will rely on frequent monitoring, and further research will be needed to improve the prediction on this group of patients. Also, the availability of paternal genotypes may also be a potential limitation in some settings.

Conclusion

Our data have demonstrated that through tiered modelling, prediction for PE is further enhanced, with 91% sensitivity in Tier 1 and a PPV of 22.94% in Tier 2.This model could be used to identify women at risk for PE who could then have tailored antenatal care, while modifiable risk factors at 15 weeks of gestation can also be addressed and novel interventions applied to reduce the risk or severity of preeclampsia.

References

1. Hanretty, K., *Obstetrics Illustrated*. 7th ed. 2009: Churchill Livingstone. 448.

2. Pfeifer, S.M., *NMS Obstetrics and Gynecology*. 6th ed. National Medical Series for Independent Study. 2007: Lippincott Williams & Wilkins. 496.

3. Briceno-Perez, C., L. Briceno-Sanabria, and P. Vigil-De Gracia, *Prediction and prevention of preeclampsia.* Hypertension in Pregnancy, 2009. **28**(2): p. 138-55.

4. Jacquemyn, Y. and O. Zemtsova, *Risk factors and prediction of preeclampsia.* Acta Clinica Belgica, 2010. **65**(1): p. 1-12.

5. Papageorghiou, A.T., et al., *Assessment of risk for the development of pre-eclampsia by maternal characteristics and uterine artery Doppler.* British Journal of Obstetrics and Gynaecology, 2005. **112**(6): p. 703-9.

6. North, R.A., et al., *Clinical risk prediction for pre-eclampsia in nulliparous women: development of model in international prospective cohort.* BMJ, 2011. **342**: p. d1875.

7. Akolekar, R., et al., *Competing risks model in early screening for preeclampsia by biophysical and biochemical markers.* Fetal Diagn Ther, 2013. **33**(1): p. 8-15.

8. Wright, D., et al., *A competing risks model in early screening for preeclampsia.* Fetal Diagn Ther, 2012. **32**(3): p. 171-8.

9. Yu, C.K., et al., *Prediction of pre-eclampsia by uterine artery Doppler imaging: relationship to gestational age at delivery and small-for-gestational age.* Ultrasound Obstet Gynecol, 2008. **31**(3): p. 310-3.

10. McCowan, L., R. North, and R. Taylor. *ACTRN12607000551493*. 2007; Available from: www.anzctr.org.autrialSearch.aspx.

11. Andraweera, P.H., et al., *Single-nucleotide polymorphisms in the KDR gene in pregnancies complicated by gestational hypertensive disorders and small-for-*

gestational-age infants. Reprod Sci, 2012. **19**(5): p. 547-54.

12. Zou, H. and T. Hastie, *Regularization and variable selection via the Elastic Net.* Journal of Royal Statistical Society, Series B, 2005. **67**(2): p. 301-320.

13. Wu, T.T., et al., *Genome-wide association analysis by lasso penalized logistic regression.* Bioinformatics, 2009. **25**(6): p. 714-721.

14. Lindgaard, G., et al., *Does Bayes have it? Decision Support Systems in diagnostic medicine.* International Journal of Industrial Ergonomics, 2009. **39**(3): p. 524-532.

15. Hall, G.H., *The clinical application of Bayes' theorem.* The Lancet, 1967. **290**(7515): p. 555-557.

16. Round, A., *Introduction to clinical reasoning.* Journal of Evaluation in Clinical Practice, 2001. **7**(2): p. 109-117.

17. Carty, D.M., C. Delles, and A.F. Dominiczak, *Novel biomarkers for predicting preeclampsia.* Trends Cardiovasc Med, 2008. **18**(5): p. 186-94.

18. Farag, K., I. Hassan, and W.L. Ledger, *Prediction of preeclampsia: can it be achieved?* Obstet Gynecol Surv, 2004. **59**(6): p. 464-82.

19. Steegers, E.A.P., et al., *Pre-eclampsia.* The Lancet, 2010. **376**(9741): p. 631-644.

20. Bodnar, L.M., et al., *The risk of preeclampsia rises with increasing prepregnancy body mass index.* Ann Epidemiol, 2005. **15**(7): p. 475-82.

21. Stone, C.D., et al., *The combined effect of maternal smoking and obesity on the risk of preeclampsia.* J Perinat Med, 2007. **35**(1): p. 28-31.

22. Ros, H.S., S. Cnattingius, and L. Lipworth, *Comparison of risk factors for preeclampsia and gestational hypertension in a population-based cohort study.* Am J Epidemiol, 1998. **147**(11): p. 1062-70.

23. Innes, K.E., et al., *A woman's own birth weight and gestational age predict her*

later risk of developing preeclampsia, a precursor of chronic disease. Epidemiology, 1999. **10**(2): p. 153-60.

24. Chesley, L.C. and D.W. Cooper, *Genetics of hypertension in pregnancy: possible single gene control of pre-eclampsia and eclampsia in the descendants of eclamptic women.* Br J Obstet Gynaecol, 1986. **93**(9): p. 898-908.

25. Esplin, M.S., et al., *Paternal and maternal components of the predisposition to preeclampsia.* N Engl J Med, 2001. **344**(12): p. 867-72.

26. Cincotta, R.B. and S.P. Brennecke, *Family history of pre-eclampsia as a predictor for pre-eclampsia in primigravidas.* Int J Gynaecol Obstet, 1998. **60**(1): p. 23-7.

27. Boyd, H.A., et al., *Associations of personal and family preeclampsia history with the risk of early-, intermediate- and late-onset preeclampsia.* Am J Epidemiol, 2013. **178**(11): p. 1611-9.

28. Eskenazi, B., L. Fenster, and S. Sidney, *A multivariate analysis of risk factors for preeclampsia.* JAMA, 1991. **266**(2): p. 237-41.

29. Mutze, S., et al., *Genes and the preeclampsia syndrome.* J Perinat Med, 2008. **36**(1): p. 38-58.

30. Valencia Villalvazo, E.Y., et al., *Analysis of polymorphisms in interleukin-10, interleukin-6, and interleukin-1 receptor antagonist in Mexican-Mestizo women with pre-eclampsia.* Genet Test Mol Biomarkers, 2012. **16**(11): p. 1263-9.

31. Sowmya, S., et al., *Evaluation of Interleukin-10 (G-1082A) Promoter Polymorphism in Preeclampsia.* J Reprod Infertil, 2013. **14**(2): p. 62-6.

32. Lau, S.Y., et al., *Tumor necrosis factor-alpha, interleukin-6, and interleukin-10 levels are altered in preeclampsia: a systematic review and meta-analysis.* Am J Reprod Immunol, 2013. **70**(5): p. 412-27.

33. Kumar, A., et al., *IL-10, TNF-alpha & IFN-gamma: potential early biomarkers for preeclampsia.* Cell Immunol, 2013. **283**(1-2): p. 70-4.

34. Boulanger, H. and M. Flamant, *[New insights in the pathophysiology of preeclampsia and potential therapeutic implications].* Nephrol Ther, 2007. **3**(7): p. 437-48.

35. Venkatesha, S., et al., *Soluble endoglin contributes to the pathogenesis of preeclampsia.* Nat Med, 2006. **12**(6): p. 642-9.

36. Driul, L., et al., *Screening for pre-eclampsia in a low-risk population at 24 weeks: uterine artery Doppler flow velocimetry and genetic variants of factor V, prothrombin and methylenetetrahydrofolate reductase.* Minerva Ginecol, 2004. **56**(5): p. 385-90.

37. Kaiser, T., S.P. Brennecke, and E.K. Moses, *Methylenetetrahydrofolate reductase polymorphisms are not a risk factor for pre-eclampsia/eclampsia in Australian women.* Gynecol Obstet Invest, 2000. **50**(2): p. 100-2.

38. Powers, R.W., et al., *Methylenetetrahydrofolate reductase polymorphism, folate, and susceptibility to preeclampsia.* J Soc Gynecol Investig, 1999. **6**(2): p. 74-9.

39. Vucic, N., et al., *Thrombophilia, preeclampsia and other pregnancy complications.* Acta Med Croatica, 2009. **63**(4): p. 297-305.

40. Kobashi, G., et al., *A1166C variant of angiotensin II type 1 receptor gene is associated with severe hypertension in pregnancy independently of T235 variant of angiotensinogen gene.* J Hum Genet, 2004. **49**(4): p. 182-6.

41. Bouba, I., et al., *Interaction between the polymorphisms of the renin-angiotensin system in preeclampsia.* Eur J Obstet Gynecol Reprod Biol, 2003. **110**(1): p. 8-11.

42. Amaral, L.M., et al., *Maternal iNOS genetic polymorphisms and hypertensive disorders of pregnancy.* J Hum Hypertens, 2012. **26**(9): p. 547-52.

43. Mazzanti, L., et al., *Nitric oxide and peroxynitrite platelet levels in gestational hypertension and preeclampsia.* Platelets, 2012. **23**(1): p. 26-35.

44. Faxen, M., H. Nisell, and K.R. Kublickiene, *Altered mRNA expression of ecNOS and iNOS in myometrium and placenta from women with preeclampsia.* Arch Gynecol Obstet, 2001. **265**(1): p. 45-50.

45. Palei, A.C., et al., *Association between matrix metalloproteinase (MMP)-2 polymorphisms and MMP-2 levels in hypertensive disorders of pregnancy.* Exp Mol Pathol, 2012. **92**(2): p. 217-21.

46. Palei, A.C., et al., *Comparative assessment of matrix metalloproteinase (MMP)-2 and MMP-9, and their inhibitors, tissue inhibitors of metalloproteinase (TIMP)-1 and TIMP-2 in preeclampsia and gestational hypertension.* Clin Biochem, 2008. **41**(10-11): p. 875-80.

47. Dekker, G., P.Y. Robillard, and C. Roberts, *The etiology of preeclampsia: the role of the father.* J Reprod Immunol, 2011. **89**(2): p. 126-32.

48. Lie, R.T., et al., *Fetal and maternal contributions to risk of pre-eclampsia: population based study.* BMJ, 1998. **316**(7141): p. 1343-7.

49. Roberts, C.T., *IFPA Award in Placentology Lecture: Complicated interactions between genes and the environment in placentation, pregnancy outcome and long term health.* Placenta, 2010. **31 Suppl**: p. S47-53.

50. Andraweera, P.H., et al., *Association of vascular endothelial growth factor +936 C/T single-nucleotide polymorphism with pregnancies complicated by small-for-gestational-age babies.* Arch Pediatr Adolesc Med, 2011. **165**(12): p. 1123-30.

51. Brantsaeter, A.L., et al., *A dietary pattern characterized by high intake of vegetables, fruits, and vegetable oils is associated with reduced risk of preeclampsia in nulliparous pregnant Norwegian women.* J Nutr, 2009. **139**(6): p. 1162-8.

52. Oken, E., et al., *Diet During Pregnancy and Risk of Preeclampsia or Gestational Hypertension.* Ann Epidemiol, 2007. **17**(9): p. 663-668.

53. Chappell, L.C., et al., *Exploration and confirmation of factors associated with uncomplicated pregnancy in nulliparous women: prospective cohort study.* BMJ,

2013. **347**: p. f6398.

54. Bujold, E., S. Roberge, and K.H. Nicolaides, *Low-dose aspirin for prevention of adverse outcomes related to abnormal placentation.* Prenat Diagn, 2014.

55. Bezerra Maia, E.H.M.S., et al., *Prevention of preeclampsia.* J Pregnancy, 2012. **2012**: p. 435090.

56. Ebrashy, A., et al., *Usefulness of aspirin therapy in high-risk pregnant women with abnormal uterine artery Doppler ultrasound at 14-16 weeks pregnancy: randomized controlled clinical trial.* Croat Med J, 2005. **46**(5): p. 826-31.

57. Thangaratinam, S., et al., *Prediction and primary prevention of pre-eclampsia.* Best Pract Res Clin Obstet Gynaecol, 2011. **25**(4): p. 419-33.

58. Hofmeyr, G.J., et al., *Calcium supplementation during pregnancy for preventing hypertensive disorders and related problems.* Cochrane Database Syst Rev, 2010(8): p. CD001059.

59. Rembold, C.M., *Number needed to screen: development of a statistic for disease screening.* BMJ, 1998. **317**.

60. Poon, L.C., et al., *Hypertensive disorders in pregnancy: screening by biophysical and biochemical markers at 11-13 weeks.* Ultrasound Obstet Gynecol, 2010. **35**(6): p. 662-70.

61. Kusanovic, J.P., et al., *A prospective cohort study of the value of maternal plasma concentrations of angiogenic and anti-angiogenic factors in early pregnancy and midtrimester in the identification of patients destined to develop preeclampsia.* J Matern Fetal Neonatal Med, 2009. **22**(11): p. 1021-38.

62. Buhimschi, C.S., et al., *Urinary angiogenic factors cluster hypertensive disorders and identify women with severe preeclampsia.* Am J Obstet Gynecol, 2005. **192**(3): p. 734-41.

63. Poon, L.C., et al., *First-trimester prediction of hypertensive disorders in pregnancy.* Hypertension, 2009. **53**(5): p. 812-8.

64. Bahado-Singh, R.O., et al., *First-trimester metabolomic detection of late-onset preeclampsia.* Am J Obstet Gynecol, 2013. **208**(1): p. 58 e1-7.

65. Albaiges, G., et al., *One-stage screening for pregnancy complications by color Doppler assessment of the uterine arteries at 23 weeks' gestation.* Obstet Gynecol, 2000. **96**(4): p. 559-64.

Chapter 8: Manuscript III – Preterm Birth Model

Prediction for Spontaneous Preterm Birth: A Tiered Modelling Approach

Abstract

Background. PTB is the leading cause of perinatal morbidity and mortality. Currently PTB prediction is mostly based on obstetric risk factors and cervical length measurement. Predicting PTB in healthy nulliparous women at a single time point is still a major challenge. Hence, we propose a tiered prediction approach from multiple pregnancy stages using a combination of clinical predictors, including BMI, family history, lifestyle and dietary factors, and single nucleotide polymorphisms (SNPs) in several related candidate genes.

Methods. A total of 2432 participants from the Australian and New Zealand cohorts of the Screening for Pregnancy Endpoints (SCOPE) study were included in the analysis, with 123 SPTB cases and 2309 term births. Two models were developed based on maternal and paternal clinical predictors at 15 weeks' gestation and adding SNP predictors at 20 weeks' gestation. At initial screening, Tier 1 has a higher sensitivity, while Tier 2 has a higher positive predictive value during later stages of pregnancy. Prediction estimates were then integrated using Bayes' theorem.

Results. 1117 women (45.9%) were 'eliminated' as low risk at Tier 1 by 15 weeks of gestation, in which 15 SPTB cases (1.3%) were missed. By 20 weeks' gestation, at Tier 2, a further 259 women (10.7%) were identified at high risk of SPTB of whom 61 (23.6%) delivered preterm.

Conclusions. The tiered model provides a reasonable prediction for SPTB that 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 PTB.

Keywords: *Preterm birth, prediction, Bayes' theorem*

Introduction

Babies born preterm have a high risk of short-term or long-term morbidity, and even death[1-4]. It is estimated that 75%of neonatal mortality is due to preterm birth, and 50% of children who have long-term neurological impairment were born preterm[5-7]. With 500,000 neonatal deaths per year worldwide resulting from preterm birth (PTB)[8], it is one of the leading causes of perinatal morbidity and mortality.

Spontaneous preterm birth (SPTB) was defined as birth at less than 37 weeks of gestation that was not a result of medical or obstetric intervention. It accounts for approximately 60 to 70% of preterm births, and is most likely due to clinical or subclinical infective processes, cervical dysfunction, poor placentation, multiple gestation, and possibly, nutritional and environmental factors [9, 10].

Methods to identify women who are at risk of delivering preterm would be highly valued in the obstetric community, as early interventions or modifiable risk factors can be addressed to reduce the risk or severity of PTB[9, 11]. However, due to the multiple aetiology of PTB and previously identified complex gene-environment interactions[12, 13], predicting which women are at risk remains a major challenge. Current approaches based on maternal history have an estimated sensitivity and specificity of only 67%and 73% [14], and for cervical length measurements during mid-trimester, 52% and 82% [15].

There has been a marked increase over the last decade in clinical and statistical prediction models developed from prospective studies and many have shown that prediction with obstetric or genetic risk factors outperforms traditional approaches[14-16]. However, while some models show promising results, with sensitivity or specificity reaching over 90%, the majority of them rely on predictors in second trimester[17-23], and prediction during early pregnancy is only modest[24-26].

Since an early prediction of risk is desired but predictions at a single time point may not be satisfactory, this paper proposes a multi-model tiered approach, based on combinations of maternal and paternal clinical and genetic factors, with individual models tailored for each stage in pregnancy.

Methods

Models were developed from the Australian and New Zealand cohorts of the SCOPE (Screening fOr Pregnancy Endpoints) study, where nulliparous women with singleton pregnancies were recruited between November 2004 and August 2008. Ethical approval was obtained from local ethics committees [New Zealand AKX/02/00/364, Australia REC 1712/5/2008] and all women provided written informed consent [27].

Women were invited to participate prior to 15 weeks' gestation when attending hospital antenatal clinics, obstetricians, general practitioners or community midwives, and were

interviewed and examined by a research midwife at 15±1 and 20±1 weeks of gestation.

Women were excluded if they were considered to be at high risk of preeclampsia, SGA or PTB due to underlying medical conditions (e.g. chronic hypertension requiring antihypertensive medication or diabetes), previous cervical knife cone biopsy, 3 terminations or 3 miscarriages, current ruptured membranes, or their pregnancy was complicated by a known major fetal anomaly or abnormal karyotype, or if they received interventions that may modify pregnancy outcome (e.g. aspirin, cervical suture).

An internet accessed central database with a complete audit trial (Medscinet^{AB}, Stockholm, Sweden)[27] were used to store details of maternal history, dietary practices and clinical measurements at pre-pregnancy, 15 weeks and 20 weeks of gestation. At 15 weeks' gestation, blood samples were obtained from women. Partners' blood samples were also obtained at some time during the women's pregnancy, as well as cord blood at birth. DNA was extracted and genotype data were obtained for 100 candidate single nucleotide polymorphisms (SNPs) using the Sequenom Mass Array Platform as previously described[28].

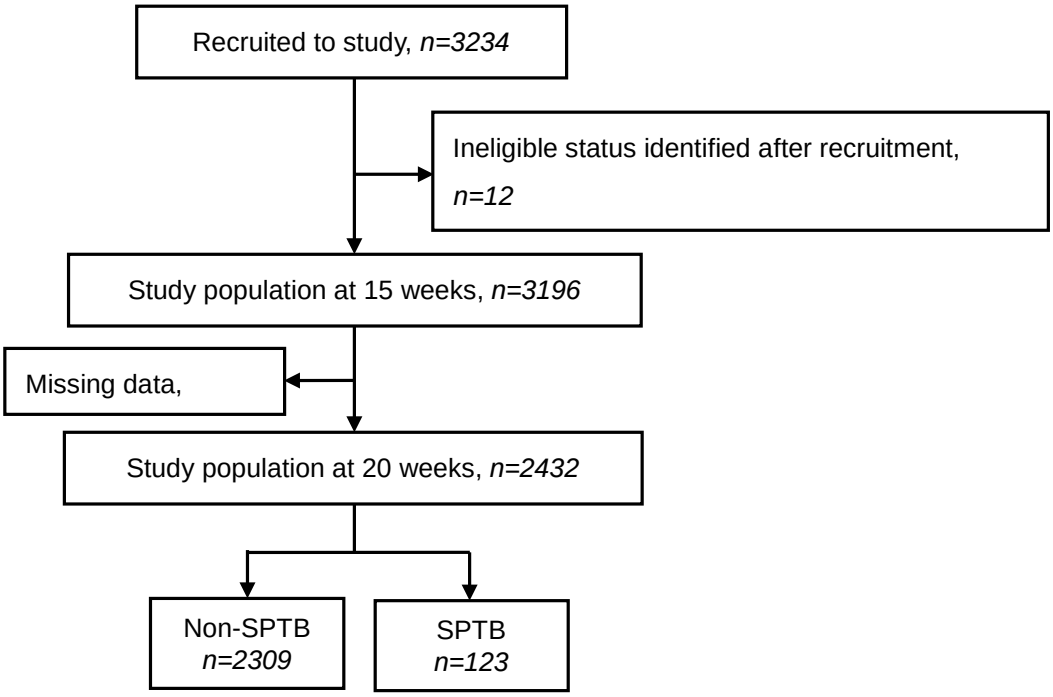

Figure 1. Participants recruited and study population

A total of 2432 participants were included in the analysis, with 123 SPTB and 2309 non-SPTB (Figure 1). Models on 2 time points: 15 weeks and 20 weeks' gestation, were developed using penalized Logistic regression, with variable selection based on Elastic-Net regularization[29, 30]. The sensitivity and specificity were calculated as measures of goodness of classification, and models were selected based on specific levels of sensitivity and specificity. The positive and negative predictive values were also obtained to assess predictive utility. Ten-fold cross validations were performed on all models using 90% of the data randomly chosen for training purposes, and validating on the remaining 10%.

Individual Model Specifications

Individual models are developed based on predictors collected at 15 weeks and 20 weeks of

gestation (Figure 2). For initial screening, at Tier 1, a high sensitivity is preferred, as the aim is to identify all patients at risk, and those who are predicted at risk (i.e. with positive test result) may benefit from more frequent monitoring. At this stage, the prediction will be based on predictors available at 15 weeks of gestation, which includes current dietary practice, pre-existing health conditions, family history, clinical measurements such as blood pressure, as well as SNP predictors.

For second tier prediction, which can be performed at 20 weeks of gestation, a higher positive predictive value (PPV), i.e. low false positive rate, is preferred to minimize the chance of unnecessary interventions. Predictors at this tier may include SNPs and transvaginal cervical length measurement.

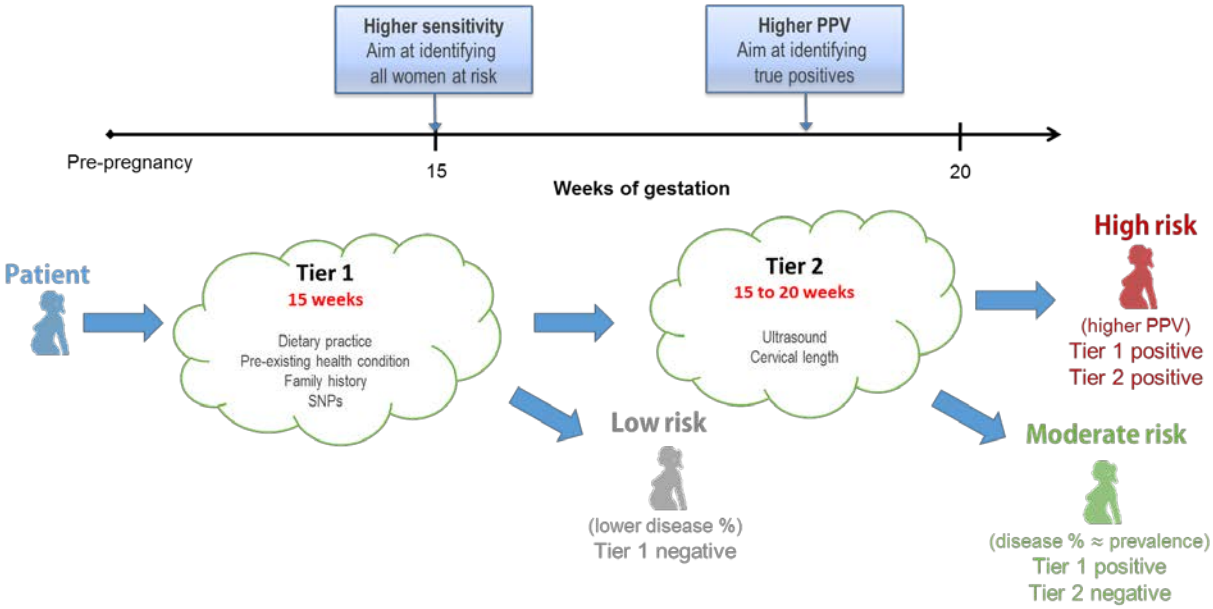

Figure 2. Tiered modelling approach.

Model Integration

For model integration, estimating a prior probability is needed to apply Bayes' theorem to obtain a post-test odds for Tier 2 based on the odds of prior 'guess' and the likelihood of current test. Bayes' theorem has been widely applied in areas of evidence-based medicine, and is also used in clinical decision support systems for individual patient risk estimation[31-33]. Here, a similar approach is applied, where the test results obtained at Tier 2 (20 weeks of gestation) is the post-test odds integrated with a pre-test probability obtained at Tier 1 (15 weeks of gestation).

Following Bayes' theorem[34]:

$$O\left(D\middle|T_{+/-}\right) = O(D)\frac{P\left(T_{+/-}\middle|D\right)}{P\left(T_{+/-}\middle|\overline{D}\right)} \quad (1)$$

The integrated post-test odds of SPTB at Tier 2, with pre-test odds obtained from Tier 1, are given by:

$$O_{\text{Tier 2}}\left(D\middle|T_{+/-}\right) = O_{\text{Tier 1}}(D) \cdot \Lambda_{\text{Tier 2}}\left(D\middle|T_{+/-}\right) \quad (2)$$

where $\Lambda_{\text{Tier 2}}\left(D\middle|T_{+/-}\right) = \frac{P\left(T_{+/-}\middle|D\right)}{P\left(T_{+/-}\middle|\overline{D}\right)}$is the likelihood of SPTB given a positive or negative test result for the current stage in pregnancy. This can be calculated from sensitivity and specificity of each test, where $\Lambda_{\text{Tier 2}}(D|T_{+}) = \frac{r}{1-s}and\Lambda_{\text{Tier 2}}(D|T_{-}) = \frac{1-r}{s}$.

Risk Classification

After the post-test odds for tier 2 is obtained, the predicted risk of all tiers will be analysed together and classify the risk of disease into 3 levels: low risk, moderate risk, and high risk

(Figure 2).

Women with a negative result at tier 1 will be considered as low risk, and do not need to go through further screening to tier 2. Since the sensitivity in tier 1 is high, the likelihood of disease in women who are predicted at low risk will be relatively low. For women who are predicted at risk in tier 1, further screening through tier 2 is recommended to identify individuals who are at high risk. Since low-risk women are already “eliminated” in tier 1, the sensitivity threshold may be relaxed in tier 2 to aim for a higher positive predictive value. Therefore, individuals who may be at higher risk (i.e. those who have a positive test result in both tier 1 and 2) may be further identified, amongst those who are predicted at risk.

As a result, the proportion of disease in the low-risk group (i.e. negative result in tier 1) will be lowest amongst the 3 risk groups, and will be at least lower than the current disease prevalence. Similarly, with a higher positive predictive value in the high-risk group, the proportion of disease will be highest, preferably more than 20%. Therefore, women with relatively lower risk are ‘eliminated’ at each tier, and tailored care may be provided according to their classified predicted risk.

Results

Of the 3234 nulliparous women recruited to the SCOPE study, follow up was complete in 3196 (98.8%) of participants (Figure 1). After omitting patients with any missing data, 123 SPTB

cases (5.06%) and 2309 non-SPTB were included in the analyses, in which 90% were used to train the two separate logistic regression models at 15 weeks and 20 weeks' gestation (Table 1).

Gravidity appears to be a consistent risk factor for all tiers, with an odds of 1.388 (95% CI 1.055-1.826; P 0.019) in Tier 1 and 1.504 (95% CI 1.109 - 2.041; P 0.009) in Tier 2. Anxiety is also a risk factor for both tiers, where a State-Trait Anxiety Inventory score[35, 36] of above 90th centile measured at 15 weeks' gestation increases the risk of SPTB 2-3 times (in Tier 1 OR 2.197, 95% CI 1.204 - 4.009, P 0.010; in Tier 2 OR 3.304, 95% CI 1.763 - 6.192, P 0.000). Every cm increase in maternal height has an estimated 3% reduced risk for SPTB in Tier 2 (at 20 weeks OR 0.959; 95% CI 0.927 - 0.992; P 0.016), while having a family history of a low birthweight baby (OR 1.627; 95% CI 1.031 - 2.568; P 0.037), SPTB (OR 1.591; 95% CI 0.911 - 2.779; P 0.103), or whether the participant's mother had PE (OR 2.014; 95% CI 0.971 - 4.178; P 0.060)increases the risk of SPTB. Despite the fact that some of the odds ratios crossed unity, they still contribute to the models.

Regarding variables related to lifestyle, low fruit consumption of less than 1 time per day during 1 month prior to pregnancy is a significant risk factor (OR 1.911; 95% CI 1.162 - 3.144; P 0.011), and using marijuana in 1st trimester significantly increases the risk of SPTB (OR 8.060; 95% CI 2.736 - 23.745; P 0.000), while having at least 800 µg of folate during 1st trimester reduces the risk of SPTB (OR 0.339; 95% CI 0.109 - 1.053; P 0.061).

At 15 weeks' gestation, using other recreational drugs or binge alcohol consumption of more than 6 units per session is a strong risk factor (OR 4.341; 95% CI 1.855 - 10.160; P 0.001), and climbing stairs more than 10 times per day at 15 weeks' gestation also increases the risk of SPTB by 2-3 times (at Tier 1 OR 2.270, 95% CI 1.282 - 4.021, P 0.005; at Tier 2 OR 3.436, 95% CI 1.819 - 6.489, P 0.000).

Interestingly, only maternal SNPs were included in the final model, with 13 SNPs in Tier 1 and 7 in Tier 2. SNPs in AGT, TCN2, uPA, IGF1R, MMP2, MMP9, and TIMP3 appear to be predictive in both tiers/timepoints (Table 1).

Table 1. Predictors for Tier 1 and 2

	Tier 1 (15 weeks of gestation)		*Tier 2 (20 weeks of gestation)*	
Predictors	**Odds (95% CI)**	**P-value**	**Odds (95% CI)**	**P-value**
Height (maternal)	0.975 (0.946 - 1.005)	0.0994	0.959 (0.927 - 0.992)	***0.0155***
BMI (maternal)	-	-	1.014 (0.977 - 1.053)	0.4508
Years of schooling	-	-	0.904 (0.757 - 1.080)	0.2673
Gravidity	1.388 (1.055 - 1.826)	***0.0192***	1.504 (1.109 - 2.041)	***0.0087***
Months to conceive	-	-	1.016 (1.000 - 1.033)	0.0569
Other recreational drug use (at 15 wks)	4.341 (1.855 - 10.160)	***0.0007***	-	-
Folate dose >800µg per day (at 1st trim)	0.339 (0.109 - 1.053)	0.0614	-	-
Fruit consumption (<1x/day at 1mth pre-preg)	1.911 (1.162 - 3.144)	***0.0108***	-	-
Marijuana (>1/day at 1st trim)	-	-	8.060 (2.736 - 23.745)	***0.0002***
Climbing stairs (>10x/day at 15wks)	2.270 (1.282 - 4.021)	***0.0049***	3.436 (1.819 - 6.489)	***0.0001***
State-Trait Anxiety Inventory (>90th centile at 15wks)	2.197 (1.204 - 4.009)	***0.0104***	3.304 (1.763 - 6.192)	***0.0002***
Living with relatives	-	-	3.824 (1.277 - 11.455)	***0.0165***
Not feeling better	1.893 (1.069 - 3.352)	***0.0286***	-	-
Cervical length (at 20 wks)	-	-	1.040 (1.010 - 1.071)	***0.0133***
Hospital admission due to Hyperemesis	2.438 (0.905 - 6.564)	0.0779	-	-
Any LLETZ treatment	2.533 (1.111 - 5.773)	***0.0270***	-	-
Metformin for PCOS (at conception)	2.732 (0.850 - 8.782)	0.0916	-	-
FH (LBW baby)‡	1.627 (1.031 - 2.568)	***0.0365***	-	-
FH (SPTB)*	-	-	1.591 (0.911 - 2.779)	0.1030
Participant's mother had PE (1x)	-	-	2.014 (0.971 - 4.178)	0.0602
Participant's mother had PE (>=2x)	-	-	2.974 (1.058 - 8.356)	***0.0387***
AGT (maternal)[SNP]	3.653 (1.134 - 11.766)	***0.0300***	3.259 (0.992 - 10.712)	***0.0340***

ADD1 (maternal)[SNP]	1.356 (0.902 - 2.038)	0.1433	-	-
BCL2 (maternal)[SNP]	1.497 (0.991 - 2.261)	0.0555	-	-
MBL2 (maternal)[SNP]	2.591 (0.846 - 7.933)	0.0955	-	-
TCN2 (maternal)[SNP]	1.455 (0.973 - 2.176)	0.0681	1.535 (0.978 - 2.407)	***0.0431***
FLT1 (maternal)[SNP]	2.533 (0.605 - 10.611)	0.2034	-	-
IGF2R (maternal)[SNP]	1.505 (1.022 - 2.217)	***0.0382***	-	-
IL1B (maternal)[SNP]	1.357 (0.913 - 2.019)	0.1314	-	-
uPA (maternal)[SNP]	2.214 (1.176 - 4.169)	***0.0139***	3.347 (1.727 - 6.487)	***0.0004***
IGF1R (maternal)[SNP]	1.403 (0.954 - 2.063)	0.0856	1.667 (1.085 - 2.560)	***0.0147***
MMP2 (maternal)[SNP]	1.844 (1.251 - 2.718)	***0.0020***	1.974 (1.286 - 3.030)	***0.0020***
MMP9 (maternal)[SNP]	1.659 (1.040 - 2.645)	***0.0337***	1.655 (0.996 - 2.751)	0.0623
TIMP3 (maternal)[SNP]	1.575 (0.908 - 2.732)	0.1063	-	-

*‡ Family history of low birthweight baby; *Family history of Spontaneous preterm birth*

Integrated Models

Of the 2432 patients analysed (Table 2), 46% were eliminated as not at risk for SPTB at first screening (Tier 1) at15 weeks of gestation, leaving 54% (1315 patients) patients requiring Tier 2 screening. These may be recommended for a higher level of attention. By Tier 2 at 20 weeks of gestation, a further 35.3% are eliminated based on new predictors at this stage, with 259 identified at high risk of SPTB and may be monitored more closely. Since the prediction at Tier 2 aims for a higher PPV, 1 in 5 patients (23.55%) identified at high risk and recommended for a higher level of care eventually delivered preterm.

Table 2. Final risk classification.

Tier 1	Tier 2	Classification	Predicted	Observed	SPTB %
-		Low risk	1117	15	1.34%
+	-	Moderate risk	1056	47	4.45%
+	+	High risk	259	61	23.55%
		Overall		123	5.06%

As expected, the sensitivity at Tier 1 would be higher, in which 108 out of 123 SPTB cases (87.8%) were identified at risk. Amongst the 1117 women ‘eliminated’ at first tier, 15 SPTB cases (1.34%) were missed, resulting in a negative predictive value of 98.7%. The proportion of SPTB increases for predicted higher risk groups.

Interestingly, women predicted to be at low risk delivered babies with a higher birthweight and gestational age on average, compared to those predicted at risk

(Figure 3). When comparing amongst SPTB cases, apart from one outlier with a birthweight less than 1000 g, the average birthweight of the low-risk group is 2572g, and the majority of low birthweight babies were successfully predicted at risk (Table 3b). The gestational age of the low-risk group is also longer, with an average of 35 weeks, and only 1 (6.67%) delivered before 28 weeks. However, although the high-risk group successfully identified some outliers for early SPTB, its prediction for early SPTB is only modest, and the moderate-risk group had a lower birthweight and shorter gestation compared to the high-risk group, within observed SPTB cases. Nevertheless, the likelihood of early SPTB amongst women identified as low-risk at Tier 1 is very low, with only 6 (0.54%) SPTB cases in all women predicted as low risk, and that the overall proportion of early SPTB is only 1.14% (12 cases) in the moderate-risk group (Table 3a).

Table 3a. Characteristics for overall cohort.

		Birthweight (g)	**Gestational age (wks)‡**			
Risk	**N**	**Mean ± SEM**	**Mean ± SEM**	**<28 wks (n=21)**	**<32wks (n=37)**	**<37 wks (n=203)**
Low	1117	3439 ± 11	39.7 ± 0.04	6 (0.54%)	9 (0.81%)	50 (4.47%)
Moderate	1056	3370 ± 12	39.4 ± 0.05	12 (1.14%)	19 (1.80%)	85 (8.05%)
High	259	3145 ± 14	38.4 ± 0.06	3 (1.16%)	9 (3.48%)	68 (26.3%)

‡includes 80 Iatrogenic PTB (35 Preeclampsia cases, 32 Small for gestational age, and 17 other complications)

Table 3b. Characteristics **within SPTB cases**.

Risk	N	Birthweight (g)	Gestational age (wks)			
		Mean ± SEM	Mean ± SEM	<28 wks (n=11)	<32wks (n=21)	<37 wks (n=123)
Low	15	2572 ± 62	35.0 ± 0.31	1 (6.67%)	1 (6.67%)	15 (100%)
Moderate	47	2197 ± 75	32.5 ± 0.42	7 (14.89%)	12 (25.53%)	47 (100%)
High	61	2418 ± 61	34.6 ± 0.27	3 (4.92%)	8 (13.12%)	61 (100%)

It is worth noting that the total number of deliveries before 37 weeks' gestation in the overall cohort was higher than that of women who experienced SPTB, as there were 80 iatrogenic PTBs included (in which 35 had Preeclampsia, 32 small for gestational age, and 17 with other complications). Interestingly, 35 out of 50 (70%) who delivered before 37 weeks' gestation amongst the low-risk group were iatrogenic PTB, and 89.7% of the high-risk group were spontaneous PTB.

Discussion

Our data have demonstrated that a tiered approach may be applied to enhance prediction by integrating risk estimates from models of different specifications. This will not only allow risk estimation or prediction at various time points, but also constant monitoring and update of predicted risk for individuals when new predictors are available or when conditions change, and hence, the level of care may be tailored for individual women.

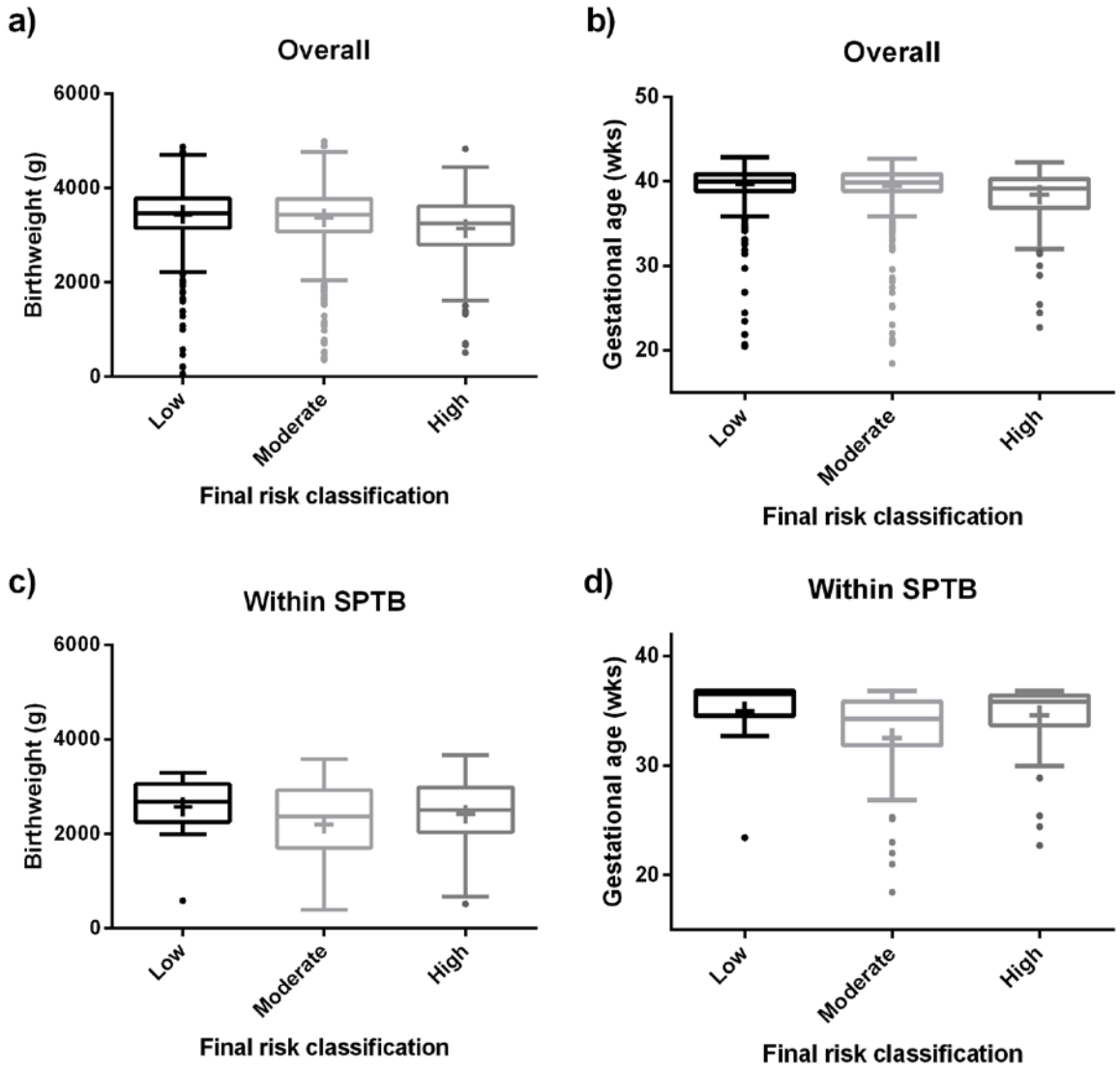

Figure 3. Final risk classification by a) birthweight (g), and b) by gestational age (wks) for all women (*includes 80 Iatrogenic PTB; 35 in low risk group*), and c) birthweight (g), d) by gestational age (wks) within observed SPTB cases.

Most clinical predictors in the models are consistent with risk factors for SPTB previously described by Murphy[12] and Dekker et al[37]. These include gravidity, family history of preterm birth, marijuana use, stress, and previous LLETZ treatment [37-39]. Similar studies have found that women with siblings born preterm have an increased

risk of giving birth preterm [39], and that the odds of PTB for nulliparous women who had one miscarriage is estimated to be 1.13, which increases further to 2.46 for women with three or more previous miscarriages [40-42].

Marijuana use is another known factor associated with adverse pregnancy outcomes, with studies showing an increased risk of PTB and SGA in women exposed to marijuana during pregnancy, with an odds of at least 1.5 adjusted for age, BMI, and smoking [43-46]. Our results from this study have also shown that a high marijuana use of more than once per day during first trimester significantly increases the odds of SPTB (OR 8.06; 95% CI 2.736 - 23.745).

It is well recognised that marijuana use is associated with low socio-economic status and stress [47-50], which are also known risk factors for PTB. It has also previously been shown that the incidence of preterm birth is higher amongst women with lower income and lower educational status[51]. Although socio-economic status was not included in the final models, measures of stress and anxiety were predictive for SPTB. Women with a State-Trait Anxiety Inventory (STAI) score [35, 36] of above the 90th centile, assessed at 15 weeks' gestation, have an estimated odds of 3.304 (95% CI 1.763 - 6.192). Self-reported 'not feeling better' also increases the risk of SPTB (odds 1.893; 95% CI 1.069 - 3.352). This is consistent with similar studies in which stress was

significantly associated with SPTB, and the estimated odds in women who had a higher STAI score was 4.8 [52-55].

Other factors such as LLETZ treatment is also associated with SPTB, with large prospective studies in Denmark and UK reporting a higher incidence of SPTB in women who have previously undergone LLETZ treatment[56, 57]. Interestingly, while moderate exercise (such as walking, yoga, or water aerobics) during pregnancy has been shown to reduce the risk of PTB with an odds of 0.91 [58], climbing stairs more than 10 times per day increased risk (odds 2.27; 95% CI 1.282 - 4.021), and is consistent with a similar study [59].

Regarding genetic risk factors, the most consistent gene reported to be associated with PTB and present in the tiered model is Interleukin-1β (*IL1β*), which encodes a pro-inflammatory cytokine that affects gestational tissues[11, 60, 61]. Other genes, including alpha-adducin (ADD1)[62], angiotensinogen (AGT)[63, 64], urokinase-plasminogen activator (uPA)[65], and matrix metallopeptidase (*MMP*)[66-68] have also been studied, but there are inconsistent results in the literature. Interestingly, other studies have reported type 1 insulin-like growth factor receptor (IGF1R)[69] and mannose-binding lectin 2 (MBL2)[70] to be associated with SPTB, but in fetal genotypes.

Application of Tiered Prediction Model

By obtaining risk estimates at each stage, the tiered prediction system can be used as a process of elimination to classify patients at risk for SPTB (Figure 4). As the initial Tier 1 screening at 15 weeks' gestation has a high sensitivity, women with potential risk of SPTB are likely to be identified at this stage, and the likelihood of SPTB given a negative test result is relatively low (NPV 98.7%). Hence, women predicted as low risk can be 'eliminated' from Tier 2 screening and continue regular antenatal visits. At this stage, modifiable risk factors including fruit consumption, folate supplementation and exercise, may be addressed in women identified at risk. Other studies have reported that an increased intake of fruit and vegetables, as well as obtaining a balanced diet, and having moderate exercise during pregnancy decreases the risk of PTB [58, 71, 72]. In addition, women predicted at risk may also benefit from more frequent monitoring.

By 18 to 20 weeks' gestation, cervical length measurement can be obtained at the mid-pregnancy transvaginal morphology scan which, in combination with other factors available, comprises the Tier 2 screening. The main aim of this tier is to identify a high-risk group amongst those predicted at risk, in which the likelihood of SPTB is highest (with PPV above 20%). While patients with negative test results at Tier 2 are

considered at moderate risk, frequent monitoring may still be beneficial, as there is still a potential risk for SPTB (estimated 4.45% SPTB cases in moderate-risk group).

At Tier 2 screening, up to 20 weeks' gestation, patients identified at high risk may benefit from a higher level of care and/or secondary prevention or intervention. Preventive strategies such as vaginal progesterone may still be administered, although a recent review reported that the optimal gestational age to commence treatment is uncertain[73], studies have shown that vaginal progesterone from 24 to 34 weeks' gestation also reduces the rate of SPTB [74-78].

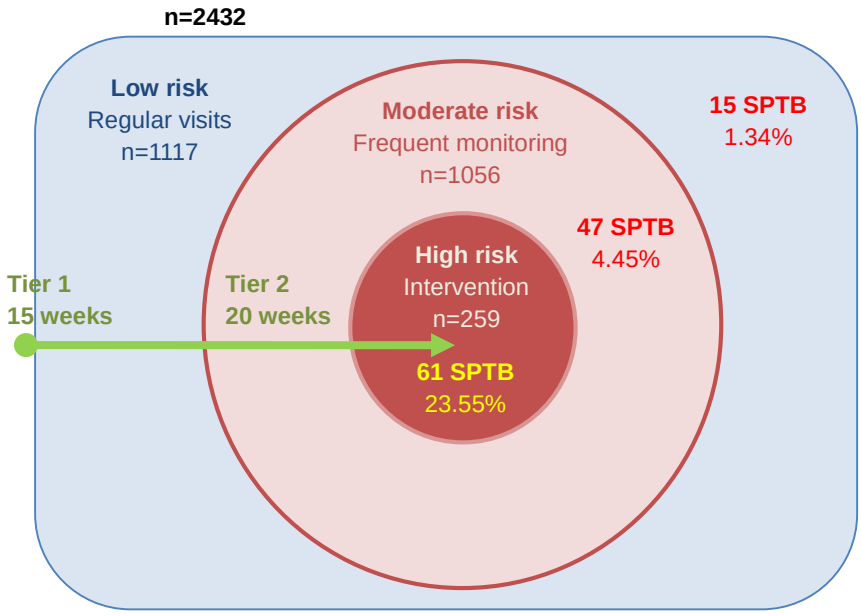

Figure 4. Process of elimination using tiered model.

Of the 2432 participants included in this analysis, 1315 women (54.1%) were predicted to be at risk at Tier 1 screening at 15 weeks' gestation. At Tier 2, up to 20 weeks' gestation, 259 women were further identified, with 61 (23.55%) of SPTB cases

identified. This means that up to 45.9% of women are ‘eliminated’ as low risk for SPTB, with only 1.34% SPTB cases missed.

Assuming all patients identified as high risk in Tier 2 are treated, with an estimated risk difference compared to non-high-risk patients of 20.7%, the number needed to treat at Tier 2 to prevent one SPTB case is 4.8. The estimated number needed to screen[79] is 169 to prevent one SPTB.

Strengths and Limitations

A major strength of the tiered model is the ability to classify risk from different tiers comprising different suites of risk factors. Thus the chance of unnecessary interventions may be minimized in patients with a negative test result ‘eliminated’ at initial screening. More importantly, modifiable risk factors may be addressed after Tier 1 screening, and preventive strategies may still be applied after Tier 2 screening to reduce the risk or severity of SPTB. In addition, since the models have been developed from a cohort of nulliparous women, the tiered prediction model may apply to women with no or unknown history of PTB.

Although the PPV may not outperform screenings from sequential fetal fibronectin at 24 to 26 weeks’ gestation in women who had a short cervix with or without uterine contractions[80, 81], the tiered model achieved a PPV of 23.55% in asymptomatic women

by 20 weeks' gestation. Thus our model has utility for early prediction in asymptomatic nulliparous women. By implementing a two-step screening process, with a high sensitivity in Tier 1 for initial screening and a higher PPV in Tier 2 dedicated to identifying high risk women, limitations of single-model predictions such as choosing the optimal threshold for sensitivity and specificity levels may be avoided.

However, the main limitation of the tiered model is the uncertainty in the moderate-risk group. With the proportion of SPTB similar to the current prevalence, at this stage, patients predicted as at moderate-risk would rely on frequent monitoring, and further research will be needed to improve the prediction in this group of patients.

Conclusion

Through a tiered integrated prediction system, the prediction of SPTB is further enhanced, as it allows for regular monitoring and revision of predicted risk throughout pregnancy. By permitting classification of patients into various levels of care, the tiered model may also assist in providing tailored antenatal care or interventions that could benefit both the mother and child. It may also be useful for avoiding unnecessary interventions for low-risk women. Finally, we have also identified modifiable predictors that could also be addressed to reduce the risk or severity of PTB.

References

1. Lepercq, J., et al., *Factors associated with preterm delivery in women with type 1 diabetes: a cohort study.* Diabetes Care, 2004. **27**(12): p. 2824-8.

2. Hibbard, J.U., et al., *Respiratory morbidity in late preterm births.* JAMA, 2010. **304**(4): p. 419-25.

3. Platzker, A.C., *Assessment of the risk of the respiratory distress syndrome with preterm delivery.* Calif Med, 1972. **117**(2): p. 49.

4. Navaei, F., et al., *Early outcome of preterm infants with birth weight of 1500 g or less and gestational age of 30 weeks or less in Isfahan city, Iran.* World J Pediatr, 2010: p. ahead of print.

5. Kramer, M.S., et al., *The contribution of mild and moderate preterm birth to infant mortality. Fetal and Infant Health Study Group of the Canadian Perinatal Surveillance System.* JAMA, 2000. **284**(7): p. 843-9.

6. Mikkola, K., et al., *Neurodevelopmental outcome at 5 years of age of a national cohort of extremely low birth weight infants who were born in 1996-1997.* Pediatrics, 2005. **116**(6): p. 1391-400.

7. Den Ouden, A.L., et al., *The relation between neonatal thyroxine levels and neurodevelopmental outcome at age 5 and 9 years in a national cohort of very preterm and/or very low birth weight infants.* Pediatr Res, 1996. **39**(1): p. 142-5.

8. Child Health Research Project, *Reducing Perinatal and Neonatal Mortality* in *Special Report*, R.E. Black and L. Kelley, Editors. 1999.

9. Honest, H., et al., *Screening to prevent spontaneous preterm birth: systematic reviews of accuracy and effectiveness literature with economic modelling.* Health Technol Assess, 2009. **13**(43): p. 1-627.

10. Lumley, J., *Defining the problem: the epidemiology of preterm birth.* BJOG, 2003. **110 Suppl 20**: p. 3-7.

11. Committee on Understanding Premature Birth and Assuring Healthy Outcomes, *Preterm Birth: Causes, Consequences, and Prevention*, ed. R.E. Behrman and A.S. Butler. 2007, Washington DC: National Academies Press.

12. Murphy, D.J., *Epidemiology and environmental factors in preterm labour.* Best Practice & Research Clinical Obstetrics & Gynaecology, 2007. **21**(5): p. 773-789.

13. Esplin, M.S. and M.W. Varner, *Genetic factors in preterm birth--the future.* BJOG, 2005. **112 Suppl 1**: p. 97-102.

14. Iams, J.D., et al., *The Preterm Prediction Study: recurrence risk of spontaneous preterm birth. National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network.* Am J Obstet Gynecol, 1998. **178**(5): p. 1035-40.

15. Fuchs, F., et al., *Ultrasound cervical length measurement for prediction of delivery before 32 weeks in women with emergency cerclage for cervical insufficiency.* Int J Gynaecol Obstet, 2010. **110**(3): p. 245-8.

16. Blondel, B., et al., *Prediction of preterm delivery: is it substantially improved by routine vaginal examinations?* Am J Obstet Gynecol, 1990. **162**(4): p. 1042-8.

17. Holst, R.M., et al., *Prediction of spontaneous preterm delivery in women with preterm labor: analysis of multiple proteins in amniotic and cervical fluids.* Obstet Gynecol, 2009. **114**(2 Pt 1): p. 268-77.

18. Hamilton, M.P., H.I. Abdalla, and C.R. Whitfield, *Significance of raised maternal serum alpha-fetoprotein in singleton pregnancies with normally formed fetuses.* Obstet Gynecol, 1985. **65**(4): p. 465-70.

19. Pieta-Dolinska, A., et al., *[Assessment of fetal fibronectin concentration in identifying patients at risk for preterm delivery].* Ginekol Pol, 2005. **76**(6): p. 431-5.

20. Paternoster, D., et al., *Phosphorylated insulin-like growth factor binding protein-*

1 in cervical secretions and sonographic cervical length in the prediction of spontaneous preterm delivery. Ultrasound Obstet Gynecol, 2009. **34**(4): p. 437-40.

21. Heine, R.P., et al., *Serial salivary estriol to detect an increased risk of preterm birth.* Obstet Gynecol, 2000. **96**(4): p. 490-7.

22. Lockwood, C.J., et al., *Increased interleukin-6 concentrations in cervical secretions are associated with preterm delivery.* Am J Obstet Gynecol, 1994. **171**(4): p. 1097-102.

23. Holzman, C., et al., *Second trimester corticotropin-releasing hormone levels in relation to preterm delivery and ethnicity.* Obstet Gynecol, 2001. **97**(5 Pt 1): p. 657-63.

24. Elaveyini, U., S.P. Devi, and K.S. Rao, *Neural networks prediction of preterm delivery with first trimester bleeding.* Arch Gynecol Obstet, 2010. **283**(5): p. 971-979.

25. Weiss, G., et al., *Elevated first-trimester serum relaxin concentrations in pregnant women following ovarian stimulation predict prematurity risk and preterm delivery.* Obstet Gynecol, 1993. **82**(5): p. 821-8.

26. Offenbacher, S., et al., *Maternal periodontitis and prematurity. Part I: Obstetric outcome of prematurity and growth restriction.* Ann Periodontol, 2001. **6**(1): p. 164-74.

27. McCowan, L., R. North, and R. Taylor. *ACTRN12607000551493*. 2007; Available from: www.anzctr.org.autrialSearch.aspx.

28. Andraweera, P.H., et al., *Single-nucleotide polymorphisms in the KDR gene in pregnancies complicated by gestational hypertensive disorders and small-for-gestational-age infants.* Reprod Sci, 2012. **19**(5): p. 547-54.

29. Zou, H. and T. Hastie, *Regularization and variable selection via the Elastic Net.* Journal of Royal Statistical Society, Series B, 2005. **67**(2): p. 301-320.

30. Wu, T.T., et al., *Genome-wide association analysis by lasso penalized logistic regression.* Bioinformatics, 2009. **25**(6): p. 714-721.

31. Lindgaard, G., et al., *Does Bayes have it? Decision Support Systems in diagnostic medicine.* International Journal of Industrial Ergonomics, 2009. **39**(3): p. 524-532.

32. Hall, G.H., *The clinical application of Bayes' theorem.* The Lancet, 1967. **290**(7515): p. 555-557.

33. Round, A., *Introduction to clinical reasoning.* Journal of Evaluation in Clinical Practice, 2001. **7**(2): p. 109-117.

34. Aitken, C.G. and D.A. Stoney, *The use of statistics in Forensic Science*. Ellis Horwood Series in Forensic Science. 1991: Ellis Horwood.

35. Spielberger, C.D., et al., *Manual for the State-Trait Anxiety Inventory*. 1983, Palo Alto, CA: Consulting Psychologists Press.

36. Spielberger, C.D., *State-Trait Anxiety Inventory*, in *The Corsini Encyclopedia of Psychology*. 2010, John Wiley & Sons, Inc.

37. Dekker, G.A., et al., *Risk factors for preterm birth in an international prospective cohort of nulliparous women.* PLoS One, 2012. **7**(7): p. e39154.

38. McCowan, L.M., et al., *Spontaneous preterm birth and small for gestational age infants in women who stop smoking early in pregnancy: prospective cohort study.* BMJ, 2009. **338**: p. b1081.

39. Bhattacharya, S., et al., *Inherited predisposition to spontaneous preterm delivery.* Obstet Gynecol, 2010. **115**(6): p. 1125-33.

40. Tulppala, M., et al., *A prospective study of 63 couples with a history of recurrent spontaneous abortion: Contributing factors and outcome of subsequent pregnancies.* Human Reproduction, 1993. **8**(5): p. 764-770.

41. Hammoud, A.O., et al., *Recurrent pregnancy loss and obstetric outcome.*

International Journal of Gynecology and Obstetrics, 2007. **96**(1): p. 28-29.

42. Shapira, E., et al., *Primary vs. secondary recurrent pregnancy loss - Epidemiological characteristics, etiology, and next pregnancy outcome.* J Perinat Med, 2012. **40**(4): p. 389-396.

43. Hatch, E.E. and M.B. Bracken, *Effect of marijuana use in pregnancy on fetal growth.* Am J Epidemiol, 1986. **124**(6): p. 986-93.

44. Hayatbakhsh, M.R., et al., *Birth outcomes associated with cannabis use before and during pregnancy.* Pediatric Research, 2012. **71**(2): p. 215-219.

45. Black, M., et al., *Outcomes of pregnancy in women using illegal drugs and in women who smoke cigarettes.* Acta Obstet Gynecol Scand, 2013. **92**(1): p. 47-52.

46. Kennare, R., A. Heard, and A. Chan, *Substance use during pregnancy: risk factors and obstetric and perinatal outcomes in South Australia.* Aust N Z J Obstet Gynaecol, 2005. **45**(3): p. 220-5.

47. Tehranifar, P., et al., *Life course socioeconomic conditions, passive tobacco exposures and cigarette smoking in a multiethnic birth cohort of U.S. women.* Cancer Causes Control, 2009. **20**(6): p. 867-76.

48. Graham, H., S.S. Hawkins, and C. Law, *Lifecourse influences on women's smoking before, during and after pregnancy.* Soc Sci Med, 2010. **70**(4): p. 582-7.

49. Passey, M.E., et al., *Tobacco, alcohol and cannabis use during pregnancy: Clustering of risks.* Drug Alcohol Depend, 2014.**134**: p. 44-50.

50. Kendzor, D.E., et al., *Mechanisms linking socioeconomic disadvantage and BMI in smokers.* Am J Health Behav, 2013. **37**(5): p. 587-98.

51. Begum, F., K. Buckshe, and J.N. Pande, *Risk factors associated with preterm labour.* Bangladesh Med Res Counc Bull, 2003. **29**(2): p. 59-66.

52. Dayan, J., et al., *Role of anxiety and depression in the onset of spontaneous preterm labor.* Am J Epidemiol, 2002. **155**(4): p. 293-301.

53. Copper, R.L., et al., *The preterm prediction study: maternal stress is associated with spontaneous preterm birth at less than thirty-five weeks' gestation. National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network.* Am J Obstet Gynecol, 1996. **175**(5): p. 1286-92.

54. Pavlov, M., et al., *Obstetric and neonatal outcome in patients with anxiety disorders.* J Matern Fetal Neonatal Med, 2013.

55. Ibanez, G., et al., *Depression and anxiety in women during pregnancy and neonatal outcome: data from the EDEN mother-child cohort.* Early Hum Dev, 2012. **88**(8): p. 643-9.

56. Poon, L.C., et al., *Large loop excision of transformation zone and cervical length in the prediction of spontaneous preterm delivery.* BJOG, 2012. **119**(6): p. 692-8.

57. Kyrgiou, M., et al., *Obstetric outcomes after conservative treatment for intraepithelial or early invasive cervical lesions: systematic review and meta-analysis.* Lancet, 2006. **367**(9509): p. 489-98.

58. Guendelman, S., et al., *Association between preterm delivery and pre-pregnancy body mass (BMI), exercise and sleep during pregnancy among working women in Southern California.* Matern Child Health J, 2013. **17**(4): p. 723-731.

59. Misra, D.P., et al., *Effects of physical activity on preterm birth.* Am J Epidemiol, 1998. **147**(7): p. 628-635.

60. Genc, M.R., et al., *Polymorphism in the interleukin-1 gene complex and spontaneous preterm delivery.* Am J Obstet Gynecol, 2002. **187**(1): p. 157-63.

61. Varner, M.W. and M.S. Esplin, *Current understanding of genetic factors in preterm birth.* BJOG, 2005. **112 Suppl 1**: p. 28-31.

62. Gibson, C.S., et al., *Genetic polymorphisms and spontaneous preterm birth.* Obstet Gynecol, 2007. **109**(2 Pt 1): p. 384-91.

63. Valdez-Velazquez, L.L., et al., *Genetic polymorphisms of the renin-angiotensin system in preterm delivery and premature rupture of membranes.* J Renin Angiotensin Aldosterone Syst, 2007. **8**(4): p. 160-8.

64. Gargano, J.W., et al., *Polymorphisms in thrombophilia and renin-angiotensin system pathways, preterm delivery, and evidence of placental hemorrhage.* Am J Obstet Gynecol, 2009. **201**(3): p. 317 e1-9.

65. Haedersdal, S., et al., *Inflammatory markers in the second trimester prior to clinical onset of preeclampsia, intrauterine growth restriction, and spontaneous preterm birth.* Inflammation, 2013. **36**(4): p. 907-13.

66. Sundrani, D., et al., *Matrix metalloproteinases-2, -3 and tissue inhibitors of metalloproteinases-1, -2 in placentas from preterm pregnancies and their association with one-carbon metabolites.* Reproduction, 2013. **145**(4): p. 401-10.

67. Kim, A., et al., *Identification of biomarkers for preterm delivery in mid-trimester amniotic fluid.* Placenta, 2013. **34**(10): p. 873-8.

68. Menon, R., et al., *Racial disparity in pathophysiologic pathways of preterm birth based on genetic variants.* Reprod Biol Endocrinol, 2009. **7**: p. 62.

69. Haataja, R., et al., *Mapping a new spontaneous preterm birth susceptibility gene, IGF1R, using linkage, haplotype sharing, and association analysis.* PLoS Genet, 2011. **7**(2): p. e1001293.

70. Bodamer, O.A., et al., *Evidence for an association between mannose-binding lectin 2 (MBL2) gene polymorphisms and pre-term birth.* Genet Med, 2006. **8**(8): p. 518-24.

71. Englund-Ogge, L., et al., *Maternal dietary patterns and preterm delivery: results from large prospective cohort study.* BMJ, 2014. **348**: p. g1446.

72. Grieger, J.A., L.E. Grzeskowiak, and V.L. Clifton, *Preconception Dietary Patterns in Human Pregnancies Are Associated with Preterm Delivery.* J Nutr, 2014.

73. Dodd, J.M., et al., *Prenatal administration of progesterone for preventing preterm birth in women considered to be at risk of preterm birth.* Cochrane Database Syst Rev, 2013. **7**: p. CD004947.

74. Romero, R., et al., *Progesterone to prevent spontaneous preterm birth.* Semin Fetal Neonatal Med, 2014. **19**(1): p. 15-26.

75. Fonseca, E.B., et al., *Progesterone and the risk of preterm birth among women with a short cervix.* N Engl J Med, 2007. **357**(5): p. 462-9.

76. Cetingoz, E., et al., *Progesterone effects on preterm birth in high-risk pregnancies: a randomized placebo-controlled trial.* Arch Gynecol Obstet, 2011. **283**(3): p. 423-9.

77. Romero, R., et al., *Vaginal progesterone in women with an asymptomatic sonographic short cervix in the midtrimester decreases preterm delivery and neonatal morbidity: a systematic review and metaanalysis of individual patient data.* Am J Obstet Gynecol, 2012. **206**(2): p. 124.e1-124.e19.

78. da Fonseca, E.B., et al., *Prophylactic administration of progesterone by vaginal suppository to reduce the incidence of spontaneous preterm birth in women at increased risk: a randomized placebo-controlled double-blind study.* Am J Obstet Gynecol, 2003. **188**(2): p. 419-24.

79. Rembold, C.M., *Number needed to screen: development of a statistic for disease screening.* BMJ, 1998. **317**.

80. Sananes, N., et al., *Prediction of spontaneous preterm delivery in singleton pregnancies: Where are we and where are we going? A review of literature.* J Obstet Gynaecol, 2014.

81. Iams, J.D., *Prediction and early detection of preterm labor.* Obstet Gynecol, 2003. **101**(2): p. 402-12.

Chapter 9: Discussion and Future Work

9.1. Final Discussion

It has long been a challenge to identify nulliparous women at risk of PE and SPTB, two leading causes of maternal and perinatal morbidity and mortality. Currently there is a need for accurate screening methods during the early stages of pregnancy to initiate effective early preventative strategies to reduce the risk or severity of disease. The aim of this study was to develop prediction tests for PE and SPTB using combinations of clinical and genotype predictors. The results have demonstrated that the prediction of preeclampsia and preterm birth is enhanced using combinations of clinical and SNP predictors through mathematical modelling. In particular, the tiered modelling approach provided results that outperform traditional predictions based on maternal history or ultrasound studies (Blondel et al., 1990; Iams et al., 1998; Papageorghiou et al., 2005; Fuchs et al., 2010; Jacquemyn et al., 2010).

The main outcomes and contributions of this study are:

- the prediction for PE and SPTB can be further enhanced with combinations of clinical and genotype predictors
- the novel tiered approach has demonstrated results that outperform traditional approaches
- initial screening from tiered models for PE and SPTB is available at 15 weeks' gestation, where modifiable risk factors can still be addressed
- preventative strategies may still be applicable after Tier 2 screening, at 15 weeks of gestation for PE and at 20 weeks of gestation for SPTB. By classifying risk levels, tailored care may be provided for individuals, and patients at higher risk of PE or

SPTB may benefit from frequent monitoring and/or preventative treatments

- when analyzing potential predictors, continued marijuana use at 20 weeks' gestation has been found to be significantly associated with SPTB, and its effect is independent of cigarette smoking and socio-economic status

9.1.1. Predictors for PE and SPTB

The final models (Table 9.1.1) included 25 predictors for PE (12 clinical and 13 SNP predictors), and 34 predictors for SPTB (21 clinical and 13 SNP predictors). All predictors for PE in both tiers can be obtained by 15 weeks' gestation, while the prediction for SPTB in Tier 2 included cervical length measurement obtained at 20 weeks of gestation. Nevertheless, initial screening for both PE and SPTB can be performed at 15 weeks' gestation, in which a low-risk group can be identified.

As discussed in Chapter 7 and 8, most predictors in the models are well recognized factors, which include age, obesity, family history, drug use, and stress (Murphy, 2007; Briceno-Perez et al., 2009; Ibanez et al., 2012). It is interesting to note that factors such as BMI, months to conceive, and fruit intake before pregnancy are predictive for both PE and SPTB.

Known predictors including height and years of schooling reduces the risk of SPTB, while BMI increases the risk of both PE and SPTB. Clinical predictors such as vaginal bleeding and mean arterial pressure at 15 weeks' gestation increases the risk of PE, while LLETZ treatment and previous hospital admission due to hyperemesis increases the risk of SPTB.

Table 9.1.1: Final model predictors for preeclampsia and spontaneous preterm birth

		Preeclampsia				Spontaneous Preterm Birth			
		Tier 1 (15 weeks of gestation)		*Tier 2 (15 weeks of gestation)*		*Tier 1 (15 weeks of gestation)*		*Tier 2 (20 weeks of gestation)*	
Predictors		**Odds (95% CI)**	**P**	**Odds (95% CI)**	**P**	**Odds (95% CI)**	**P**	**Odds (95% CI)**	**P**
Demographic	Age (maternal)	0.962 (0.933 - 0.992)	***0.0141***	-		-		-	
	Height (maternal)	-		-		0.975 (0.946 - 1.005)	0.0994	0.959 (0.927 - 0.992)	***0.0155***
	BMI (maternal)	1.059 (1.031 - 1.087)	***0.0000***	1.058 (1.018 - 1.099)	***0.004***	-		1.014 (0.977 - 1.053)	0.4508
	Years of schooling	-		-		-		0.904 (0.757 - 1.080)	0.2673
Clinical	Gravidity	-		-		1.388 (1.055 - 1.826)	***0.0192***	1.504 (1.109 - 2.041)	***0.0087***
	Miscarriage ≤10wks	0.422 (0.193 - 0.922)	***0.0304***	0.365 (0.109 - 1.224)	0.1027	-		-	
	Months to conceive	0.418 (0.218 - 0.802)	***0.0088***	0.377 (0.150 - 0.951)	***0.0387***	-		1.016 (1.000 - 1.033)	0.0569
	Hospital admission due to Hyperemesis	-		-		2.438 (0.905 - 6.564)	0.0779	-	
	Any LLETZ treatment	-		-		2.533 (1.111 - 5.773)	***0.027***	-	
	Metformin for PCOS (at conception)	-		-		2.732 (0.850 - 8.782)	0.0916	-	
	Vaginal bleeding ≥5 days	2.272 (1.214 - 4.250)	***0.0102***	-		-		-	
	MAP (at 15 wks)	1.062 (1.040 - 1.084)	***0.0000***	1.073 (1.041 - 1.106)	***0.0000***	-		-	
	Cervical length (at 20 wks)	-		-		-		1.040 (1.010 - 1.071)	***0.0133***

Lifestyle	Fruit consumption (<1x/day at 1mth pre-preg)	1.336 (0.859 - 2.080)	0.1989	-		1.911 (1.162 - 3.144)	***0.0108***	-	
	Folate dose >800µg per day (at 1st trim)	-		-		0.339 (0.109 - 1.053)	0.0614	-	
	Alcohol consumption (1st trim)	1.002 (0.991 - 1.014)	0.6766	0.944 (0.886 - 1.007)	0.0803	-		-	
	Cigarettes per day (at 15 wks)	0.951 (0.893 - 1.013)	0.1183	-		-		-	
	Other recreational drug use (at 15 wks)	-		-		4.341 (1.855 - 10.160)	***0.0007***	-	
	Marijuana (>1/day at 1st trim)	-		-		-		8.060 (2.736 - 23.745)	***0.0002***
	Climbing stairs (>10x/day at 15wks)	-		-		2.270 (1.282 - 4.021)	***0.0049***	3.436 (1.819 - 6.489)	***0.0001***
	Living with relatives	-		-		-		3.824 (1.277 - 11.455)	***0.0165***
	State-Trait Anxiety Inventory (>90th centile at 15wks)	-		-		2.197 (1.204 - 4.009)	***0.0104***	3.304 (1.763 - 6.192)	***0.0002***
	Not feeling better	-		-		1.893 (1.069 - 3.352)	***0.0286***	-	
Family History	Participant's birthweight	1.000 (0.999 - 1.000)	***0.0105***	1.000 (0.999 - 1.000)	0.0508	-		-	
	Participant's mother had PE (1x)	-		-		-		2.014 (0.971 - 4.178)	0.0602
	Participant's mother had PE (>=2x)	-		-		-		2.974 (1.058 - 8.356)	***0.0387***
	Family history (Low birthweight baby)	-		-		1.627 (1.031 - 2.568)	***0.0365***	-	
	Family history of SPTB	-		-		-		1.591 (0.911 - 2.779)	0.103
	Family history of PE	2.030 (1.338 - 3.080)	***0.0009***	2.889 (1.565 - 5.332)	***0.0007***	-		-	
	Family history (CH)	1.143	0.4312	-		-		-	

		(0.820 - 1.594)				
Genotype	AGT (maternal)[SNP]	-	-		3.653 (1.134 - 11.766)	***0.03***	3.259 (0.992 - 10.712)	***0.034***
	AGTR1 (maternal)[SNP]	-	0.243 (0.055 - 1.068)	0.0611	-		-	
	ADD1 (maternal)[SNP]	-	-		1.356 (0.902 - 2.038)	0.1433	-	
	BCL2 (maternal)[SNP]	-	-		1.497 (0.991 - 2.261)	0.0555	-	
	MBL2 (maternal)[SNP]	-	-		2.591 (0.846 - 7.933)	0.0955	-	
	TCN2 (maternal)[SNP]	-	-		1.455 (0.973 - 2.176)	0.0681	1.535 (0.978 - 2.407)	***0.0431***
	FLT1 (maternal)[SNP]	-	-		2.533 (0.605 - 10.611)	0.2034	-	
	IGF2R (maternal)[SNP]	-	-		1.505 (1.022 - 2.217)	***0.0382***	-	
	IL1B (maternal)[SNP]	-	-		1.357 (0.913 - 2.019)	0.1314	-	
	uPA (maternal)[SNP]	-	-		2.214 (1.176 - 4.169)	***0.0139***	3.347 (1.727 - 6.487)	***0.0004***
	IGF1R (maternal)[SNP]	-	-		1.403 (0.954 - 2.063)	0.0856	1.667 (1.085 - 2.560)	***0.0147***
	MMP2 (maternal)[SNP]	-	-		1.844 (1.251 - 2.718)	***0.002***	1.974 (1.286 - 3.030)	***0.002***
	MMP9 (maternal)[SNP]	-	-		1.659 (1.040 - 2.645)	***0.0337***	1.655 (0.996 - 2.751)	0.0623
	TIMP3 (maternal)[SNP]	-	-		1.575 (0.908 - 2.732)	0.1063	-	
	IL10 (maternal)[SNP]	-	0.513 (0.312 - 0.846)	***0.0089***	-		-	
	MTHFR (maternal)[SNP]	-	3.424 (1.730 - 6.776)	***0.0004***	-		-	

	PGF (maternal)[SNP]	-	2.151 (1.032 - 4.486)	***0.0411***	-	-
	PLG (maternal)[SNP]	-	1.745 (1.078 - 2.825)	***0.0235***	-	-
	INSR (maternal)[SNP]	-	0.556 (0.236 - 1.307)	0.1782	-	-
	NOS2A (paternal)[SNP]	-	0.578 (0.342 - 0.978)	***0.0411***	-	-
	TP53 (paternal)[SNP]	-	1.625 (0.992 - 2.662)	0.0536	-	-
	MTHFR (paternal)[SNP]	-	1.763 (0.881 - 3.527)	0.1092	-	-
	INS (paternal)[SNP]	-	2.699 (1.157 - 6.298)	***0.0217***	-	-
	TGFB (paternal)[SNP]	-	1.906 (0.706 - 5.142)	0.2029	-	-
	PGF (paternal)[SNP]	-	0.521 (0.297 - 0.914)	***0.0231***	-	-
	MMP2 (paternal)[SNP]	-	2.355 (1.076 - 5.157)	***0.0322***	-	-

Regarding lifestyle factors, a low fruit intake of less than one serve per day in the month pre-pregnancy increases the risk of PE and SPTB, and having a folate supplement of at least 800µg per day during first trimester reduces the risk of SPTB. Interestingly, drug use and psychological factors such as anxiety are predictive for SPTB, but not PE.

When comparing prediction accuracy through sensitivity, specificity, and AUC, it is apparent that SNP predictors have predictive value for both PE and SPTB, with an improved AUC (Fig. 9.1.1). Interestingly, only maternal SNPs were included in the SPTB model, suggesting that genetic factors in the mother but not placenta may have a higher predictive value for SPTB. However, for PE paternal SNPs also contribute to risk prediction suggesting factors in the placenta confer risk.

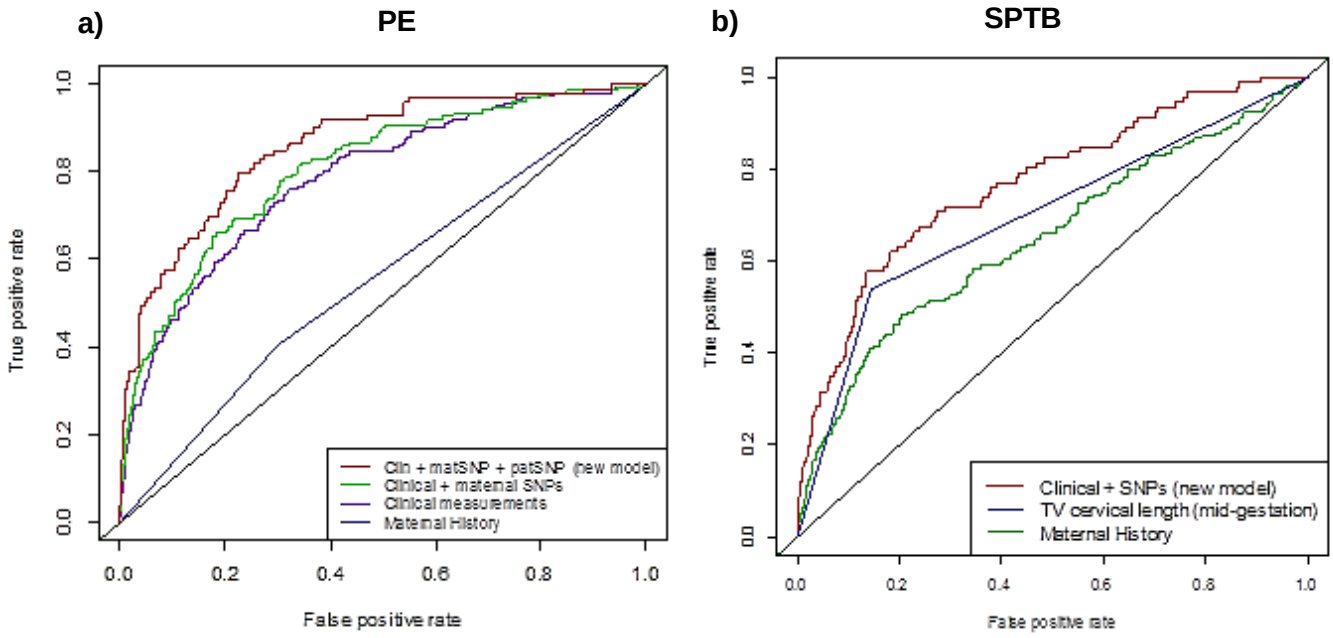

Fig. 9.1.1: ROC curves for a) PE and b) SPTB models
red=clinical and SNP predictors; blue= traditional approach (maternal history for PE, TV cervical length for SPTB)

In addition to identifying potential predictors, this study has also performed a detailed analysis of the effects of marijuana use on pregnancy complications (Chapter 6). In view of the fact that there is a continuing increase in the number of women of

reproductive age being exposed to marijuana, and more recently the legalization of marijuana in the US, this is of great concern to public health.

The results demonstrated clearly that marijuana use increases the risk of SPTB, and its effect is independent of cigarette smoking and socio-economic status. Moreover, there was a higher proportion of early SPTB, delivered before 34 weeks of gestation, amongst women who continued to use marijuana at 20 weeks' gestation, with a significant increase in risk (Odds 5.13; 95% CI 2.48 to 11.89), compared to women who did not use marijuana.

All results are in agreement with the literature that PE and SPTB are complex diseases that do not solely depend on clinical or genetic factors, but involve a mixture of both genetic and environmental effects. By understanding the relationship or associations of risk factors, more accurate prediction may be achieved.

9.1.2. Prediction with Tiered Models

In summary of the tiered prediction models for PE and SPTB (Chapter 7 and 8), this section will focus on the prediction results of both models. It needs to be noted that while the results for both tiered models are summarized together, they are independent predictions, and have not been designed for 'combined' prediction of PE and SPTB. In other words, the prediction outcome for PE has no 'knowledge' of SPTB, and vice versa, even though some predictors overlap. Nevertheless, it is of interest to observe the prediction results where predictions for both PE and SPTB are desired. Hence, this summary provides an insight into the prediction results when both tiered models are applied simultaneously.

A total of 2284 patients were analyzed with both tiered models (Table 9.1.2). The rows indicate the number of patients predicted at low, moderate, or high risk, while the columns show the number of observed cases. Of the 392 patients predicted as low-risk in both PE and SPTB models, there are 5 SPTB cases and 9 PE cases, resulting in 3.45% of PE or SPTB cases missed. It is worth noting that the joint incidence of either a PE or SPTB case is 10.3% in SCOPE.

With 55 PE and 46 SPTB cases amongst the 1421 women identified to be at moderate risk of any PE or SPTB, the percentage of PE or SPTB is 7.11%. Similarly, amongst the 170 predicted as either at high risk of PE or SPTB only, 18 had PE and 17 had SPTB, resulting in a positive predictive value of 20.6%. Interestingly, there is only 1 PE case misclassified as at risk of SPTB.

There are 84 PE and/or SPTB cases amongst the 297 women predicted at high risk of either PE or SPTB, with a moderate risk of SPTB or PE, or at high risk for both PE and SPTB. This results in a positive predictive value of 29.3%.

Table 9.1.1: Predicted risk vs. true cases of preeclampsia and preterm birth

Predicted Risk		Non-case	PE	SPTB	PE& SPTB	Total	Disease %
Low		392	9	5	0	406	3.45%
Moderate	**PE only**	531	23	8	0	562	
	SPTB only	310	1	11	0	322	
	PE& SPTB	479	31	27	0	537	
						1421	7.11%
High	**PE only**	75	18	0	0	93	
	SPTB only	59	1	17	0	77	
						170	21.18%
	PE & moderate SPTB	93	30	2	1	126	
	SPTB & moderate PE	93	3	26	0	122	
	PE& SPTB	17	10	10	2	39	
						287	29.27%
Total		2049	126	106	3	2284	

Amongst patients who are predicted at high risk for any PE or SPTB (Fig. 9.1.2), 39 (8.5%) were predicted at high risk for both PE and SPTB, in which 22 patients (56.4%) had PE or SPTB. From the prediction results, it appears that while there are a few overlaps between the outcomes PE and SPTB, the tiered models are still able to provide reasonable prediction of distinct cases of both PE and SPTB.

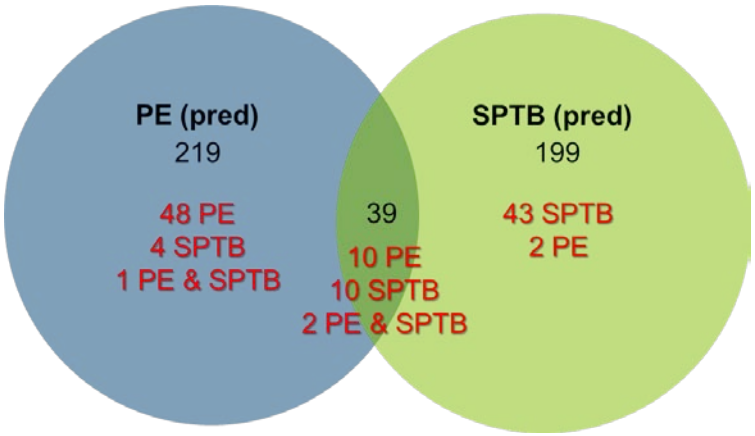

Fig. 9.1.2: Venn diagram of patients predicted as high risk for PE or SPTB
black=number predicted at risk; red=number of observed cases

9.1.3. Model Comparison

A graph comparing the sensitivity, specificity and area under ROC curve of the top 40 models developed in this study and a few traditional models are shown in Fig. 9.1.3 and Fig. 9.1.4 for PE and SPTB, ordered by sensitivity and specificity (r + s) (as discussed in Section 4.5.1). As expected, models that contain both clinical factors and SNPs have the best accuracy, in particular those that include both maternal and paternal predictors.

Interestingly, logistic regression appears to outperform other classification and clustering methods discussed in Chapter 4. This may be due to the fact that a major challenge for applying clustering methods in this study is the lack of distinct features that separate PE or SPTB cases from non-cases. Furthermore, diseases with a low prevalence less than 10% may render grouping or discriminating wide-spread or sparse

cases inefficient. In contrast to other approaches, the “fuzzy” approach in logistic regression that allows for modeling of continuous odds had the advantage of performing sensitivity analysis for various probability thresholds, which provides more flexibility on customized level of sensitivity and specificity that is essential for establishing the tiered model.

Comparing the performance of the models, the tiered models had the best accuracy out of all models developed in this study and outperformed traditional approaches (indicated by red dotted lines labelled Maternal history, Mid-gestation Ultrasound Doppler, and Cervical length in Fig. 9.1.3 and Fig. 9.1.4) based on maternal history of a previous PE or PTB (Papageorghiou et al., 2005; Bittar et al., 2007), or mid-gestation ultrasound Doppler reported in the literature. With a PPV of 22.9%, the tiered model for PE not only outperforms the current approach using ultrasound Doppler at mid-gestation, with a PPV of 22.6% (Albaiges et al., 2000), but also provides an earlier prediction available at 15 weeks gestation. Similarly for SPTB, the tiered model achieved a PPV of 23.6% (note this is in nulliparous women), which outperforms the current approach using cervical length measurements with an estimated PPV of 20.8% in symptomatic women (Lim et al., 2011).

It is worth noting that although some of the prediction models published (discussed in Chapter 2) reported a higher predictive performance, many of the prediction models are for severe cases of disease such as early-onset PE rather than all PE as described in this thesis, or based on symptomatic women, which therefore cannot be performed during early stages of pregnancy. In contrast, the tiered model developed in this study is applicable to all pregnant women, as the prediction is independent of a previous pregnancy or symptoms of PE or PTB.

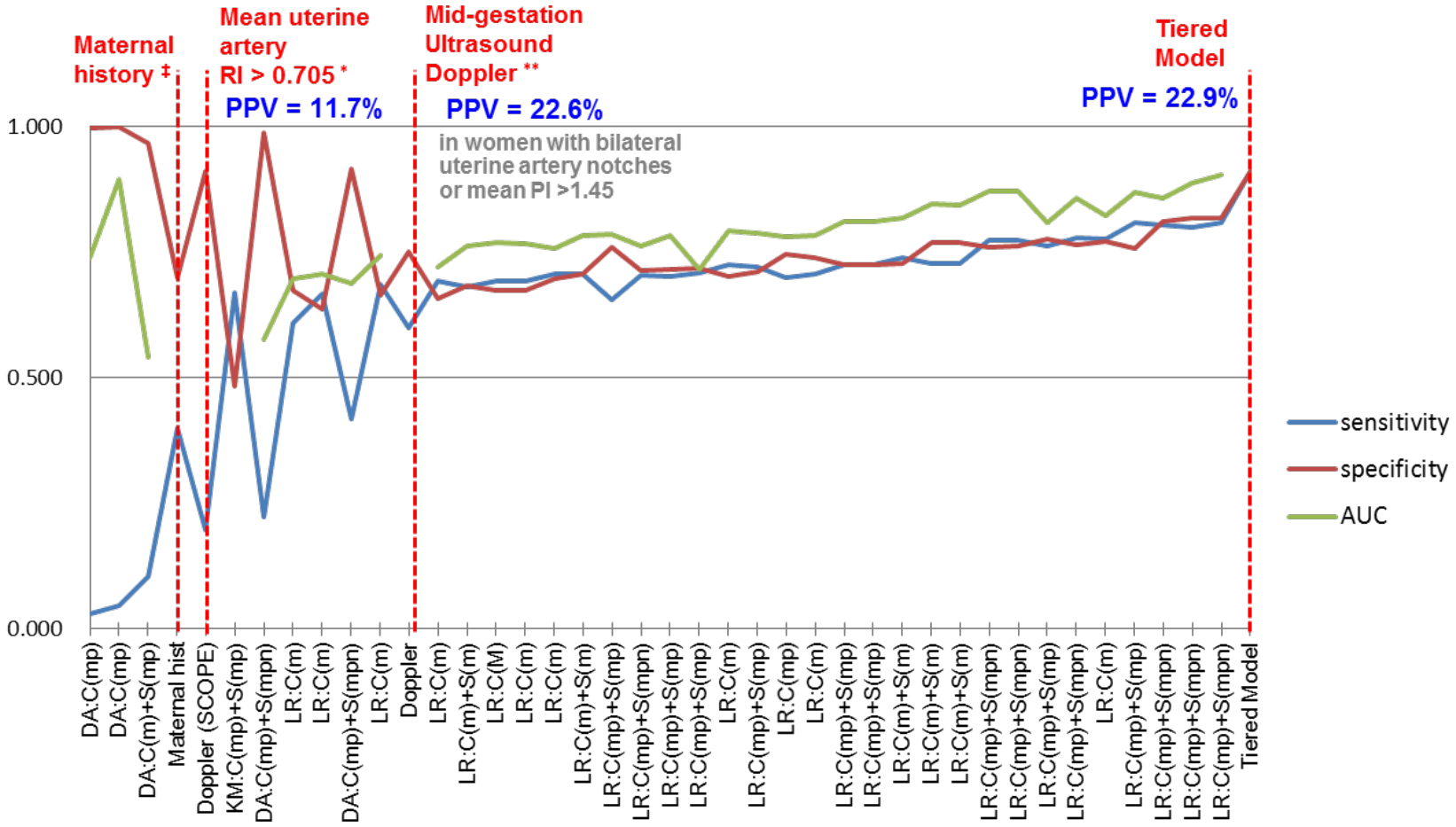

Fig. 9.1.3: Top 40 models developed in this study for preeclampsia (sorted by r+s) compared with current approaches

*‡history of a previous PE (Papageorghiou et al., 2005); *estimate from SCOPE data; **abnormal Doppler (Albaiges et al., 2000)*
LR=Logistic regression; DA=Discriminant analysis; C(..)=clinical predictors; S(..)=SNPs; m= maternal factors; p= paternal factors; n=neonatal factors

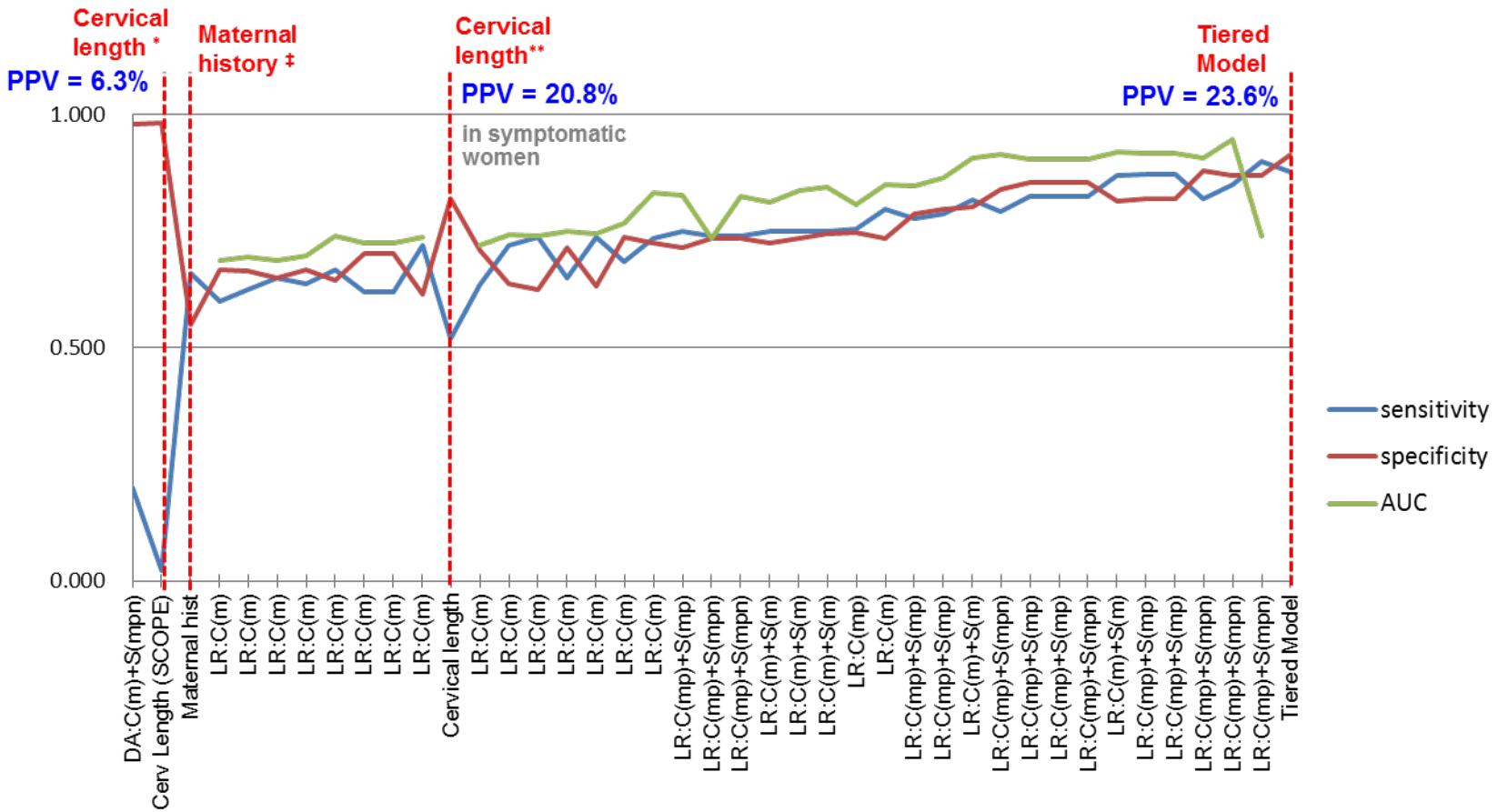

Fig. 9.1.4: Top 40 models developed in this study for preterm birth (sorted by r+s) compared with current approaches

**Cervical length <=25mm (estimate from SCOPE data); ‡history of a previous PTB (Bittar et al., 2007); ** cervical length <=25mm (Lim et al., 2011)*
LR=Logistic regression; DA=Discriminant analysis; C(..)=clinical predictors; S(..)=SNPs; m= maternal factors; p= paternal factors; n=neonatal factors

9.2. Strengths and Limitations

A major strength of this study is the wealth of information recorded and the data quality of the SCOPE database, in which a clearly defined population of nulliparous women was recruited, and rigorous data monitoring was performed to reduce data entry or transcription errors. With details of antenatal visits, as well as lifestyle and psychological status available at pre-pregnancy, 15 weeks and 20 weeks of gestation, models can be developed from predictors obtained at different stages of pregnancy. However, as the lifestyle factors such as fruit intake and drug use are self-reported, there may be a potential recall bias.

Regarding the prediction of PE and SPTB, the ability of the tiered model to classify risk levels is a major strength, where an initial screening is available at 15 weeks of gestation to identify women considered as low-risk, and the chance of unnecessary interventions may be minimized in this group of patients. At the same time, modifiable predictors, including fruit intake and exercise, may be addressed in patients predicted at risk to reduce the risk or severity of disease. Moreover, preventative treatments such as calcium supplementation or aspirin for PE or progesterone for SPTB, may still be administered after Tier 2 screening available by 15 weeks' for PE and 20 weeks of gestation for SPTB.

On the other hand, the uncertainty of the moderate-risk group is a limitation. As discussed in Chapters 7 and 8, with the proportion of disease close to its current prevalence, women classified at moderate risk will rely on frequent monitoring, and further research will benefit the prediction for this group of women.

9.3. Future Work

To further enhance the prediction of PE and SPTB, alternative approaches such as Structural Equation Modelling (SEM), Bayesian Network Analysis (BN), and Hidden Markov Model (HMM) can be considered (Baum, 1972; Castillo et al., 1997; Kline, 2010). These methods provide a probabilistic approach to prediction, where the association of predictors may be modelled through a sequence of ‘states’, with the ability to account for latent or hidden variables.

In addition, graphical modeling may also be applied to enhance the understanding of associations behind predictors of PE and SPTB. This can be visualized through Directed Acyclic Chain Graphs (DACG), where the causal and non-causal relationships between variables (or predictors) are shown through nodes and edges, based on their conditional dependency (Thulasiraman et al., 1992; Edwards, 2000).

It will also be of interest to expand the prediction models to other pregnancy complications, and to develop an ‘integrated’ prediction system.

References

Albaiges, G., Missfelder-Lobos, H., et al. (2000). "One-stage screening for pregnancy complications by color Doppler assessment of the uterine arteries at 23 weeks' gestation." Obstet Gynecol **96**(4): 559-564.

Baum, L. E. (1972). "An Inequality and Associated Maximization Technique in Statistical Estimation of Probabilistic Functions of a Markov Process." Inequalities **3**: 1-8.

Bittar, R. E., da Fonseca, E. B., et al. (2007). "Predicting preterm delivery in asymptomatic patients with prior preterm delivery by measurement of cervical length and phosphorylated insulin-like growth factor-binding protein-1." Ultrasound Obstet Gynecol **29**(5): 562-567.

Blondel, B., Le Coutour, X., et al. (1990). "Prediction of preterm delivery: is it substantially improved by routine vaginal examinations?" Am J Obstet Gynecol **162**(4): 1042-1048.

Briceno-Perez, C., Briceno-Sanabria, L., et al. (2009). "Prediction and prevention of preeclampsia." Hypertension in Pregnancy **28**(2): 138-155.

Castillo, E., Gutiérrez, J. M., et al. (1997). Learning Bayesian Networks. Expert Systems and Probabilistic Network Models. New York, Springer-Verlag**:** 481-528.

Edwards, D. (2000). Introduction to Graphical Modelling. New York, Springer-Verlag.

Fuchs, F., Bouyer, J., et al. (2010). "Ultrasound cervical length measurement for prediction of delivery before 32 weeks in women with emergency cerclage for cervical insufficiency." Int J Gynaecol Obstet **110**(3): 245-248.

Iams, J. D., Goldenberg, R. L., et al. (1998). "The Preterm Prediction Study: recurrence risk of spontaneous preterm birth. National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network." Am J Obstet Gynecol **178**(5): 1035-1040.

Ibanez, G., Charles, M. A., et al. (2012). "Depression and anxiety in women during pregnancy and neonatal outcome: data from the EDEN mother-child cohort." Early

Hum Dev **88**(8): 643-649.

Jacquemyn, Y. and Zemtsova, O. (2010). "Risk factors and prediction of preeclampsia." Acta Clinica Belgica **65**(1): 1-12.

Kline, R. B. (2010). Principles and Practice of Structural Equation Modeling, The Guilford Press.

Lim, K., Butt, K., et al. (2011). "SOGC Clinical Practice Guideline. Ultrasonographic cervical length assessment in predicting preterm birth in singleton pregnancies." J Obstet Gynaecol Can **33**(5): 486-499.

Murphy, D. J. (2007). "Epidemiology and environmental factors in preterm labour." Best Practice & Research Clinical Obstetrics & Gynaecology **21**(5): 773-789.

Papageorghiou, A. T., Yu, C. K., et al. (2005). "Assessment of risk for the development of pre-eclampsia by maternal characteristics and uterine artery Doppler." British Journal of Obstetrics and Gynaecology **112**(6): 703-709.

Thulasiraman, K. and Swamy, M. N. S. (1992). 5.7 Acyclic Directed Graphs. Graphs: Theory and Algorithms, John Wiley and Son**:** 118.

Appendix

I. List of SNPs

No.	Gene	RS no.	No.	Gene	RS no.
1	ACE	rs4343	51	IL4	rs2243250
2	ADD1	rs4961	52	IL6	rs1800795
3	ADRB2	rs1042714	53	INS	rs3842752
4	AGT	rs699	54	INSR	rs1051690
5	AGT	rs4762	55	INSR	rs2059806
6	AGTR1	rs5186	56	IRS2	rs1865434
7	AGTR2	rs1403543	57	KDR	rs2071559
8	AGTR2	rs11091046	58	KDR	rs2305948
9	AGTR2	rs12710567	59	LGALS13	rs3764843
10	AMEL	Amelogenin	60	LIN28	rs12747426
11	ANGPT1	rs2507800	61	MAD1L1	rs1801368
12	ANXA5	rs17551751	62	MBL2	rs1800450
13	BAX	rs4645878	63	MDM2	rs2279744
14	BCL2	rs2279115	64	MMP2	rs243865
15	COL4A2	rs41315048	65	MMP2	rs2285053
16	COX2	rs5275	66	MMP9	rs3918242
17	COX2	rs20417	67	MTHFD1	rs2236225
18	CPB2	rs3742264	68	MTHFR	rs1801131
19	CYP11A1	rs4887139	69	MTHFR	rs1801133
20	CYP11A1	rs8039957	70	MTR	rs1805087
21	CYP24A1	rs2248137	71	MTRR	rs1801394
22	ENG	rs10987759	72	NAT1	rs1057126
23	F2	rs1799963	73	NAT2	rs1208
24	F5	rs6025	74	NOS2A	rs1137933
25	FLT1	FLT1C677T	75	PAI1	rs1799768
26	FTO	rs9939609	76	PAI1	rs1799889
27	GSTP1	rs1695	77	PAI2	rs6098
28	GSTT1	rs2266637	78	PAI2	rs6103
29	H19	rs217727	79	PCSK4	rs791470
30	H19	rs2839701	80	PGF	rs1042886

31	HIF1a	rs10873142	**81**	PLG	rs4252114
32	HIF1a	rs11549465	**82**	PLG	rs2859879
33	IGF1	rs5742620	**83**	PTEN	rs1234220
34	IGF1	rs7965399	**84**	PTEN	rs2673832
35	IGF1	rs12579108	**85**	REN	rs5707
36	IGF1R	rs2229765	**86**	TCN2	rs1801198
37	IGF1R	rs11247361	**87**	TGFB	rs1800469
38	IGF2	rs680	**88**	THBS1	rs2228262
39	IGF2	rs3741204	**89**	TIMP2	rs8179090
40	IGF2AS	rs1003484	**90**	TIMP3	rs5749511
41	IGF2AS	rs1004446	**91**	TP53	rs1042522
42	IGF2R	rs2274849	**92**	uPA	rs2227564
43	IL10	rs1800871	**93**	uPA	UPA4065
44	IL10	rs1800872	**94**	UPAR	rs4251923
45	IL10	rs1800896	**95**	VDR	rs2228570
46	IL1A	rs17561	**96**	VDR	rs7975232
47	IL1A	rs1800587	**97**	VEGF	rs699947
48	IL1B	rs16944	**98**	VEGF	rs3025039
49	IL1B	rs3136558	**99**	VTN	rs704
50	IL1RN	rs454078	**100**	XRCC2	rs3218536

II. Significant subgroup differences between Adelaide and Auckland SCOPE pregnancies

Variable	Adelaide Mean ± SE / %	Auckland Mean ± SE / %	P
Age of participants approached, includes recruited and excluded; these data can be used to compare age of women declined to participate with recruits age	23.76 ± 0.09	30.39 ± 0.08	0.000
Number of years ago that participant migrated to current country	11.90 ± 0.37	8.78 ± 0.33	0.010
Total years of schooling (primary and secondary, not pre-school or tertiary)	11.67 ± 0.02	12.59 ± 0.01	0.000
Maternal Socioeconomic index (SEI) calculated using the New Zealand Socioeconomic Index guide (Galbraith C, Jenkin G, Davis P, Coope P, New Zealand Social Economic Index 1996 Users Guide, Statistics New Zealand, Wellington, New Zealand)	27.76 ± 0.19	47.93 ± 0.26	0.000
Number people sharing current accommodation	2.65 ± 0.02	2.40 ± 0.02	0.000
number of individuals supported by participant and partner's income	1.91 ± 0.01	2.06 ± 0.01	0.000
Participant’s birthweight (g)	3240.22 ± 10.69	3314.02 ± 9.82	0.001
Participant’s partner’s birthweight (g). Partner refers to biological father of current fetus.	3421.54 ± 11.96	3486.32 ± 11.14	0.005
Participant’s gestation at delivery (wks)	39.41 ± 0.04	39.94 ± 0.03	0.000
Number of previous miscarriages at <=10 wks gestation with a different man from one who has fathered the current pregnancy	0.04 ± 0.00	0.02 ± 0.00	0.006
Number of D&C or surgical terminations of pregnancy i.e. Number of cervical dilatations	0.22 ± 0.01	0.14 ± 0.01	0.000
Number of cervical dilatations associated with a termination of pregnancy	0.12 ± 0.01	0.06 ± 0.00	0.000
Partner number' for father of fetus in current pregnancy. Partner refers to men with whom the woman has had a pregnancy, not the number of sexual partners	1.16 ± 0.01	1.13 ± 0.01	0.016
Age at menarche (years)	12.63 ± 0.03	12.80 ± 0.03	0.003
Number of colposcopies	0.04 ± 0.01	0.09 ± 0.01	0.000

Number of LLETZ treatments	0.03 ± 0.00	0.05 ± 0.00	0.007
Number of laser treatments	0.01 ± 0.00	0.04 ± 0.00	0.000
Number of colposcopies where the last colposcopy was 7-12 months before conception	0.00 ± 0.00	0.01 ± 0.00	0.031
Number of colposcopies where the last colposcopy was >12 months before conception	0.02 ± 0.00	0.05 ± 0.00	0.000
Number of laser where the last colposcopy was >12 months before conception	0.01 ± 0.00	0.04 ± 0.00	0.000
Number of LLETZ where the last colposcopy was >12 months before conception	0.02 ± 0.00	0.04 ± 0.00	0.020
Months of sexual relationship prior to conception with the biological father of the baby; conceived on 1st intercourse or with donor sperm=0.03 months	40.21 ± 0.67	64.79 ± 0.80	0.000
Months used barrier contraception (condoms or diaphragm) in relationship with biological father of baby before conception	7.10 ± 0.29	16.72 ± 0.49	0.000
Months of sexual relationship without barrier contraception with biological father of baby	33.11 ± 0.63	48.07 ± 0.75	0.000
Frequency of sexual intercourse with biological father of baby per month in the 3 months prior to conception	17.83 ± 0.30	11.26 ± 0.16	0.000
Total number of exposures to sperm from biological father of baby prior to conception	500.06 ± 13.25	477.74 ± 9.27	0.009
Frequency of sexual intercourse per month in the 1st trimester	7.96 ± 0.18	5.57 ± 0.12	0.000
Duration of hyperemesis (weeks); if no hyperemesis=0	0.45 ± 0.03	0.32 ± 0.03	0.014
Weight loss associated with hyperemesis (kg); if no hyperemesis then 0	2.86 ± 0.25	4.83 ± 0.21	0.000
number of vaginal bleeds commencing <=6 weeks	0.08 ± 0.01	0.13 ± 0.01	0.000
number of vaginal bleeds commencing >12 weeks	0.05 ± 0.00	0.03 ± 0.00	0.043
number of vaginal bleeds lasting 2-4 days before 15w SCOPE visit	0.07 ± 0.01	0.09 ± 0.01	0.026
Total days of vaginal bleeding before 15w SCOPE visit	0.80 ± 0.07	0.99 ± 0.06	0.048
Total days of vaginal bleeding before 15w SCOPE visit, spotting or light	0.75 ± 0.06	0.90 ± 0.06	0.043
Total days of vaginal bleeding at or before 6 weeks gestation	0.31 ± 0.04	0.58 ± 0.05	0.000

Total days of vaginal bleeding after 12 weeks gestation	0.11 ± 0.01	0.07 ± 0.01	0.044
Gestational age when 1st vaginal bleed occurred before 15w SCOPE visit	8.45 ± 0.12	7.30 ± 0.11	0.000
Gestational age when 2nd vaginal bleed occurred before 15w SCOPE visit	10.67 ± 0.22	9.20 ± 0.22	0.002
Gestational age when last vaginal bleed occurred before 15w SCOPE visit	9.32 ± 0.13	8.33 ± 0.12	0.000
Duration 1st vaginal bleed (days) before 15w SCOPE visit	0.68 ± 0.06	0.81 ± 0.05	0.049
Number of episodes of spotting vaginal bleeding at or before 6w gestation	0.06 ± 0.00	0.09 ± 0.01	0.004
Number of episodes of light vaginal bleeding at or before 6w gestation	0.01 ± 0.00	0.03 ± 0.00	0.003
Number of episodes of spotting vaginal bleeding after 12w gestation	0.04 ± 0.00	0.02 ± 0.00	0.012
Total duration of spotting vag bleeding at or before 6w gestation	0.25 ± 0.04	0.40 ± 0.04	0.004
Total duration of light vag bleeding at or before 6w gestation	0.04 ± 0.01	0.15 ± 0.03	0.003
Total duration of spotting or light vag bleeding at or before 6w gestation	0.28 ± 0.04	0.55 ± 0.05	0.000
Total duration of spotting vag bleeding after 12w gestation	0.08 ± 0.01	0.04 ± 0.01	0.012
Total number of vag bleeding at or before 12 weeks gestation	0.21 ± 0.01	0.27 ± 0.01	0.009
Total duration of vag bleeding at or before 12w gestation (days)	0.69 ± 0.06	0.92 ± 0.06	0.005
Folate dose (µg per day) prior to pregnancy	173.76 ± 6.00	502.77 ± 9.66	0.000
Folate dose (µg per day) in 1st trimester	518.24 ± 5.63	823.02 ± 10.68	0.000
Folate dose (µg per day) at 15w SCOPE visit	466.81 ± 6.29	560.13 ± 10.18	0.000
number of cigarettes per day in the 3 months pre-pregnancy	6.27 ± 0.17	1.19 ± 0.07	0.000
number of cigarettes per day in the 1st trimester	4.54 ± 0.14	0.79 ± 0.05	0.000
Number of weeks of cigarette exposure in pregnancy prior to 15w SCOPE visit	4.63 ± 0.12	1.04 ± 0.06	0.000
number of cigarettes per day at 15w SCOPE visit	1.98 ± 0.08	0.16 ± 0.02	0.000

Total number of cigarettes a woman was exposed to in the 1st trimester	394.86 ± 13.23	52.32 ± 4.08	0.000
units of alcohol per week in the 3 months pre-pregnancy (1unit=10ml)	4.52 ± 0.24	4.55 ± 0.17	0.000
units of alcohol per week in the 1st trimester	3.45 ± 0.20	3.16 ± 0.16	0.000
Number of weeks of alcohol exposure in pregnancy prior to 15w SCOPE visit	2.49 ± 0.07	3.02 ± 0.07	0.000
Total units of alcohol a woman was exposed in the 1st trimester	24.05 ± 2.02	18.60 ± 0.93	0.000
Gestation ceased other recreational drugs (binge drinking ie>=6 units/session or illicit drugs)	1.56 ± 0.07	0.59 ± 0.04	0.000
Number of times marijuana was taken in the 1st trimester	26.27 ± 3.55	0.29 ± 0.07	0.000
Gestation marijuana ceased in pregnancy	0.97 ± 0.06	0.16 ± 0.02	0.000
Number of times amphetamines was taken in the 1st trimester	0.18 ± 0.05	0.02 ± 0.01	0.000
Gestation amphetamines ceased in pregnancy	0.14 ± 0.02	0.02 ± 0.01	0.000
Number of times herbal highs was taken in the 1st trimester	0.00 ± 0.00	0.01 ± 0.00	0.002
Gestation herbal highs ceased in pregnancy	0.00 ± 0.00	0.04 ± 0.01	0.001
Number of times substance P/ crystal meth or amphetamines was taken in the 1st trimester	0.18 ± 0.05	0.03 ± 0.01	0.001
Gestation substance P/ crystal meth or amphetamines ceased in pregnancy	0.12 ± 0.02	0.03 ± 0.01	0.002
gestation of 1st scope visit at '15w'	15.52 ± 0.01	15.42 ± 0.01	0.000
1st systolic BP at 15w SCOPE visit measured by mercury or aneroid sphygmomanometer	109.63 ± 0.19	107.50 ± 0.20	0.000
1st MAP (mean arterial pressure) BP at 15w SCOPE visit measured by mercury or aneroid sphygmomanometer	79.61 ± 0.14	79.08 ± 0.15	0.047
2nd systolic BP at 15w SCOPE visit measured by mercury or aneroid sphygmomanometer	109.33 ± 0.18	106.71 ± 0.19	0.000
2nd MAP (mean arterial pressure) at 15w SCOPE visit measured by mercury or aneroid sphygmomanometer	79.41 ± 0.14	78.58 ± 0.14	0.003
mean systolic BP from the 1st and 2nd BP recordings at 15w SCOPE visit	109.48 ± 0.18	107.11 ± 0.19	0.000
3rd diastolic BP measurement at 15w SCOPE visit using a single recording with Microlife 3AC1-2	71.09 ± 0.21	69.10 ± 0.21	0.000

4th diastolic BP measurement at 15w SCOPE visit by MaM measurement (3x) using Microlife 3AC1-2	69.20 ± 0.18	68.18 ± 0.19	0.009
pulse per minute at 15w SCOPE visit	83.67 ± 0.19	73.48 ± 0.17	0.000
Weight at 15w SCOPE visit (kg)	72.38 ± 0.33	68.44 ± 0.22	0.000
BMIat15w SCOPE visit	27.05 ± 0.12	24.78 ± 0.07	0.000
Height at 15w SCOPE visit (cm)	163.36 ± 0.11	166.18 ± 0.11	0.000
Measured Sitting Height at 15w SCOPE visit (cm)	128.72 ± 0.06	132.51 ± 0.09	0.000
Stool Height at 15w SCOPE visit (cm)	45.99 ± 0.01	45.56 ± 0.03	0.000
Calculated leg length at 15w SCOPE visit (cm)	80.65 ± 0.08	79.24 ± 0.09	0.000
Waist at 15w SCOPE visit (cm)	89.38 ± 0.24	84.14 ± 0.18	0.000
Waist Height Ratio at 15w SCOPE visit	0.55 ± 0.00	0.51 ± 0.00	0.000
Hip circumference at 15w SCOPE visit (cm)	106.05 ± 0.24	101.91 ± 0.17	0.000
Head circumference (cm) at 15w SCOPE visit	55.87 ± 0.03	56.03 ± 0.03	0.006
15w SCOPE visit waist hip ratio	0.84 ± 0.00	0.83 ± 0.00	0.000
Random glucose measured by glucometer at 15w SCOPE visit (mmol/L)	5.48 ± 0.01	5.29 ± 0.02	0.000
Hours worked in paid employment per week evaluated at 15w SCOPE visit	22.75 ± 0.30	36.38 ± 0.25	0.000
Hours studying per week evaluated at 15w SCOPE visit	1.71 ± 0.10	1.62 ± 0.10	0.018
Hours exercising/gardening per week evaluated at 15w SCOPE visit	3.19 ± 0.08	3.44 ± 0.05	0.000
How many hours of standing on weekdays on average at 15w SCOPE visit	5.71 ± 0.05	3.80 ± 0.05	0.000
How many hours of standing on weekend day on average at 15w SCOPE visit	5.14 ± 0.05	4.19 ± 0.04	0.000
How many hours of sleeping during day on weekdays on average at 15w SCOPE visit	0.98 ± 0.03	0.36 ± 0.01	0.000
How many hours of sleeping at night on weekdays on average at 15w SCOPE visit	8.35 ± 0.03	8.18 ± 0.02	0.001

Behavioural response to pregnancy: Limiting Behaviour Score evaluated at 15w SCOPE visit	7.22 ± 0.07	8.56 ± 0.06	0.000
Social support (listening ears and practical support scores added) evaluated at 15w SCOPE visit	2.78 ± 0.02	3.06 ± 0.02	0.000
Participant's booking platelets x10^9/L before 20 weeks	256.38 ± 0.93	266.82 ± 1.00	0.000
Participant's ferritin (ug/L) before 20 weeks	76.15 ± 1.63	67.95 ± 1.78	0.002
Participant's risk of trisomy on screening	7654.87 ± 84.23	3890.08 ± 70.02	0.000
Biobank Cholesterol	5.48 ± 0.02	5.41 ± 0.02	0.018
Biobank HDL	1.73 ± 0.01	1.83 ± 0.01	0.000
Biobank LDL	3.05 ± 0.01	2.92 ± 0.01	0.000
Biobank Total cholesterol: HDL cholesterol ratio	3.28 ± 0.01	3.05 ± 0.01	0.000
Participant's gestation (in weeks) of SCOPE 20w visit	20.27 ± 0.01	20.02 ± 0.01	0.000
number of cigarettes per day in the week prior to 20w SCOPE visit	1.86 ± 0.08	0.17 ± 0.02	0.000
Units of Alcohol per week at 20w SCOPE visit	0.06 ± 0.02	0.12 ± 0.01	0.000
1st systolic BP at 20w SCOPE visit measured by mercury or aneroid sphygmomanometer	112.57 ± 0.19	108.51 ± 0.19	0.000
1st MAP (mean arterial pressure) BP at 20w SCOPE visit measured by mercury or aneroid sphygmomanometer	81.27 ± 0.14	79.58 ± 0.14	0.000
2nd systolic BP at 20w SCOPE visit measured by mercury or aneroid sphygmomanometer	112.10 ± 0.18	107.91 ± 0.19	0.000
2nd MAP (mean arterial pressure) at 20w SCOPE visit measured by mercury or aneroid sphygmomanometer	80.98 ± 0.14	79.12 ± 0.14	0.000
3rd systolic BP measurement at 20w SCOPE visit using a single recording with Microlife 3AC1-2	117.94 ± 0.29	114.90 ± 0.29	0.000
3rd diastolic BP measurement at 20w SCOPE visit using a single recording with Microlife 3AC1-2	73.28 ± 0.22	68.90 ± 0.20	0.000
4th systolic BP measurement at 20w SCOPE visit by MaM measurement (3x) using Microlife 3AC1-2	114.83 ± 0.23	112.29 ± 0.25	0.000
4th diastolic BP measurement at 20w SCOPE visit by MaM measurement (3x) using Microlife 3AC1-2	71.17 ± 0.20	68.25 ± 0.19	0.000
mean systolic BP from the 1st and 2nd BP recordings at 20w SCOPE visit	112.32 ± 0.18	108.21 ± 0.19	0.000

pulse per minute at 20w SCOPE visit	87.81 ± 0.19	76.12 ± 0.18	0.000
Weight at 20w SCOPE visit (kg)	74.83 ± 0.34	70.71 ± 0.22	0.000
Weight change between 15w and 20w SCOPE visit (kg)	2.61 ± 0.03	2.50 ± 0.03	0.000
Random glucose measured by glucometer at 20w SCOPE visit (mmol/L)	5.70 ± 0.02	5.41 ± 0.02	0.000
ph high vaginal swab measured using Hydrion paper (pH range 4.0-5.5)	4.66 ± 0.01	4.52 ± 0.01	0.000
weight of vaginal swab secretions collected by WeckCel swab (g)	0.06 ± 0.00	0.10 ± 0.00	0.000
Hours worked in paid employment per week evaluated at 20w SCOPE visit	21.31 ± 0.31	35.80 ± 0.25	0.000
Hours studying per week evaluated at 20w SCOPE visit	1.26 ± 0.08	1.44 ± 0.09	0.000
Hours relaxing per week evaluated at 20w SCOPE visit	16.48 ± 0.23	14.44 ± 0.16	0.027
Hours exercising/gardening per week evaluated at 20w SCOPE visit	3.41 ± 0.08	3.59 ± 0.05	0.000
Score for 'Need for recovery' scale measured at 20w SCOPE visit	4.05 ± 0.05	3.53 ± 0.05	0.000
How many hours of standing on weekdays on average at 20w SCOPE visit	5.63 ± 0.06	3.90 ± 0.05	0.000
How many hours of standing on weekend day on average at 20w SCOPE visit	5.07 ± 0.05	4.27 ± 0.04	0.000
How many hours of sleeping during day on weekdays on average at 20w SCOPE visit	0.91 ± 0.03	0.30 ± 0.01	0.000
How many hours of sleeping during day on weekend day on average at 20w SCOPE visit	1.29 ± 0.03	1.02 ± 0.02	0.014
How many hours of sleeping at night on weekdays on average at 20w SCOPE visit	8.19 ± 0.03	8.06 ± 0.02	0.004
Behavioural response to pregnancy: Limiting Behaviour Score evaluated at 20w SCOPE visit	6.28 ± 0.06	6.94 ± 0.06	0.000
Score for 'Short form State-Trait Anxiety Inventory evaluated at 20w SCOPE visit	32.95 ± 0.21	31.68 ± 0.19	0.017
Social support (listening ears and practical support scores added) evaluated at 20w SCOPE visit	2.69 ± 0.02	3.06 ± 0.02	0.000
Gestational age (weeks) of anatomy/growth scan 19-21w	20.28 ± 0.01	19.90 ± 0.01	0.000
Biparietal diameter on 19-21w scan	47.67 ± 0.05	47.27 ± 0.05	0.000

Biparietal diameter on 19-21w scan transformed to multiple of median (MoM) for gestational age	0.99 ± 0.00	1.01 ± 0.00	0.000
Head circumference on 19-21w scan	177.38 ± 0.17	173.71 ± 0.18	0.000
Abdominal circumference on 19-21w scan	157.12 ± 0.18	153.62 ± 0.20	0.000
Femur length on 19-20w scan	32.85 ± 0.05	32.38 ± 0.05	0.000
Femur length on 19-20w scan transformed to MoM for gestational age	1.00 ± 0.00	1.01 ± 0.00	0.000
Gestation age (weeks) of 19-21w Doppler	20.29 ± 0.01	19.92 ± 0.01	0.000
Umbilical artery Resistance Index (RI) measured using Doppler ultrasound at 19-21w	0.75 ± 0.00	0.72 ± 0.00	0.000
Umbilical artery Doppler RI at 19-21w, transformed to MoM by gestation	1.03 ± 0.00	0.99 ± 0.00	0.000
Right uterine RI at 19-21w	0.60 ± 0.00	0.54 ± 0.00	0.000
Left uterine RI at 19-21w	0.60 ± 0.00	0.55 ± 0.00	0.000
Mean uterine artery RI at 19-21w, note when only 1 uterine RI available this was used as the mean RI	0.60 ± 0.00	0.54 ± 0.00	0.000
Mean uterine artery RI at 19-21w transformed into MoM	1.06 ± 0.00	0.96 ± 0.00	0.000
Shortest transvaginal cervical lengh (mm)	40.38 ± 0.16	41.21 ± 0.13	0.000
Random glucose (mmol/L) result	7.03 ± 0.64	4.43 ± 0.32	0.050
Gestation age (weeks) of oral glucose tolerance test (OGTT)	28.49 ± 0.15	30.56 ± 0.16	0.000
OGTT fasting glucose result (mmol/L)	4.41 ± 0.03	4.25 ± 0.02	0.000
GTT 1 hour glucose result (mmol/L)	7.82 ± 0.09	8.32 ± 0.08	0.004
Gestation age (weeks) GDM diagnosed	28.19 ± 0.33	30.86 ± 0.40	0.000
Last recorded weight prior to delivery	83.14 ± 0.37	80.27 ± 0.27	0.010
Gestational age of Last recorded weight prior to delivery	36.66 ± 0.08	38.49 ± 0.08	0.000
Maximum systolic BP in the last 2 weeks of pregnancy prior to the onset of labour	129.07 ± 0.28	121.40 ± 0.26	0.000

Maximum diastolic BP in the last 2 weeks of pregnancy prior to the onset of labour	77.45 ± 0.21	75.24 ± 0.20	0.000
Systolic BP which is linked with the maximum diastolic BP prior to onset of labour	125.94 ± 0.28	119.98 ± 0.24	0.000
Highest pre-labour 24h protein excretion at the end of pregnancy	0.88 ± 0.19	1.03 ± 0.14	0.003
Last Haemoglobin (g/L) pre-IV fluids at the end of pregnancy	118.56 ± 0.19	123.06 ± 0.18	0.000
Last Haematocrit (g/L) pre-IV fluids at the end of pregnancy	0.36 ± 0.00	0.37 ± 0.00	0.000
Last Platelets pre-IV fluids at the end of pregnancy	239.36 ± 1.08	230.39 ± 1.03	0.000
Total duration of 2nd stage of labour (minutes)	62.21 ± 1.11	84.89 ± 1.36	0.000
Total Duration of Stage 1 plus stage 2 of labour (hours)	8.14 ± 0.08	8.85 ± 0.09	0.000
Maximum temperature in labour (oC)	36.72 ± 0.01	36.79 ± 0.01	0.001
Placental weight (grams); they were not routinely weighed at most centres	573.22 ± 2.84	640.98 ± 2.81	0.000
Apgar score at 1 minute	8.05 ± 0.03	8.49 ± 0.02	0.000
Apgar score at 5 minutes	8.96 ± 0.02	9.64 ± 0.01	0.000
Baby birthweight (grams)	3335.01 ± 11.45	3418.04 ± 10.35	0.000
Customised birthweight centile adjusted for mothers height, weight at 15w visit, ethnicity, sex and weight of baby and gestation at delivery of baby; all mothers were 0 parity	46.38 ± 0.52	49.04 ± 0.51	0.010
Baby Head circumference (cm)	34.32 ± 0.04	34.85 ± 0.04	0.000
Baby Length (cm)	49.03 ± 0.06	50.77 ± 0.06	0.000
Baby Mid arm circumference (cm)	10.60 ± 0.02	10.84 ± 0.02	0.000
Number of Days baby spent in hospital following birth	4.97 ± 0.34	4.13 ± 0.09	0.000
Fasting Glucose (mmol/L) if GDM diagnosed in pregnancy and GTT done postpartum	5.11 ± 0.27	4.44 ± 0.06	0.012
Gestation age (weeks) of regular contractions (defined as contractions >=1 every 10 minutes when in preterm labour leading to birth) in women who had spontaneous PTB	32.81 ± 0.37	34.43 ± 0.25	0.015
Highest systolic BP on admission to hospital	132.91 ± 0.67	127.01 ± 0.70	0.000

Highest systolic BP antepartum	140.30 ± 0.82	131.60 ± 0.80	0.000
Highest systolic BP postpartum	138.94 ± 0.75	131.86 ± 0.88	0.000
systolic BP which is linked with the maximum diastolic BP on admission to hospital	132.57 ± 0.66	126.48 ± 0.68	0.000
Maximum diastolic BP antepartum	86.25 ± 0.60	83.72 ± 0.59	0.015
Systolic BP which is linked with the maximum diastolic BP on admission to hospital	137.65 ± 0.78	129.88 ± 0.73	0.000
Maximum diastolic BP postpartum	83.98 ± 0.51	81.50 ± 0.59	0.005
systolic BP which is linked with the maximum diastolic BP antepartum	135.88 ± 0.76	129.41 ± 0.82	0.000
Max dBP antepartum or postpartum	77.74 ± 0.22	75.35 ± 0.20	0.000
Highest systolic BP either antepartum or postpartum	129.61 ± 0.30	121.64 ± 0.27	0.000
Highest pulse on admission to hospital	85.32 ± 0.46	80.78 ± 0.40	0.000
Haemoglobin (g/L) Lowest antepartum	118.29 ± 0.61	124.35 ± 0.57	0.000
Haematocrit (PCV) Lowest antepartum	0.35 ± 0.00	0.37 ± 0.00	0.000
WCC (x109/L) Highest antepartum	14.24 ± 0.19	12.89 ± 0.19	0.000
WCC (x109/L) highest value postpartum	15.00 ± 0.29	13.86 ± 0.30	0.031
Platelets Highest antepartum value	254.32 ± 3.05	237.88 ± 2.90	0.004
Urine protein creatinine ratio (mg/mmol) Lowest antepartum	64.72 ± 13.00	123.04 ± 20.60	0.000
Urine protein creatinine ratio (mg/mmol) Highest antepartum	135.84 ± 21.07	139.84 ± 21.49	0.004
Maximum spot urine protein creatinine ratio (mg/mmol) at end of pregnancy	58.61 ± 8.00	79.30 ± 9.48	0.000
24h urinary protein (g/24h) Lowest antepartum	1.09 ± 0.26	1.18 ± 0.17	0.004
24h urinary protein (g/24h) Highest antepartum	1.21 ± 0.26	1.23 ± 0.17	0.033
Maximum 24h urinary protein (g/24h) at end of pregnancy	0.88 ± 0.18	1.01 ± 0.13	0.007

Creatinine (mmol/L) Lowest antepartum	51.59 ± 0.82	65.28 ± 0.81	0.000
Creatinine (mmol/L) Highest antepartum	59.26 ± 0.92	70.92 ± 0.94	0.000
Creatinine (mmol/L) highest value postpartum	61.96 ± 1.30	74.47 ± 1.44	0.000
ALT (IU/L) Lowest antepartum	14.42 ± 0.90	22.79 ± 3.27	0.001
GGT (IU/L) highest value postpartum	19.94 ± 2.51	43.05 ± 5.02	0.003
Albumin (g/L) Lowest antepartum	28.99 ± 0.31	33.09 ± 0.28	0.000
Albumin (g/L) Highest antepartum	31.00 ± 0.33	34.38 ± 0.26	0.000
Albumin (g/L) lowest value postpartum	26.09 ± 0.59	30.52 ± 0.75	0.005
Haptoglobin (mg/dL) Lowest antepartum	0.97 ± 0.09	0.40 ± 0.08	0.049
APTT (sec) Highest antepartum	30.16 ± 0.45	27.03 ± 0.30	0.000
PR Highest antepartum	1.43 ± 0.30	0.91 ± 0.01	0.001
D-dimer (ug/L) Highest antepartum	764.13 ± 151.87	1329.00 ± 13.90	0.022
Number of day unit visits	1.01 ± 0.07	0.53 ± 0.08	0.000
Number of antenatal hospitalization days	2.14 ± 0.19	1.51 ± 0.14	0.000
Number of days in high dependency unit	0.40 ± 0.05	0.06 ± 0.01	0.000
Participant's study number	2920.19 ± 22.02	1818.06 ± 20.65	0.000
Smoked during the 1st trimester	39.0%	12.0%	0.000
Any use of amphetamines during pregnancy	1.8%	0.4%	0.000
Consumed/inhaled/injected other recreational drugs or binge alcohol consumption (>=6 units/session) in the 1st trimester; screening variable which includes binge drinking (>=6 units/session) plus illicit drugs; details of individual drugs recorded separately	19.8%	10.7%	0.000
Using marijuana at 15w SCOPE visit	3.6%	0.5%	0.000
Any use of marijuana during pregnancy	10.3%	2.2%	0.000

Used any marijuana in the 1st trimester	10.3%	2.2%	0.000
Consumed/inhaled/injected other recreational drugs or binge alcohol consumption (>=6 units/session) in the 1st trimester combining 'unknown' (n=0) with NO; screening variable which includes binge drinking (>=6 units/session) plus illicit drugs; details of individual drugs recorded separately	19.8%	10.7%	0.000
Any alcohol consumption in 1st trimester	38.3%	53.0%	0.000
Consumed/inhaled/injected other recreational drugs or binge alcohol consumption (>=6 units/session) at 15w SCOPE visit; screening variable which includes binge drinking (>=6 units/session) plus illicit drugs; details of individual drugs recorded separately	4.3%	0.6%	0.000
Consumed/inhaled/injected other recreational drugs or binge alcohol consumption (>=6 units/session) at gestation of 1st SCOPE visit with unknown (n=0) combined with NO; screening variable which includes binge drinking (>=6 units/session) plus illicit drugs; details of individual drugs recorded separately	4.3%	0.6%	0.000
Any herbal highs in the 1st trimester	0.0%	0.9%	0.003
Pulse >=84/min at 15w SCOPE visit	52.8%	15.1%	0.000
Homeopathic used as alternative therapy at 15w SCOPE visit	0.0%	2.6%	0.000
Smoked at 15w SCOPE visit	23.9%	3.9%	0.000
Any use of herbal highs during pregnancy	0.0%	0.9%	0.003
waist at 15w SCOPE visit >= 94 cm	34.2%	16.0%	0.000
waist at 15w SCOPE visit > 90th centile according to ethnicity	17.7%	5.6%	0.000
Any amphetamines in the 1st trimester	1.8%	0.4%	0.000
High binge alcohol consumption in 1st trimester (defined as >1 binge per week)	1.7%	0.1%	0.000
high (>=3 times per day) green leafy vegetables consumption in the month prior to conception	2.6%	13.5%	0.000
Pulse <=60/min at 15w SCOPE visit	0.9%	11.3%	0.000
Any proteinuria at 15w SCOPE visit	0.0%	3.1%	0.000
Short Height (<161 cm)	33.0%	19.2%	0.000
Nutritional supplements used as alternative therapy at 15w SCOPE visit	1.9%	9.2%	0.000
Acupuncture used as alternative therapy at 15w SCOPE visit	0.4%	3.6%	0.000

Low green (<3x times/mth) leafy vegetables consumption in pregnancy prior to 15w SCOPE visit	9.7%	4.6%	0.000
Any folate intake in 1st trimester	88.9%	96.4%	0.000
Alternative therapies used at 15w SCOPE visit	11.9%	29.2%	0.000
Naturopathic supplements used as alternative therapy at 15w SCOPE visit	0.9%	2.9%	0.000
Massage used as alternative therapy at 15w SCOPE visit	4.7%	7.9%	0.001
Yoga used as alternative therapy at 15w SCOPE visit	0.2%	3.8%	0.000
High(>=3 times per day) green leafy vegetables consumption in pregnancy prior to 15w SCOPE visit	3.2%	12.5%	0.000
Any folate intake prior to pregnancy	31.0%	66.8%	0.000
Any folate intake at 15w SCOPE visit	78.4%	73.0%	0.001
Low(<3x times/mth) green leafy vegetables consumption in the month prior to conception	11.5%	2.7%	0.000
Any hard drug (cocaine, substance P, amphetamines or opiates) use in pregnancy.	2.1%	0.8%	0.003
Any use of substance P/ crystal meth or amphetamines during pregnancy	1.9%	0.6%	0.001
Any substance P/ crystal meth or amphetamines in the 1st trimester	1.9%	0.6%	0.001
Score for 'Short form State-Trait Anxiety Inventory evaluated at 15w SCOPE visit >90th centile	9.0%	6.6%	0.017
Behavioural response to pregnancy: Limiting Behaviour Score evaluated at 15w SCOPE visit >90th Centile	6.1%	8.1%	0.043
Never used a computer in the last month evaluated at 15w SCOPE visit	25.8%	3.6%	0.000
Edinburgh Postnatal Depression Score evaluated at 15w SCOPE visit >90th C	11.4%	8.7%	0.015
Watching >=5h TV per day evaluated at 15w SCOPE visit	22.1%	8.1%	0.000
Any computer usage in last month evaluated at 15w SCOPE visit	74.2%	96.4%	0.000
Snored most nights (binary) evaluated at 15w SCOPE visit	18.4%	10.2%	0.000
Never undertakes exercise in pregnancy evaluated at 15w SCOPE visit	15.5%	5.3%	0.000
Any hospital admissions between the 15w and 20w SCOPE visit	3.7%	1.1%	0.000
Any Trichomonas Vaginalis infection on vaginal swab at <20 weeks	99.3%	88.8%	0.000

Any Gonorrhea infection on vaginal swab at <20 weeks	99.3%	88.8%	0.000
Alternative therapies used at 20w SCOPE visit	6.8%	23.2%	0.000
Any Staph aureus OR E coli OR Ureaplasma infection on vaginal swab at <20 weeks	99.3%	88.8%	0.000
Severity of vaginal infection on vaginal swab at <20 weeks (compressed normal flora and Blank/no details)	0.3%	3.1%	0.000
Naturopathic supplements used as alternative therapy at 20w SCOPE visit	0.2%	1.5%	0.001
Nutritional supplements used as alternative therapy at 20w SCOPE visit	1.1%	7.2%	0.000
Herbal treatment used as alternative therapy at 20w SCOPE visit	0.1%	0.9%	0.008
Yoga used as alternative therapy at 20w SCOPE visit	0.1%	4.1%	0.000
Acupuncture used as alternative therapy at 20w SCOPE visit	0.0%	1.6%	0.000
Massage used as alternative therapy at 20w SCOPE visit	2.7%	6.1%	0.000
Smoking at 20w SCOPE visit (week prior to interview)	22.3%	3.8%	0.000
Homeopathic used as alternative therapy at 20w SCOPE visit	0.0%	1.4%	0.000
Any alcohol consumption at 20w SCOPE visit (week prior to interview)	3.6%	9.9%	0.000
Consumed/inhaled/injected other recreational drugs or binge alcohol consumption (>=6 units/session) between 15w and 20w SCOPE visit; screening variable for illicit drugs or binge alcohol; details of individual drugs recorded separately	3.8%	0.5%	0.000
high vaginal swab taken at 20w SCOPE visit	98.4%	66.1%	0.000
fibronectin swab taken at 20w SCOPE visit	1.7%	0.2%	0.000
heparin taken at 20w SCOPE visit	99.9%	99.0%	0.005
Any marijuana taken between 15w and 20w SCOPE visit	3.3%	0.5%	0.000
Consumed/inhaled/injected other recreational drugs or binge alcohol consumption (>=6 units/session) between 15w and 20w SCOPE visit with UNKNOWN (n=0) and NO combined; screening variable for illicit drugs or binge alcohol; details of individual drugs recorded separately	3.8%	0.5%	0.000
citrate taken at 20w SCOPE visit	98.8%	57.5%	0.000
'Need for recovery' scale: Worn out at end of day evaluated at 20w SCOPE visit	66.4%	51.7%	0.000

Need for recovery' scale: Feels exhausted at end of working day evaluated at 20w SCOPE visit	54.2%	43.1%	0.000
Participant with Mixed Ethnicity (main and other ethnicity is not the same)	30.9%	13.2%	0.000
Participant's main Ethnicity is Indian	0.3%	3.8%	0.000
Participant's main Ethnicity is Asian	3.6%	5.3%	0.040
Participant's main Ethnicity is Pacific Island	0.1%	2.1%	0.000
Participant's main Ethnicity is Caucasian	91.7%	84.0%	0.000
Participant's Other Ethnicity is Caucasian	28.8%	8.8%	0.000
Participant's main Ethnicity is Other / African	3.9%	1.6%	0.000
Participant's main Ethnicity is Other	3.5%	1.4%	0.000
Participant's main Ethnicity is Maori	0.4%	3.3%	0.000
Need for recovery' scale: Feeling fresh after dinner evaluated at 20w SCOPE visit	51.8%	37.1%	0.000
Snored most nights (binary) evaluated at 20w SCOPE visit	21.0%	12.4%	0.000
Need for recovery' scale: During last part of day, suboptimal performance at job due to fatigue evaluated at 20w SCOPE visit	33.4%	38.3%	0.020
Need for recovery' scale: Able to relax only on second day off evaluated at 20w SCOPE visit	30.5%	19.8%	0.000
Need for recovery' scale: Takes over 1hr to recover after work evaluated at 20w SCOPE visit	47.8%	42.2%	0.011
Never undertakes exercise in pregnancy evaluated at 20w SCOPE visit	13.4%	3.8%	0.000
Watching >=5h TV per day in the last month, evaluated at 20w SCOPE visit	20.9%	5.5%	0.000
Behavioural response to pregnancy: Limiting Behaviour Score >90th Centile, evaluated at 20w SCOPE visit	6.7%	8.8%	0.050
Any computer usage in last month evaluated at 20w SCOPE visit	73.3%	96.4%	0.000
Score for 'Short form State-Trait Anxiety Inventory >90th centile, evaluated at 20w SCOPE visit	12.3%	8.3%	0.000
Edinburgh Postnatal Depression Score >90th C, evaluated at 20w SCOPE visit	12.5%	8.3%	0.000
Abdominal circumference on 19-21w scan <142 (<10thC)	5.4%	13.3%	0.000

Head circumference on 19-21w scan transformed to Z score for gestational age <10th Centile	14.4%	11.3%	0.015
Biparietal diameter on 19-21w scan transformed to multiple of median (MoM) for gestational age <10th Centile	12.1%	9.5%	0.022
Head circumference on 19-21w scan <165 mm (<10th C)	8.3%	17.3%	0.000
Biparietal diameter on 19-21w scan transformed to Z score for gestational age <10th centile Z score	17.2%	8.6%	0.000
Femur length on 19-20w scan transformed to Z score for gestational age <10th C	11.2%	7.1%	0.000
Femur length on 19-20w scan transformed to Z score for gestational age <10th C	13.1%	6.7%	0.000
Femur length on 19-20w scan <30 mm (<10th C)	8.5%	12.3%	0.001
Any proteinuria at the end of pregnancy (pre-labour) measured by either dipstick or PCR or 24h urine based on data in form 24 only (not case forms)	24.7%	4.8%	0.000
Any other infections in pregnancy between 20w SCOPE and delivery	24.4%	6.6%	0.000
Gastroenteritis in pregnancy between 20w SCOPE and delivery	9.4%	2.4%	0.000
Pyelonephritis infection in pregnancy between 20w SCOPE and delivery	0.7%	0.1%	0.028
Proven Vaginal Candida infection in pregnancy between 20w SCOPE and delivery	26.3%	6.8%	0.000
Onset of delivery was induction and mode of delivery either pre-labour LSCS or LSCS in labour	10.4%	7.4%	0.005
Mean uterine artery RI at 19-21w>90th centile	11.3%	6.0%	0.000
Mean uterine artery RI at 19-21w >=75th centile	65.6%	82.0%	0.000
Right notch at 19-21w, compressed	12.4%	19.7%	0.000
Any vaginal bleeding in pregnancy since 2nd visit	9.2%	4.7%	0.000
Flu/Respiratory tract infection between 20w SCOPE and delivery	25.0%	8.0%	0.000
Umbilical artery Doppler RI at 19-21w, transformed to MoM by gestation >90th Centile	18.6%	6.8%	0.000
Umbilical artery Doppler RI at 19-21w >90th centile	18.2%	6.8%	0.000
Bilateral notch at 19-21w	6.9%	13.3%	0.000
Left notch at 19-21w, compressed	12.8%	23.4%	0.000

Unilateral notch at 19-21w	11.3%	17.3%	0.000
Umbilical artery end diastolic flow velocity at 19-21w compressed categories	1.6%	0.4%	0.001
Umbilical artery end diastolic flow velocity at 19-21w, as collected	1.7%	0.4%	0.001
Mean uterine artery RI at 19-21w MoM>90th centile	11.3%	6.2%	0.000
Gestation screened for gestational diabetes	98.0%	91.2%	0.000
Duration of ruptured membranes before the onset of labour is >6 hours	14.4%	17.8%	0.019
Who performed the Baby Measurements	23.7%	28.8%	0.002
Cord Blood-EDTA Taken	82.0%	61.2%	0.000
Baby Length measured in neonatometer	66.8%	46.0%	0.000
Apgar score < 7 at 5 minutes	2.2%	0.8%	0.002
Baby admitted to Neonatal unit	21.1%	6.7%	0.000
Baby Buccal Swab Taken	1.7%	32.9%	0.000
Baby Oragene Saliva Taken	0.0%	1.6%	0.000
Exclusively breastfeeding on discharge	64.2%	76.2%	0.000
Spontaneous preterm birth according to SCOPE definition	5.9%	4.3%	0.046
Preeclampsia according to SCOPE definition	8.0%	4.2%	0.000
Hypertensive SGA defined by SGA <10th customised birthweight centile and mother has either chronic hypertension, gestational hypertension or preeclampsia	4.0%	2.1%	0.003
Significant proteinuria at the end of pregnancy measured by either dipstick (>=2+) or PCR (>=30 mg/mmol) or 24h urine (>=0.3g/24h)	8.8%	5.2%	0.000
Term Preeclampsia, defined as Preeclampsia and delivered >=37weeks	6.3%	2.9%	0.000
Spontaneous preterm birth defined as spontaneous preterm labour or preterm premature rupture of membranes (PPROM) resulting in preterm birth at <37 weeks.	6.0%	4.3%	0.046
Preeclampsia defined as gestational hypertension with either proteinuria or multi-system disease.	8.0%	4.2%	0.000

Severe maternal preeclampsia defined as sustained severe hypertension (2 or more BP recordings with systolic BP >=170 mm Hg or diastolic BP >=110mm Hg) or multisystem disease	4.3%	2.4%	0.004
Uncomplicated pregnancy within a clinical framework.	52.5%	68.0%	0.000
Uncomplicated pregnancy for laboratory studies, with rigorously defined normal pregnancy	45.5%	64.0%	0.000
Placental abruption defined as retroplacental clot at delivery at delivery or seen on ultrasound scan pre-delivery or diagnosed by clinical criteria including one or more of the following: vaginal bleeding with uterine tenderness and/or evidence of fetal compromise.	1.5%	0.5%	0.007
Baby SGA by population centile using Beeby's scale adjusted for sex and gestation.	12.8%	8.7%	0.000
Gestational hypertension defined as systolic BP>=140 mmHg or diastolic BP>=90 mmHg on at least 2 occasions after the 20 week visit before onset of labour	10.2%	5.6%	0.000
Other pregnancy complication; this would include other complications not listed above such as antepartum haemorrhage, asthma exacerbation, pyelonephritis.	20.0%	11.8%	0.000
Participant's Other Ethnicity is Maori	0.1%	1.9%	0.000
Previous ectopic pregnancy with same man who has fathered the current pregnancy	0.3%	1.1%	0.042
Participant’s gestation at delivery <34wk	2.6%	1.2%	0.004
Birthweight confirmed by checking health records	18.0%	13.2%	0.000
Paid employment working part-time at 15w SCOPE visit	27.7%	10.0%	0.000
Unemployed or Sickness Beneficiary at 15w SCOPE visit	22.3%	2.4%	0.000
Student at 15w SCOPE visit	4.5%	2.5%	0.003
Paid employment working full-time at 15w SCOPE visit	41.6%	81.2%	0.000
Works in a paid job at 15w SCOPE visit	69.3%	91.2%	0.000
Socioeconomic Index (SEI) <24	48.2%	7.8%	0.000
Participant’s gestation at delivery confirmed by checking health or other written record	67.0%	33.1%	0.000
Participant's ethnicity Pacific Islander (recorded as Pacific Islander under either Main Ethnicity or Other Ethnicity)	0.1%	3.1%	0.000
Participant's Other Ethnicity is Pacific Island	0.0%	1.2%	0.000

Participant's ethnicity Caucasian (recorded as Caucasian under either Main Ethnicity or Other Ethnicity)	93.3%	86.2%	0.000
Participant's ethnicity Indian (recorded as Indian under either Main Ethnicity or Other Ethnicity)	0.5%	3.9%	0.000
Participant's ethnicity Maori (recorded as Maori under either Main Ethnicity or Other Ethnicity)	0.5%	5.2%	0.000
Participant's ethnicity Other (recorded as 'other' under either Main Ethnicity or Other Ethnicity)	4.9%	2.1%	0.000
Participant is an immigrant	10.6%	31.5%	0.000
Less than 12 years of schooling	73.8%	34.9%	0.000
Participant immigrated <= 1 year ago	1.9%	4.7%	0.000
Participant immigrated <= 2 years ago	2.7%	8.5%	0.000
Currently attends university	1.8%	4.6%	0.000
Participant's partner is an immigrant. Partner refers to biological father of fetus.	10.4%	30.9%	0.000
Participant immigrated <= 5 years ago	4.2%	16.9%	0.000
Any tertiary education at a university or other post school institution	64.0%	87.1%	0.000
Children in household	16.6%	6.2%	0.000
Participant’s birthweight<2500g	8.0%	6.0%	0.044
Birthweight Confirmed	89.9%	59.0%	0.000
Participant born preterm (<37 weeks)	7.5%	4.4%	0.000
Either IVF or ICSI to conceive current pregnancy	2.9%	5.3%	0.002
Fertility treatment to conceive current pregnancy	4.7%	9.0%	0.000
Clomiphene to assist conception of current pregnancy	1.9%	4.0%	0.002
ICSI to conceive current pregnancy	1.1%	2.7%	0.005
IVF to conceive current pregnancy	2.9%	5.3%	0.002
Any previous pregnancy loss with a different man from one who has fathered the current pregnancy	14.0%	11.3%	0.030
Young age (<=10 years) at menarche	5.7%	3.2%	0.001

Had either LLETZ, laser or cryotherapy treatment for CIN/abnormal smear	3.7%	8.5%	0.000
Last colposcopy >12 months before conception current pregnancy	2.0%	4.6%	0.000
Had laser treatment for CIN/abnormal smear	0.9%	3.7%	0.000
Donor sperm or donor egg used in this pregnancy	0.2%	0.8%	0.035
Any previous termination or miscarriage <=10wks gestation or an ectopic pregnancy with a different man from one who has fathered the current pregnancy	11.4%	9.0%	0.032
History of abnormal smear that led to colposcopy with or without additional treatment such as LLETZ, laser, diathermy; may or may not have had cervical intraepithelial neoplasia (CIN); binary variable	6.0%	13.8%	0.000
Had LLETZ treatment for CIN	2.7%	4.6%	0.009
Underwent colposcopy for abnormal smear or CIN	2.5%	5.9%	0.000
Any previous miscarriage at <=10 wks gestation with a different man from one who has fathered the current pregnancy	3.7%	2.1%	0.008
Duration of sex without contraception with father of baby before current pregnancy= 1 day	4.3%	11.7%	0.000
Any cervical dilatation	18.5%	11.9%	0.000
Participant's mother had any history of pregnancy induced hypertension	11.8%	6.2%	0.000
Wheezing in the past year	17.4%	10.7%	0.000
Currently using B2-agonist inhaler (only includes short acting B2 agonists like salbutamol)	14.6%	9.0%	0.000
Participant's mother had any history of pregnancy induced hypertension (GH or preeclampsia, but unsure which condition), gestational hypertension or preeclampsia	22.3%	13.6%	0.000
Participant's mother had any miscarriages	34.1%	27.6%	0.000
Any family history of miscarriage(s)	39.3%	34.5%	0.007
On metformin for PCOS prior to/at conception	0.3%	1.9%	0.000
Self reported polycystic ovarian syndrome compressed categories of 'unsure and no'=NO and Yes by scan or blood tests=YES	4.0%	8.3%	0.000
Months of sexual relationship prior to conception with the biological father of the baby <=6 months	13.9%	5.4%	0.000
Continued on metformin in 1st trimester after missed period/confirmed pregnancy	0.1%	0.8%	0.018

Received fertility treatment for PCOS prior to/at conception	1.0%	2.3%	0.014
Months of sexual relationship prior to conception with the biological father of the baby <=3 months	5.2%	2.1%	0.000
Use of barrier contraception (condoms or diaphragm) with biological father of baby	59.3%	71.3%	0.000
Had last laserRx>12 months before conception current pregnancy	0.8%	3.6%	0.000
Had last LLETZ Rx>12 months before conception current pregnancy	2.3%	3.8%	0.027
Months of sexual relationship without barrier contraception with biological father of baby <=6 months	24.3%	17.7%	0.000
Diagnosed and treated for anemia prior to pregnancy (self reported)	9.3%	15.4%	0.000
Number of episodes of sexual intercourse per month in 1st trimester <=3	34.4%	46.9%	0.000
Self reported previous >1 urinary tract infection (confirmed by MSU)	28.4%	31.8%	0.046
Self reported hypertension (on more than 1 occasion) while on oral contraception	0.7%	2.4%	0.001
Diagnosed asthma	33.1%	20.1%	0.000
Family history of PET, i.e. participant's mother or sister had had PET	14.0%	7.9%	0.000
Any family history of PIH	13.1%	7.5%	0.000
participant's father has chronic hypertension	18.1%	23.9%	0.000
Participant's mother had any history of LBW baby	17.0%	13.3%	0.005
Family history of GH, participant's mother or sister has had GH	11.9%	5.7%	0.000
Participant's mother had any GDM	5.2%	1.5%	0.000
Participant's mother had any spontaneous PTB	11.8%	9.2%	0.021
Strong family history (2 or more members of immediate family -mother or sisters) of PIH	0.9%	0.2%	0.014
Family history of cerebrovascular accident (participant's mother, father, sibling)	3.3%	6.4%	0.000
Family history of chronic hypertension (participant's mother, father, sibling)	34.9%	42.4%	0.000
Family history of ischaemic heart disease (participant's mother, father, sibling)	11.0%	18.2%	0.000
Participant's mother had any PTB (spontaneous or iatrogenic)	16.9%	11.1%	0.000

participant's father has had a VTE	1.9%	3.3%	0.027
Family history of PIH (i.e. GH or preeclampsia, but unsure which condition), gestational hypertension or preeclampsia; family members are participant's mother and/or sisters	26.3%	16.6%	0.000
participant's father has IHD	8.1%	13.8%	0.000
Participant's mother had any history of PET	11.8%	6.3%	0.000
Any vaginal bleeding at or before 6w gestation	7.6%	12.0%	0.000
participant's father has had a CVA	1.6%	4.5%	0.000
Participant's father deceased due to either CH, IHD, VTE, CVA, diabetes; if died from other conditions=No; unknown fused with NO	1.7%	3.7%	0.002
Participant's mother or father has one or more of type 2 diabetes, chronic hypertension, CVA and IHD	43.6%	51.6%	0.000
Participant's father has one or more of type 2 diabetes, chronic hypertension, CVA and IHD	25.8%	34.6%	0.000
Family history of LBW baby, i.e. participant's mother or sister had had LBW baby	19.7%	15.4%	0.002
Any sister had a history of GH	2.9%	1.6%	0.020
Family history (mother or sisters) of GH or PET	21.1%	12.3%	0.000
FH spontaneous PTB (i.e. participant's mother or sister(s) had a spontaneous PTB)	14.4%	11.8%	0.034
FH GDM (i.e. participant's mother or sister(s) had GDM)	6.7%	2.4%	0.000
Number of episodes of light vaginal bleeding at or before 6w gestation (categorized)	1.5%	3.1%	0.005
Hyperemesis: repeated vomiting in pregnancy not due to other causes (e.g., gastroenteritis) requiring either inpatient admission, IV fluids, nasogastric feeding or vomiting associated with loss >5% of booking weight	6.4%	4.2%	0.010
FH all PTB(spontaneous or iatrogenic) i.e. participant's mother or sister(s) delivered >=2 babies preterm (spont or iatrogenic)	20.0%	14.5%	0.000
participant's mother has IHD	3.1%	5.1%	0.009
Participant's mother had any history of GH	9.4%	4.3%	0.000
Demi-vegetarian (compressed into 2 groups with not vegetarian and demi-vegetarian compressed into one group and the remaining types of vegetarians into 2nd group)	0.9%	2.0%	0.031

Any hospital admissions due to medical reasons other than asthma before 15w SCOPE visit	1.4%	0.3%	0.002
Low (<3x times/mth) fruit consumption in pregnancy prior to 15w SCOPE visit	10.1%	1.3%	0.000
Did NOT consume other fish or seafood in pregnancy prior to 15w SCOPE visit	42.4%	26.5%	0.000
High consumption oily fish (which is high in omega 3 long chain fatty acids) >=3 times a week in pregnancy prior to 15w SCOPE visit	3.4%	5.9%	0.002
high (>=3 times per day) fruit consumption in pregnancy prior to 15w SCOPE visit	15.7%	54.5%	0.000
Consumption other fish or seafood >=3 times a week in the month prior to conception	1.5%	2.8%	0.025
high (>=3 times per day) fruit consumption in the month prior to conception	6.6%	31.2%	0.000
Low (<3x times/mth) fruit consumption in the month prior to conception	22.2%	3.5%	0.000
NOT consumed other fish or seafood in the month prior to conception	38.5%	16.4%	0.000
Any hospital admissions before the 15w SCOPE visit	8.6%	3.1%	0.000
Any light vag bleeding at or before 6w gestation	1.5%	3.1%	0.005
Any spotting vag bleeding after 12w gestation	3.5%	2.1%	0.018
Any hospital admissions due to vaginal bleeding before 15w SCOPE visit	3.1%	0.7%	0.000
Gastroenteritis in pregnancy before 15w SCOPE visit	7.6%	4.7%	0.001
Any spotting vag bleeding at or before 6w gestation	5.7%	8.4%	0.005
Flu/respiratory tract infection in pregnancy before 15w SCOPE visit	28.8%	21.3%	0.000
Any hospital admissions due to hyperemesis before 15w SCOPE visit	3.4%	1.1%	0.000
Any infection (urti or uti or pyelnephritis or gastro or vag candida or other infections) in pregnancy before 15w SCOPE visit	44.1%	37.4%	0.000
Rarely consumed oily fish (which is high in omega 3 long chain fatty acids) <=3 times per month prior to conception	70.4%	58.1%	0.000
Did NOT consume oily fish (which is high in omega 3 long chain fatty acids) in pregnancy prior to 15w SCOPE visit	48.5%	41.0%	0.000
Did NOT consume oily fish (which is high in omega 3 long chain fatty acids) in the month prior to conception	38.9%	24.2%	0.000

High consumption oily fish (which is high in omega 3 long chain fatty acids) >=3 times a week in the month prior to conception	6.3%	9.7%	0.001

III. Exploratory Analysis of Preeclampsia

Variable	Uncomplicated pregnancy N	Preeclampsia N	Uncomplicated pregnancy Mean ± SE	Preeclampsia Mean ± SE	P
Age of participants	1984	178	28.2 ± 5.70	26.7 ± 5.72	0.0006
Total years of schooling (primary and secondary, not pre-school or tertiary)	1984	178	12.3 ± 1.05	12.1 ± 1.08	0.0139
Maternal Socioeconomic index (SEI)	1984	178	41.7 ± 16.41	36.7 ± 15.56	0.0001
Number people sharing current accommodation	1984	178	2.5 ± 1.12	2.6 ± 1.36	0.0313
Participant's birthweight (g)	1872	167	3332.8 ± 529.93	3174.2 ± 550.38	0.0002
Participant's gestation at delivery (wks)	1922	174	39.8 ± 1.79	39.5 ± 2.11	0.0452
Frequency of sexual intercourse with biological father of baby per month in the 3 months prior to conception	1979	178	13.2 ± 11.65	16.7 ± 17.66	0.0003
Duration of hyperemesis (weeks); if no hyperemesis=0	1984	178	0.3 ± 1.64	0.6 ± 2.26	0.0290
number of vaginal bleeds lasting >=10 days before 15w SCOPE visit	1984	178	0.0 ± 0.13	0.0 ± 0.19	0.0495
Number of episodes of vag spotting that last >=5 days before 15w SCOPE visit	1984	178	0.0 ± 0.16	0.1 ± 0.24	0.0064
Number of hospital admissions due to hyperemesis before 15w SCOPE visit	1984	178	0.0 ± 0.22	0.1 ± 0.49	0.0323
Folate dose (µg per day) in 1st trimester	1984	178	730.1 ± 550.73	621.8 ± 261.88	0.0042
Folate dose (µg per day) at 15w SCOPE visit	1984	178	532.4 ± 512.75	455.6 ± 320.60	0.0431
Number of weeks of alcohol exposure in pregnancy prior to 15w SCOPE visit	1967	176	2.9 ± 3.71	2.1 ± 3.47	0.0093
Gestation amphetamines ceased in pregnancy	1984	178	0.1 ± 0.68	0.2 ± 1.27	0.0464
Gestation substance P/ crystal meth or amphetamines ceased in	1984	178	0.1 ± 0.62	0.2 ± 1.27	0.0317

pregnancy					
1st systolic BP at 15w SCOPE visit measured by mercury or aneroid sphygmomanometer	1984	178	106.9 ± 10.46	114.0 ± 11.26	0.0000
1st diastolic BP at 15w SCOPE visit measured by mercury or aneroid sphygmomanometer	1984	178	63.9 ± 7.92	69.3 ± 8.40	0.0000
1st MAP (mean arterial pressure) BP at 15w SCOPE visit measured by mercury or aneroid sphygmomanometer	1984	178	78.2 ± 7.80	84.2 ± 8.46	0.0000
2nd systolic BP at 15w SCOPE visit measured by mercury or aneroid sphygmomanometer	1984	178	106.2 ± 9.96	113.3 ± 10.48	0.0000
2nd diastolic BP at 15w SCOPE visit measured by mercury or aneroid sphygmomanometer	1984	178	63.5 ± 7.71	69.0 ± 8.61	0.0000
2nd MAP (mean arterial pressure) at 15w SCOPE visit measured by mercury or aneroid sphygmomanometer	1984	178	77.8 ± 7.53	83.8 ± 8.34	0.0000
mean systolic BP from the 1st and 2nd BP recordings at 15w SCOPE visit	1984	178	106.6 ± 9.85	113.6 ± 10.53	0.0000
mean diastolic BP from the 1st and 2nd BP recordings at 15w SCOPE visit	1984	178	63.7 ± 7.51	69.1 ± 8.19	0.0000
pulse per minute at 15w SCOPE visit	1982	178	75.9 ± 10.75	80.9 ± 10.92	0.0000
Weight at 15w SCOPE visit (kg)	1984	178	68.6 ± 13.55	76.2 ± 19.25	0.0000
BMI at 15w SCOPE visit	1984	178	25.0 ± 4.66	28.3 ± 6.83	0.0000
Height at 15w SCOPE visit (cm)	1984	178	165.5 ± 6.55	163.9 ± 6.30	0.0020
Measured Sitting Height at 15w SCOPE visit (cm)	1979	178	131.4 ± 4.98	130.6 ± 4.91	0.0485
Stool Height at 15w SCOPE visit (cm)	1981	178	45.7 ± 1.58	45.9 ± 1.38	0.0218
Waist at 15w SCOPE visit (cm)	1980	177	84.9 ± 10.89	91.6 ± 14.81	0.0000
Waist Height Ratio at 15w SCOPE visit	1980	177	0.5 ± 0.07	0.6 ± 0.09	0.0000
Hip circumference at 15w SCOPE visit (cm)	1980	176	102.5 ± 10.41	108.3 ± 13.65	0.0000
Mid upper arm circumference (cm) at 15w SCOPE visit	1972	176	28.7 ± 3.67	30.6 ± 4.55	0.0000
15w SCOPE visit waist hip ratio	1980	176	0.8 ± 0.07	0.8 ± 0.07	0.0113

Variable	Category	Uncomplicated pregnancy N	Preeclampsia N	Uncomplicated pregnancy %	Preeclampsia %	P
Ethnicity of participants	Caucasian	1740	152	87.7%	85.4%	Reference
	Maori	50	5	2.5%	2.8%	0.7767
	Pacific Islander	18	4	0.9%	2.3%	0.0950
	South East Asian	97	7	4.9%	3.9%	0.6333
	Indian subcontinent	38	5	1.9%	2.8%	0.3966
	African ancestry	3	1	0.2%	0.6%	0.2474
	Middle-eastern	12	0	0.6%	0.0%	0.9751
	Hispanic	5	0	0.3%	0.0%	0.9839
	Aboriginal	7	3	0.4%	1.7%	0.0222
	Other	14	1	0.7%	0.6%	0.8463
Participant's main Ethnicity is Pacific Island	No (Not Pacific Islander)	1964	172	99.0%	96.6%	Reference
	Yes (Pacific Islander)	20	6	1.0%	3.4%	0.0091
Years of schooling in categories	<12	321	46	16.2%	25.8%	Reference
	12,13	1608	130	81.1%	73.0%	0.0017
	>13	55	2	2.8%	1.1%	0.0626
University education status	No	1010	111	50.9%	62.4%	Reference
	Dropped out	81	12	4.1%	6.7%	0.3583
	Still attending	68	6	3.4%	3.4%	0.6157

	Graduated	825	49	41.6%	27.5%	0.0005
Tertiary education (i.e. post school education) other than university	No	1048	84	52.8%	47.2%	Reference
	Dropped out	94	10	4.7%	5.6%	0.4205
	Still attending	95	17	4.8%	9.6%	0.0051
	Graduated	747	67	37.7%	37.6%	0.5099
Current work situation at 15w SCOPE visit	Full time work	1376	114	69.4%	64.0%	Reference
	Part time work	315	26	15.9%	14.6%	0.9868
	Student	60	9	3.0%	5.1%	0.1091
	Homemaker	70	6	3.5%	3.4%	0.9379
	Unemployed	145	21	7.3%	11.8%	0.0273
	Sickness beneficiary	10	2	0.5%	1.1%	0.2590
	Other (e.g.) voluntary work	8	0	0.4%	0.0%	0.9797
Unemployed or Sickness Beneficiary at 15w SCOPE visit	No	1829	155	92.2%	87.1%	Reference
	Yes	155	23	7.8%	12.9%	0.0189
Work status at 15w SCOPE visit (3 groups)	No paid work	293	38	14.8%	21.4%	Reference
	Part time work	315	26	15.9%	14.6%	0.0907
	Full time work	1376	114	69.4%	64.0%	0.0237
Works in a paid job at 15w SCOPE visit	No paid work	293	38	14.8%	21.4%	Reference
	Paid work	1691	140	85.2%	78.7%	0.0204
Codes for maternal occupation using the New Zealand Socioeconomic Index guide	Elementary occupations	46	8	2.3%	4.5%	Reference
	Plant/machine operators	28	5	1.4%	2.8%	0.9659

(Galbraith C, Jenkin G, Davis P, Coope P, New Zealand Social Economic Index 1996 Users Guide, Statistics New Zealand, Wellington, New Zealand)	Trade workers	13	2	0.7%	1.1%	0.8854
	Agriculture/fishery workers	13	1	0.7%	0.6%	0.4609
	Service/sales workers	434	49	21.9%	27.5%	0.2939
	Clerks	273	36	13.8%	20.2%	0.5121
	Armed forces	4	0	0.2%	0.0%	0.9768
	Associate professional/technical	333	23	16.8%	12.9%	0.0357
	Professionals	503	31	25.4%	17.4%	0.0147
	Legislators, administrators, managers	337	23	17.0%	12.9%	0.0333
Participant and partner's income grouped into categories	<$25K	152	23	9.3%	14.4%	Reference
	$25K-$74K	555	61	34.1%	38.1%	0.2210
	$75-$124K	741	63	45.5%	39.4%	0.0262
	>$125K	182	13	11.2%	8.1%	0.0392
Participant born preterm (<37 weeks)	No	1876	161	95.7%	92.0%	Reference
	Yes	85	14	4.3%	8.0%	0.0297
Participant's small for gestational age (SGA by population centiles) status at birth	Woman not SGA at birth	1602	132	80.8%	74.2%	Reference
	Unknown	129	11	6.5%	6.2%	0.9165
	Woman was SGA at birth	253	35	12.8%	19.7%	0.0102
Previous Etopic Pregnancy	No	1965	173	99.0%	97.2%	Reference
	Yes	19	5	1.0%	2.8%	0.0314
Any previous miscarriage at >10 weeks	No	1943	170	97.9%	95.5%	Reference
	Yes	41	8	2.1%	4.5%	0.0421

Previous ectopic pregnancy with same man who has fathered the current pregnancy	No	1970	174	99.3%	97.8%	Reference
	Yes	14	4	0.7%	2.3%	0.0403
A single previous miscarriage at <=10 wks gestation with same man who has fathered the current pregnancy	No	1814	172	91.4%	96.6%	Reference
	Yes - 1x	170	6	8.6%	3.4%	0.0195
Any previous miscarriage at <=10 wks gestation with same man who has fathered the current pregnancy	No	1796	170	90.5%	95.5%	Reference
	Yes	188	8	9.5%	4.5%	0.0306
Months of sexual relationship prior to conception with the biological father of the baby in categories	<=3 month	53	11	2.7%	6.2%	Reference
	>3 and <=12month	267	30	13.5%	16.9%	0.1093
	>12 and <=36month	501	43	25.3%	24.2%	0.0163
	>36month	1160	94	58.6%	52.8%	0.0069
Months of sexual relationship prior to conception with the biological father of the baby <=3 months	> 3 months or unknown	1928	167	97.3%	93.8%	Reference
	<=3 months	53	11	2.7%	6.2%	0.0104
Months of sexual relationship prior to conception with the biological father of the baby <=6 months	> 6 months or unknown	1838	152	92.8%	85.4%	Reference
	<=6 months	143	26	7.2%	14.6%	0.0006
Months of sexual relationship without barrier contraception with biological father of baby in categories	<=3 month	227	32	11.5%	18.0%	Reference
	>3 and <=6 month	143	19	7.2%	10.7%	0.8479
	>6 and <=12month	206	10	10.4%	5.6%	0.0045
	>12 and <=36month	511	46	25.8%	25.8%	0.0656
	>36month	894	71	45.1%	39.9%	0.0109
Months of sexual relationship without barrier contraception with biological father of baby <=3 months	No >3m	1754	146	88.5%	82.0%	Reference
	Yes =<3m	227	32	11.5%	18.0%	0.0111

Months of sexual relationship without barrier contraception with biological father of baby <=6 months	No >6m	1611	127	81.3%	71.4%	Reference
	Yes =<6m	370	51	18.7%	28.7%	0.0015
Self reported hypertension (on more than 1 occasion) while on oral contraception	No	1963	173	98.9%	97.2%	Reference
	Yes	21	5	1.1%	2.8%	0.0486
Doctor diagnosed asthma	NO	1535	124	77.4%	69.7%	Reference
	YES	449	54	22.6%	30.3%	0.0204
Severity of asthma in 3 grades	No Asthma	1535	124	77.4%	69.7%	Reference
	Mild asthma	364	48	18.4%	27.0%	0.0064
	Moderate asthma	79	5	4.0%	2.8%	0.6040
	Severe asthma	6	1	0.3%	0.6%	0.5041
Number of sisters who had GH	None of the Sister	1950	170	98.3%	95.5%	Reference
	1 Sister	33	7	1.7%	3.9%	0.0359
	>= 2 Sisters	1	1	0.1%	0.6%	0.0850
Any sister had a history of GH	No	1950	170	98.3%	95.5%	Reference
	Yes	34	8	1.7%	4.5%	0.0133
Family history of GH, participant's mother or sister has had GH	No	1843	157	92.9%	88.2%	Reference
	Yes	141	21	7.1%	11.8%	0.0244
Strong family history (2 or more members of immediate family -mother or sisters) of GH	No	1980	175	99.8%	98.3%	Reference
	Yes	4	3	0.2%	1.7%	0.0054
Strong family history (2 or more members of immediate family -mother or sisters) of recurrent GH	No	1984	178	100.0%	100.0%	Reference
	Yes	0	0	0.0%	0.0%	0.0054

Participant's biological mother had preeclampsia	No	1754	142	88.4%	79.8%	Reference
	participant's mother had PET 1x	116	18	5.9%	10.1%	0.0152
	participant's mother had PET >=2x	29	9	1.5%	5.1%	0.0006
Participant's mother had a history of PE	No	1754	142	88.4%	79.8%	Reference
	participant's mother had PET 1x	116	18	5.9%	10.1%	0.0152
	participant's mother had PET >=2x	29	9	1.5%	5.1%	0.0006
Participant's mother had any history of PE	No, participant's mother never had PET	1839	151	92.7%	84.8%	Reference
	Yes, participant's mother had PET	145	27	7.3%	15.2%	0.0003
Participant's mother had any history of recurrent PE	No, participant's mother never had recurrent PET	1955	169	98.5%	94.9%	Reference
	Yes, participant's mother had recurrent PET	29	9	1.5%	5.1%	0.0010
Number of sisters who had PE	No sister had PET	1938	168	97.7%	94.4%	Reference
	1 Sister had PET	43	10	2.2%	5.6%	0.0061
	>= 2 Sisters had PET	3	0	0.2%	0.0%	0.9810
Any sister has a history of PE	No	1938	168	97.7%	94.4%	Reference
	Yes	46	10	2.3%	5.6%	0.0102
Family history of recurrent PET, i.e. participant's mother or sister had had recurrent PE	No	1946	169	98.1%	94.9%	Reference
	Yes	38	9	1.9%	5.1%	0.0082
Family history (mother or sisters) of recurrent GH or recurrent PE	No Strong Family History of rec_GH or PET or Ecl	1929	166	97.2%	93.3%	Reference

	Yes, Strong Family' History of rec_GH or PET or Ecl	55	12	2.8%	6.7%	0.0047
Participant's biological mother had LBW baby	No	1680	135	84.7%	75.8%	Reference
	participant's mother had LBW baby 1x	189	27	9.5%	15.2%	0.0103
	participant's mother had LBW baby >=2x	58	8	2.9%	4.5%	0.1634
Participant's mother had a history of LBW baby	No	1680	135	84.7%	75.8%	Reference
	participant's mother had LBW baby 1x	189	27	9.5%	15.2%	0.0103
	participant's mother had LBW baby >=2x	58	8	2.9%	4.5%	0.1634
Participant's mother had any history of LBW baby	No, participant's mother never had LBW baby	1737	143	87.6%	80.3%	Reference
	Yes, participant's mother had LBW baby	247	35	12.5%	19.7%	0.0068
Family history of LBW baby, i.e. participant's mother or sister had had LBW baby	No	1695	135	85.4%	75.8%	Reference
	Yes	289	43	14.6%	24.2%	0.0008
Number of sisters with a history of PTB (all-spontaneous or iatrogenic)	No sister had a PTB (spont or iatrogenic)	1915	165	96.5%	92.7%	Reference
	1 Sister had PTB (spont or iatrogenic)	66	13	3.3%	7.3%	0.0085
	>= 2 Sisters had PTB (spont or iatrogenic)	3	0	0.2%	0.0%	0.9810
	Yes	69	13	3.5%	7.3%	0.0124
FH all PTB (spontaneous or iatrogenic) i.e. participant's mother or sister(s) delivered >=2 babies preterm (spont or iatrogenic)	No	1692	141	85.3%	79.2%	Reference
	Yes	292	37	14.7%	20.8%	0.0319
participant's father has chronic hypertension	No	1446	118	72.9%	66.3%	Reference

	participant's father has chronic hypertension	401	48	20.2%	27.0%	0.0336
participant's father has ischaemic heart disease (IHD)	No	1651	133	83.2%	74.7%	Reference
	participant's father has IHD	208	34	10.5%	19.1%	0.0006
participant's father has IHD	No	1651	133	83.2%	74.7%	Reference
	participant's father has IHD	208	34	10.5%	19.1%	0.0006
Participant's father has one or more of type 2 diabetes, chronic hypertension, CVA and IHD	No	1392	110	70.2%	61.8%	Reference
	Yes	592	68	29.8%	38.2%	0.0209
Participant's mother or father has one or more of type 2 diabetes, chronic hypertension, CVA and IHD	No	1060	78	53.4%	43.8%	Reference
	Yes	924	100	46.6%	56.2%	0.0144
participant's sibling(s) has type 2 diabetes	No	1882	168	94.9%	94.4%	Reference
	participant's sibling(s) has type 2 diabetes	7	3	0.4%	1.7%	0.0239
participant's sibling(s) has type 2	No	1882	168	94.9%	94.4%	Reference
	participant's sibling(s) has type 2 diabetes	7	3	0.4%	1.7%	0.0239
Family history of chronic hypertension (participant's mother, father, sibling)	No FH CH	1235	96	62.3%	53.9%	Reference
	yes, FH CH	749	82	37.8%	46.1%	0.0295
Family history of ischaemic heart disease (participant's mother, father, sibling)	No FH IHD	1706	138	86.0%	77.5%	Reference
	yes, FH IHD	278	40	14.0%	22.5%	0.0025
Hyperemesis: repeated vomiting in pregnancy not due to other causes (e.g., gastroenteritis) requiring either inpatient admission, IV fluids, nasogastric feeding or vomiting associated with loss >5% of booking weight	NO	1895	163	95.5%	91.6%	Reference
	YES	89	15	4.5%	8.4%	0.0207

Vomiting continuing at 15 week interview	No Hyperemesis at or before 15w vst	1895	163	95.5%	91.6%	Reference
	Hyperemesis before 15w vst, but ceased before 15w vst	42	6	2.1%	3.4%	0.2532
	Hyperemesis at or before 15w vst and continuing at 15w vst	47	9	2.4%	5.1%	0.0319
number of vaginal bleeds commencing <=6w, categorised	No bleeding	1792	157	90.3%	88.2%	Reference
	1x bleeding	179	16	9.0%	9.0%	0.9417
	>=2x bleeding	13	5	0.7%	2.8%	0.0055
number of vaginal bleeds 5-9 days before 15w SCOPE visit, categorised	No bleeding	1949	171	98.2%	96.1%	Reference
	1x bleeding	32	7	1.6%	3.9%	0.0315
	>=2x bleeding	3	0	0.2%	0.0%	0.9810
number of vaginal bleeds >=10 days before 15w SCOPE visit, categorised	No bleeding	1951	171	98.3%	96.1%	Reference
	1x bleeding	32	7	1.6%	3.9%	0.0313
	>=2x bleeding	1	0	0.1%	0.0%	0.9751
Gestational age when 1st vaginal bleed occurred before 15w SCOPE visit, categorised	No Bleeding	1606	137	81.0%	77.0%	Reference
	1st bleed between 1-6wk	192	21	9.7%	11.8%	0.3133
	1st bleed between 7-12wk	153	13	7.7%	7.3%	0.9895
	1st bleed after 12wk	33	7	1.7%	3.9%	0.0323
Gestational age when 2nd vaginal bleed occurred before 15w SCOPE visit, categorised	No Bleeding	1902	166	95.9%	93.3%	Reference
	2nd bleed between 1-6wk	14	5	0.7%	2.8%	0.0075
	2nd bleed between 7-12wk	56	6	2.8%	3.4%	0.6390

	2nd bleed after 12wk	12	1	0.6%	0.6%	0.9647
Any vaginal bleeding continuing for at least 5 days before 15w SCOPE visit	No	1917	165	96.6%	92.7%	Reference
	Yes	67	13	3.4%	7.3%	0.0096
Any vaginal bleeding continuing for at least 10 days before 15w SCOPE visit	No	1951	171	98.3%	96.1%	Reference
	Yes	33	7	1.7%	3.9%	0.0370
Any spotting or light vaginal bleeding continuing for at least 5 days before 15w SCOPE visit	No	1919	166	96.7%	93.3%	Reference
	Yes	65	12	3.3%	6.7%	0.0195
Any spotting or light vaginal bleeding continuing for at least 10 days before 15w SCOPE visit	No	1953	171	98.4%	96.1%	Reference
	Yes	31	7	1.6%	3.9%	0.0262
Number of hospital admissions due to hyperemesis before 15w SCOPE visit, categorised	No admission due to hyperemesis (includes women with no hospital admission for any reason)	1949	170	98.2%	95.5%	Reference
	1x admission due to Hyperemesis	30	7	1.5%	3.9%	0.0213
	>=2x admission due to Hyperemesis	5	1	0.3%	0.6%	0.4499
Any hospital admissions due to hyperemesis before 15w SCOPE visit	No admission due to hyperemesis (includes women with no hospital admission for any reason)	1949	170	98.2%	95.5%	Reference
	Admission due to Hyperemesis	35	8	1.8%	4.5%	0.0160
Frequency consumed fruit in the month prior to conception, compressed categories (5 gps)	>=1x per day	1290	90	65.0%	50.6%	Reference
	3-6x per week	262	29	13.2%	16.3%	0.0394
	1-2x per week	269	37	13.6%	20.8%	0.0010

	1-3x per month	96	15	4.8%	8.4%	0.0069
	Never	67	7	3.4%	3.9%	0.3269
high (>=3 times per day) fruit consumption in the month prior to conception	No	1472	151	74.2%	84.8%	Reference
	Consumed fruit >=3x per day	512	27	25.8%	15.2%	0.0020
Frequency consumed fruit in pregnancy prior to 15w SCOPE visit, compressed categories (5 gps)	>=1x per day	1537	119	77.5%	66.9%	Reference
	3-6x per week	253	31	12.8%	17.4%	0.0310
	1-2x per week	121	17	6.1%	9.6%	0.0308
	1-3x per month	45	8	2.3%	4.5%	0.0355
	Never	28	3	1.4%	1.7%	0.5973
Frequency consumed green leafy vegetables in the month prior to conception, compressed categories (5 gps)	>=1x per day	1062	69	53.5%	38.8%	Reference
	3-6x per week	569	65	28.7%	36.5%	0.0018
	1-2x per week	250	32	12.6%	18.0%	0.0026
	1-3x per month	65	8	3.3%	4.5%	0.1056
	Never	38	4	1.9%	2.3%	0.3717
high (>=3 times per day) green leafy vegetables consumption in the month prior to conception	No	1762	169	88.8%	94.9%	Reference
	Consumed >=3x green leafy veges per day	222	9	11.2%	5.1%	0.0137
Frequency consumed green leafy vegetables in pregnancy prior to 15w SCOPE visit, compressed categories (5 gps)	>=1x per day	969	63	48.8%	35.4%	Reference
	3-6x per week	586	67	29.5%	37.6%	0.0021
	1-2x per week	304	34	15.3%	19.1%	0.0149
	1-3x per month	70	8	3.5%	4.5%	0.1535
	Never	55	6	2.8%	3.4%	0.2492

High (>=3 times per day) green leafy vegetables consumption in pregnancy prior to 15w SCOPE visit	No	1774	170	89.4%	95.5%	Reference
	Consumed >=3x green leafy veges per day	210	8	10.6%	4.5%	0.0124
	1-3/mth	486	49	39.8%	42.6%	0.0140
	1 or 2/wk	340	34	27.9%	29.6%	0.0208
	3 or 4/wk	59	10	4.8%	8.7%	0.0034
	5 or 6/wk	8	4	0.7%	3.5%	0.0004
	1-2/day	10	3	0.8%	2.6%	0.0096
	3-4/day	3	0	0.3%	0.0%	0.9820
Frequency consumed burger in the month prior to conception	Never	314	15	25.7%	13.0%	Reference
	1-3x per month	486	49	39.8%	42.6%	0.0139
	1-2x per week	340	34	27.9%	29.6%	0.0208
	3-6x per week	67	14	5.5%	12.2%	0.0002
	>=1x per day	13	3	1.1%	2.6%	0.0230
Frequency consumed burger in pregnancy prior to 15w SCOPE visit	Never	369	18	30.2%	15.7%	Reference
	1-3x per month	503	55	41.2%	47.8%	0.0040
	1-2x per week	303	34	24.8%	29.6%	0.0057
	3-6x per week	41	8	3.4%	7.0%	0.0023
	>=1x per day	6	0	0.5%	0.0%	0.9832
Frequency consumed fried chicken in the month prior to conception	Never	745	52	61.0%	45.2%	Reference
	1-3x per month	351	52	28.8%	45.2%	0.0003
	1-2x per week	110	11	9.0%	9.6%	0.3004

	3-6x per week	12	0	1.0%	0.0%	0.9840
	>=1x per day	3	0	0.3%	0.0%	0.9920
Frequency consumed fried chicken in pregnancy prior to 15w SCOPE visit , categorised compressed	Never	873	67	71.5%	58.3%	Reference
	1-3x per month	288	43	23.6%	37.4%	0.0013
	1-2x per week	54	3	4.4%	2.6%	0.5942
	3-6x per week	4	2	0.3%	1.7%	0.0323
	>=1x per day	2	0	0.2%	0.0%	0.9847
Frequency consumed hot chips/french fries in the month prior to conception, compressed categorised	Never	228	10	18.7%	8.7%	Reference
	1-3x per month	550	44	45.2%	38.3%	0.0941
	1-2x per week	349	44	28.7%	38.3%	0.0034
	3-6x per week	76	16	6.2%	13.9%	0.0002
	>=1x per day	15	1	1.2%	0.9%	0.6988
Frequency consumed hot chips/french fries in pregnancy prior to 15w SCOPE visit	Never	206	17	16.9%	14.8%	Reference
	1-3x per month	534	39	43.7%	33.9%	0.6858
	1-2x per week	370	41	30.3%	35.7%	0.3279
	3-6x per week	98	18	8.0%	15.7%	0.0262
	>=1x per day	13	0	1.1%	0.0%	0.9831
Chinese treatment used as alternative therapy at 15w SCOPE visit	No alternative therapy used or no chinese treatment	1980	176	99.8%	98.9%	Reference
	Yes	4	2	0.2%	1.1%	0.0470
units of alcohol per week in the 3 months pre-pregnancy divided in 3 grades of severity	No Alcohol	575	75	29.0%	42.1%	Reference
	Low Alcohol consumption (<=2units/day or	1307	91	65.9%	51.1%	0.0001

	>=14units/week)					
	High Alcohol consumption (>2 units/day or >14 units/wk)	102	12	5.1%	6.7%	0.7538
units of alcohol per week in the 1st trimester (categories)	No alcohol	1012	115	51.0%	64.6%	Reference
	2 月 01 日	337	18	17.0%	10.1%	0.0038
	7 月 03 日	398	28	20.1%	15.7%	0.0285
	Aug-14	164	8	8.3%	4.5%	0.0242
	>14	73	9	3.7%	5.1%	0.8241
Alcohol consumption status at 15w SCOPE visit in 3 groups (consumption during pregnancy only)	No alcohol in pregnancy	1002	113	50.5%	63.5%	Reference
	Quit alcohol in pregnancy	888	58	44.8%	32.6%	0.0011
	Continuing to drink alcohol at 15 weeks	94	7	4.7%	3.9%	0.3045
Any alcohol consumption in 1st trimester	No	1012	115	51.0%	64.6%	Reference
	Yes	972	63	49.0%	35.4%	0.0006
Chronic hypertension' defined as repeated systolic BP>=140 mmHg or repeated diastolic BP>=140 mmHg at 1st SCOPE visit	No	1980	175	99.8%	98.3%	Reference
	Yes	4	3	0.2%	1.7%	0.0054
Pulse <=60/min at 15w SCOPE visit	No	1798	174	90.7%	97.8%	Reference
	Yes	184	4	9.3%	2.3%	0.0035
Short Height (<161 cm)	>=161 cm	1531	125	77.2%	70.2%	Reference
	<161 cm	453	53	22.8%	29.8%	0.0369
Snored most nights, evaluated at 15w SCOPE visit	No	1338	115	67.4%	64.6%	Reference
	Yes	215	35	10.8%	19.7%	0.0020

Snored most nights (binary) evaluated at 15w SCOPE visit	No or Unknown	1769	143	89.2%	80.3%	Reference
	Yes	215	35	10.8%	19.7%	0.0005
Engaged in less vigorous exercise (the woman did not breathe harder or puff or pant) in the last month, evaluated at 15w SCOPE visit	Never	453	54	23.0%	30.3%	Reference
	Once a week	557	47	28.3%	26.4%	0.0987
	2-3 / wk	621	50	31.5%	28.1%	0.0565
	4-6 x /wk	171	17	8.7%	9.6%	0.5344
	Daily	150	7	7.6%	3.9%	0.0230
	More than once a day	19	3	1.0%	1.7%	0.6594
Engaged in less vigorous exercise (the woman did not breathe harder or puff or pant) in the last month, evaluated at 15w SCOPE visit, compressed into 3 categories	Never	453	54	23.0%	30.3%	Reference
	1-3 times/ week	1178	97	59.8%	54.5%	0.0383
	>=4 times / wk	340	27	17.3%	15.2%	0.0992
Number of times climbed stairs in the last month, evaluated at 15w SCOPE visit	Never	711	81	36.1%	45.5%	Reference
	<10x/day	1056	85	53.6%	47.8%	0.0327
	>=10x/day	205	12	10.4%	6.7%	0.0370
Number of hours spent using a computer per day in the month prior to the 15±1 week SCOPE interview	None	194	28	9.8%	15.7%	Reference
	<2h	557	49	28.2%	27.5%	0.0487
	2-4h	290	25	14.7%	14.0%	0.0759
	5-6h	329	25	16.7%	14.0%	0.0268
	>6h	603	51	30.6%	28.7%	0.0320
Any computer usage in last month evaluated at 15w SCOPE visit	NO	194	28	9.8%	15.7%	Reference
	YES	1779	150	90.2%	84.3%	0.0142

Never used a computer in the last month evaluated at 15w SCOPE visit	NO	1779	150	90.2%	84.3%	Reference
	YES	194	28	9.8%	15.7%	0.0142
Behaviour Responses to Pregnancy': Put parts of life on hold since pregnant evaluated at 15w SCOPE visit	Not at all	510	56	25.9%	31.5%	Reference
	Rarely	586	58	29.7%	32.6%	0.5980
	Some days	572	46	29.0%	25.8%	0.1345
	Most days	235	14	11.9%	7.9%	0.0479
	Every day	68	4	3.5%	2.3%	0.2419
Behaviour Responses to Pregnancy': Not slowed down since pregnant evaluated at 15w SCOPE visit	Not at all	349	23	17.7%	12.9%	Reference
	Rarely	390	34	19.8%	19.1%	0.3174
	Some days	571	53	29.0%	29.8%	0.1857
	Most days	506	43	25.7%	24.2%	0.3420
	Everyday	156	25	7.9%	14.0%	0.0035
Behaviour Responses to Pregnancy': Gone to bed during day evaluated at 15w SCOPE visit	Not at all	384	29	19.5%	16.3%	Reference
	Rarely	560	40	28.4%	22.5%	0.8255
	Some days	799	76	40.5%	42.7%	0.3093
	Most days	184	27	9.3%	15.2%	0.0185
	Everyday	45	6	2.3%	3.4%	0.2318
I have felt better than ever in pregnancy evaluated at 15w SCOPE visit, compressed categories	Not at all	454	33	23.1%	18.5%	Reference
	Sometimes	1224	109	62.1%	61.2%	0.3246
	At least most days	292	36	14.8%	20.2%	0.0363
STAI: I feel content evaluated at 15w SCOPE	Very much	878	88	44.5%	49.4%	Reference

visit	Moderately	698	46	35.4%	25.8%	0.0264
	Somewhat	325	38	16.5%	21.4%	0.4516
	Not at all	72	6	3.7%	3.4%	0.6744
Depression Scale: Anxious for no reason evaluated at 15w SCOPE visit	Not at all	409	29	20.7%	16.3%	Reference
	Not much	666	62	33.8%	34.8%	0.2438
	Sometimes	791	72	40.1%	40.5%	0.2737
	Quite a lot	107	15	5.4%	8.4%	0.0425
Depression Scale: Things getting on top evaluated at 15w SCOPE visit	"No, coping well"	410	35	20.8%	19.7%	Reference
	"No, mostly coping"	970	88	49.2%	49.4%	0.7702
	"Yes, sometimes not coping"	566	48	28.7%	27.0%	0.9773
	"Yes, mostly not coping"	27	7	1.4%	3.9%	0.0156
Depression Scale: So unhappy that been crying evaluated at 15w SCOPE visit	Never	1029	81	52.2%	45.5%	Reference
	Only occasionally	811	79	41.1%	44.4%	0.1964
	Quite often	112	17	5.7%	9.6%	0.0211
	Most of the time	21	1	1.1%	0.6%	0.6255
Support people around to provide emotional support evaluated at 15w SCOPE visit	All the time	1339	114	67.9%	64.0%	Reference
	Most of the time	465	51	23.6%	28.7%	0.1521
	Sometimes	136	9	6.9%	5.1%	0.4813
	Seldom	27	2	1.4%	1.1%	0.8506
	Never	4	2	0.2%	1.1%	0.0422
Current work situation at 20w SCOPE visit	Full time work	1315	104	67.8%	59.4%	Reference

	Part time work	316	31	16.3%	17.7%	0.3141
	Student	53	2	2.7%	1.1%	0.3091
	Homemaker	87	10	4.5%	5.7%	0.2842
	Unemployed	150	27	7.7%	15.4%	0.0004
	Sickness beneficiary	6	1	0.3%	0.6%	0.4920
	Other (e.g.) voluntary work	12	0	0.6%	0.0%	0.9753
Paid employment working full-time at 20w SCOPE visit	No	624	71	32.2%	40.6%	Reference
	Yes	1315	104	67.8%	59.4%	0.0243
Unemployed or Sickness Beneficiary at 20w SCOPE visit	No	1783	147	92.0%	84.0%	Reference
	Yes	156	28	8.1%	16.0%	0.0005
Work status at 20w SCOPE visit (3 groups)	No paid work	308	40	15.9%	22.9%	Reference
	Part time work	316	31	16.3%	17.7%	0.2662
	Full time work	1315	104	67.8%	59.4%	0.0116
Hyperemesis continuing at SCOPE 20w visit	No Hyperemesis	1816	154	93.6%	88.0%	Reference
	Hyperemesis at or before 15w visit, no vomiting afterwards	70	12	3.6%	6.9%	0.0296
	Hyperemesis at or before 15w vst, ongoing between 15w and 20w visit	16	3	0.8%	1.7%	0.2113
	New onset vomiting between 15w and 20w visit	38	6	2.0%	3.4%	0.1646
Any vaginal bleeding between 15w and 20w SCOPE visits	No	1890	165	97.4%	94.3%	Reference
	Yes	50	10	2.6%	5.7%	0.0198

Any single episode of vaginal bleeding between 15w and 20w SCOPE visits	No	1893	166	97.6%	94.9%	Reference
	Yes	47	9	2.4%	5.1%	0.0361
Any spotting or light vaginal bleeding between 15w-20w SCOPE visit	No	1891	166	97.5%	94.9%	Reference
	Yes	49	9	2.5%	5.1%	0.0469
Any vaginal bleeding between 13 weeks' gestation and 20w SCOPE visit (could be recorded at either 15w or 20w visit)	No	1844	160	95.1%	91.4%	Reference
	Yes	96	15	5.0%	8.6%	0.0422
Any hospital admissions due to hyperemesis between the 15w and 20w SCOPE visits	No admission due to hyperemesis (includes women with no hospital admission for any reason)	1938	173	99.9%	98.9%	Reference
	Admission due to Hyperemesis	2	2	0.1%	1.1%	0.0160
Frequency consumed oily fish between 15w and 20w SCOPE visits, 3 severity grades	Often	112	3	5.8%	1.7%	Reference
	Moderate	538	54	27.8%	30.9%	0.0283
	Rarely	1286	118	66.4%	67.4%	0.0378
High consumption oily fish (>=3 times a week) between 15w and 20w SCOPE visits	No	1824	172	94.2%	98.3%	Reference
	Yes	112	3	5.8%	1.7%	0.0330
Frequency consumed fruit between the 15w and 20w SCOPE visits, compressed categories (5 gps)	>=1x per day	1566	125	80.7%	71.4%	Reference
	3-6x per week	189	21	9.7%	12.0%	0.1825
	1-2x per week	120	18	6.2%	10.3%	0.0192
	1-3x per month	38	10	2.0%	5.7%	0.0012
	Never	27	1	1.4%	0.6%	0.4525
Low fruit consumption (<3x times/mth) between	No	1875	164	96.7%	93.7%	Reference

the 15w and 20w SCOPE visit	Yes	65	11	3.4%	6.3%	0.0495
Frequency consumed green leafy vegetables between 15w and 20w SCOPE visits	>=1x per day	1079	78	55.6%	44.6%	Reference
	3-6x per week	525	55	27.1%	31.4%	0.0437
	1-2x per week	265	35	13.7%	20.0%	0.0050
	1-3x per month	37	3	1.9%	1.7%	0.8512
	Never	34	4	1.8%	2.3%	0.3684
	1-3/mth	433	46	36.3%	41.4%	0.0262
	1 or 2/wk	300	32	25.2%	28.8%	0.0379
	3 or 4/wk	23	6	1.9%	5.4%	0.0035
	5 or 6/wk	3	0	0.3%	0.0%	0.9816
	1-2/day	2	1	0.2%	0.9%	0.0884
Frequency consumed burgers between the 15w and 20w SCOPE visits	Never	431	26	36.2%	23.4%	Reference
	1-3x per month	433	46	36.3%	41.4%	0.0262
	1-2x per week	300	32	25.2%	28.8%	0.0379
	3-6x per week	26	6	2.2%	5.4%	0.0068
	>=1x per day	2	1	0.2%	0.9%	0.0884
Frequency consumed fried chicken between the 15w and 20w SCOPE visits	Never	943	74	79.2%	66.7%	Reference
	1-3x per month	204	33	17.1%	29.7%	0.0012
	1-2x per week	39	3	3.3%	2.7%	0.9740
	3-6x per week	4	1	0.3%	0.9%	0.3028
	>=1x per day	1	0	0.1%	0.0%	0.9836

Frequency consumed curries between 15w and 20w SCOPE visits	Never	700	77	58.7%	69.4%	Reference
	1-3x per month	358	29	30.0%	26.1%	0.1784
	1-2x per week	109	3	9.1%	2.7%	0.0204
	3-6x per week	14	1	1.2%	0.9%	0.6786
	>=1x per day	11	1	0.9%	0.9%	0.8561
Chinese treatment used as alternative therapy at 20w SCOPE visit	No	1939	173	100.0%	98.9%	Reference
	Yes	1	2	0.1%	1.1%	0.0113
Any alcohol consumption at 20w SCOPE visit (week prior to interview)	No	1793	169	92.4%	96.6%	Reference
	Yes	147	6	7.6%	3.4%	0.0484
	Yes	1371	138	70.7%	78.9%	0.0227
	Yes	1474	151	76.0%	86.3%	0.0024
Main activities at work evaluated at 20w SCOPE visit	"Administrative, sitting activities"	578	53	35.2%	39.6%	Reference
	Sitting and some walking	539	30	32.8%	22.4%	0.0345
	Standing	32	1	2.0%	0.8%	0.2933
	Standing/walking	330	32	20.1%	23.9%	0.8113
	Standing/walking/intermittent exercise	158	16	9.6%	11.9%	0.7399
	Regular exercise	6	2	0.4%	1.5%	0.1195
PSS: Upset because of something that happened unexpectedly evaluated at 20w SCOPE visit	Never	299	36	15.5%	20.9%	Reference
	Almost never	776	55	40.1%	32.0%	0.0185
	Sometimes	679	63	35.1%	36.6%	0.2366
	Fairly often	153	16	7.9%	9.3%	0.6561

	Very often	27	2	1.4%	1.2%	0.5193
PSS: Confident about ability to handle personal problems evaluated at 20w SCOPE visit	Never	622	56	32.2%	32.6%	Reference
	Almost never	863	66	44.7%	38.4%	0.3883
	Sometimes	335	31	17.3%	18.0%	0.9066
	Fairly often	73	14	3.8%	8.1%	0.0194
	Very often	40	5	2.1%	2.9%	0.5069
Snores most night, evaluated at 20w SCOPE visit	No	1214	100	62.8%	57.8%	Reference
	Yes	259	36	13.4%	20.8%	0.0111
Number of hours spent using a computer per day in the last month, evaluated at 20w SCOPE visit	None	183	29	9.5%	16.8%	Reference
	<2h	564	42	29.2%	24.3%	0.0032
	2-4h	295	29	15.3%	16.8%	0.0870
	5-6h	371	37	19.2%	21.4%	0.0793
	>6h	521	36	26.9%	20.8%	0.0017
Any computer usage in last month evaluated at 20w SCOPE visit	NO	183	29	9.5%	16.8%	Reference
	YES	1751	144	90.5%	83.2%	0.0026
'Behaviour Responses to Pregnancy' :Obliged to carry out responsibilities no matter how bad she feels, evaluated at 20w SCOPE visit	Not at all	476	51	24.6%	29.5%	Reference
	Rarely	576	43	29.8%	24.9%	0.0945
	Some days	551	60	28.5%	34.7%	0.9356
	Most days	260	15	13.5%	8.7%	0.0415
	Everyday	70	4	3.6%	2.3%	0.2398
'Behaviour Responses to Pregnancy' :Avoided usual activities, evaluated at 20w SCOPE visit	Not at all	523	58	27.1%	33.5%	Reference

	Rarely	785	59	40.6%	34.1%	0.0442
	Some days	522	49	27.0%	28.3%	0.4130
	Most days	86	6	4.5%	3.5%	0.2970
	Everyday	16	1	0.8%	0.6%	0.5814
STAI: I feel worried, evaluated at 20w SCOPE visit	Very much	1008	108	52.2%	62.8%	Reference
	Moderately	677	48	35.1%	27.9%	0.0221
	Somewhat	203	14	10.5%	8.1%	0.1344
	Not at all	43	2	2.2%	1.2%	0.2532
Social support (listening ears and practical support scores added) categorised	2	1045	105	54.1%	60.7%	Reference
	3	340	19	17.6%	11.0%	0.0224
	4	338	34	17.5%	19.7%	0.9957
	5	119	10	6.2%	5.8%	0.6042
	>5	91	5	4.7%	2.9%	0.1996
Femur length on 19-20w scan transformed to Z score for gestational age <10th Centile	No	1792	155	91.9%	87.1%	Reference
	Yes	158	23	8.1%	12.9%	0.0290
Umbilical artery Doppler RI at 19-21w, transformed to MoM by gestation >90th Centile	No	1715	147	89.8%	84.5%	Reference
	yes	194	27	10.2%	15.5%	0.0295
Umbilical artery Doppler RI at 19-21w >90th centile	No (Normal)	1719	147	90.1%	84.5%	Reference
	Yes (Abnormal)	190	27	10.0%	15.5%	0.0227
Right notch at 19-21w	Absent	1517	118	79.8%	68.2%	Reference
	Present	301	47	15.8%	27.2%	0.0001

	Indeterminate	83	8	4.4%	4.6%	0.5749
	Present	301	47	15.2%	26.4%	0.0001
Left notch at 19-21w	Absent	1453	114	76.5%	66.7%	Reference
	Present	355	48	18.7%	28.1%	0.0028
	Indeterminate	92	9	4.8%	5.3%	0.5428
Bilateral notch at 19-21w	Not bilateral Notch	1770	145	90.5%	81.5%	Reference
	Bilateral Notch	185	33	9.5%	18.5%	0.0002
Any vaginal bleeding in pregnancy (includes any bleeding recorded at 1st or 2nd SCOPE visit in addition to bleeding since 2nd visit)	No	1574	129	79.3%	72.5%	Reference
	Yes	410	49	20.7%	27.5%	0.0328
Urinary tract infection (lower) in pregnancy between 20w SCOPE and delivery	No	1925	165	97.0%	92.7%	Reference
	Yes	59	13	3.0%	7.3%	0.0029
Pyelonephritis infection in pregnancy between 20w SCOPE and delivery	No	1984	178	100.0%	100.0%	Reference
	Yes	0	0	0.0%	0.0%	0.0029
Gastroenteritis in pregnancy between 20w SCOPE and delivery	No	1894	162	95.5%	91.0%	Reference
	Yes	90	16	4.5%	9.0%	0.0098
Proven Vaginal Candida infection in pregnancy between 20w SCOPE and delivery	No	1735	143	87.5%	80.3%	Reference
	Yes	249	35	12.6%	19.7%	0.0077
Asthma exacerbation in pregnancy	Not applicable (did not have Asthma)	1533	124	77.3%	69.7%	Reference
	No	394	50	19.9%	28.1%	0.0109
	"Yes no oral steroids	53	4	2.7%	2.3%	0.8954
	"Yes x1,used oral steroids	1	0	0.1%	0.0%	0.9891

	Yes x >=2 oral steroids	2	0	0.1%	0.0%	0.9846
Apgar score < 7 at 5 minutes	No	1974	174	99.6%	97.8%	Reference
	Yes	8	4	0.4%	2.3%	0.0049
Other pregnancy complication; this would include other complications such as antepartum haemorrhage, asthma exacerbation, pyelonephritis.	No	1828	172	92.1%	96.6%	Reference
	Yes	156	6	7.9%	3.4%	0.0346

IV. Significant Cochran-Mantel-Haenszel (CMH) Test Results for PE

Variable	P	Variable	P
Participant's birthweight (g)	0.004	Family history (mother or sisters) of recurrent GH or recurrent PET	0.0050
number of vaginal bleeds lasting >=10 days before 15w SCOPE visit	0.019	Participant's mother had any history of LBW baby	0.0244
Number of episodes of vag spotting that last >=5 days before 15w SCOPE visit	0.002	Family history of LBW baby, i.e. participant's mother or sister had had LBW baby	0.0036
Folate dose (?g per day) at 15w SCOPE visit	0.039	Any sister with a history of PTB (all-spontaneous or iatrogenic)	0.0230
number of cigarettes per day in the 3 months pre-pregnancy	0.020	participant's father has chronic hypertension	0.0107
number of cigarettes per day in the 1st trimester	0.016	Participant's father has one or more of type 2 diabetes, chronic hypertension, CVA and IHD	0.0033
Total number of cigarettes a woman was exposed to in the 1st trimester	0.042	Participant's mother or father has one or more of type 2 diabetes, chronic hypertension, CVA and IHD	0.0035
Number of weeks of alcohol exposure in pregnancy prior to 15w SCOPE visit	0.019	participant's sibling(s) has type 2 diabetes	0.0178
Number of times other recreational drugs was taken over the 3m pre-pregnancy	0.050	Family history of chronic hypertension (participant's mother, father, sibling)	0.0083
Number of times other recreational drugs was taken in the 1st trimester	0.001	Family history of ischaemic heart disease (participant's mother, father, sibling)	0.0002
Gestation other recreational drugs ceased in pregnancy	0.001	Family history of cerebrovascular accident (participant's mother, father, sibling)	0.0417
pulse per minute at 15w SCOPE visit	0.044	Any vaginal bleeding continuing for at least 5 days before 15w SCOPE visit	0.0040
Stool Height at 15w SCOPE visit (cm)	0.016	Any vaginal bleeding continuing for at least 10 days before 15w SCOPE visit	0.0186
Total days of vaginal bleeding between 15w and 20w SCOPE visits	0.018	Any spotting or light vaginal bleeding continuing for at least 5 days before 15w SCOPE visit	0.0092

Duration 1st vaginal bleed (days) between 15w and 20w SCOPE visits	0.018	Any spotting or light vaginal bleeding continuing for at least 10 days before 15w SCOPE visit	0.0105
Participant's main Ethnicity is Pacific Island	0.0008	high (>=3 times per day) fruit consumption in pregnancy prior to 15w SCOPE visit	0.0168
Participant's ethnicity Pacific Islander (recorded as Pacific Islander under either Main Ethnicity or Other Ethnicity)	0.0082	Any alcohol consumption in 1st trimester	0.0095
Participant's birthweight<2500g	0.0036	Any other recreational drugs in the 1st trimester	0.0015
Previous Etopic Pregnancy	0.0179	Any use of other recreational drugs during pregnancy	0.0015
Any previous miscarriage at >10 weeks	0.0458	Chronic hypertension' defined as repeated systolic BP>=140 mmHg or repeated diastolic BP>=140 mmHg at 1st SCOPE visit	0.0024
Previous ectopic pregnancy with same man who has fathered the current pregnancy	0.0195	Pulse <=60/min at 15w SCOPE visit	0.0247
A single previous miscarriage at <=10 wks gestation with same man who has fathered the current pregnancy	0.0223	Pulse >=84/min at 15w SCOPE visit	0.0037
Any previous miscarriage at <=10 wks gestation with same man who has fathered the current pregnancy	0.0383	Snored most nights (binary) evaluated at 15w SCOPE visit	0.0061
Months of sexual relationship prior to conception with the biological father of the baby <=6 months	0.0112	Any vaginal bleeding between 15w and 20w SCOPE visits	0.0181
Months of sexual relationship without barrier contraception with biological father of baby <=3 months	0.0165	Any single episode of vaginal bleeding between 15w and 20w SCOPE visits	0.0365
Months of sexual relationship without barrier contraception with biological father of baby <=6 months	0.0083	high fruit consumption (>=3 times per day) between the 15w and 20w SCOPE visit	0.0177
Self reported hypertension (on more than 1 occasion) while on oral contraception	0.0160	Chinese treatment used as alternative therapy at 20w SCOPE visit	0.0011
Participant's mother had any history of pregnancy induced hypertension (GH or preeclampsia, but unsure which condition), gestational hypertension or preeclampsia	0.0010		
Any sister had a history of GH	0.0474		
Strong family history (2 or more members of immediate family -mother or sisters) of GH	0.0127		
Participant's mother had any history of PET	0.0054		

Participant's mother had any history of recurrent PET	0.0025
Any sister has a history of PET	0.0296
Family history of PET, i.e. participant's mother or sister had had PET	0.0009
Family history of recurrent PET, i.e. participant's mother or sister had had recurrent PET	0.0165
Family history (mother or sisters) of GH or PET	0.0010

V. Exploratory Analysis of Preterm Birth

Variable	Uncomplicated pregnancy N	Preterm birth N	Uncomplicated pregnancy Mean ± SE	Preterm birth Mean ± SE	P
Number of years ago that participant migrated to current country	492	27	9.3 ± 9.41	13.2 ± 11.46	0.0446
Total years of schooling (primary and secondary, not pre-school or tertiary)	1984	156	12.3 ± 1.05	12.0 ± 1.32	0.0005
Participant's birthweight (g)	1872	141	3332.8 ± 529.93	3228.7 ± 619.42	0.0263
Participant's partner's birthweight (g). Partner refers to biological father of current fetus.	1702	130	3485.4 ± 588.07	3375.3 ± 612.95	0.0402
Participant's gestation at delivery (wks)	1922	145	39.8 ± 1.79	39.3 ± 2.30	0.0012
Gravidity	1984	156	1.3 ± 0.62	1.5 ± 0.76	0.0015
Number of D&C or surgical terminations of pregnancy i.e. Number of cervical dilatations	1984	156	0.2 ± 0.43	0.3 ± 0.56	0.0003
Duration of sex without contraception before conception with father of baby	1980	156	5.7 ± 11.58	9.1 ± 16.63	0.0012
number of vaginal bleeds before 15w SCOPE visit	1984	156	0.3 ± 0.61	0.4 ± 0.82	0.0045
number of vaginal bleeds before 15w SCOPE visit, light severity	1984	156	0.0 ± 0.24	0.1 ± 0.46	0.0076
number of vaginal bleeds before 15w SCOPE visit, mod-heavy severity	1984	156	0.0 ± 0.15	0.1 ± 0.31	0.0095
number of vaginal bleeds commencing >12 weeks	1984	156	0.0 ± 0.17	0.1 ± 0.33	0.0001
number of vaginal bleeds lasting <=1 day before 15w SCOPE visit	1984	156	0.1 ± 0.43	0.2 ± 0.56	0.0394
Total days of vaginal bleeding before 15w SCOPE visit	1984	156	0.8 ± 3.16	1.5 ± 4.90	0.0146
Total days of vaginal bleeding before 15w SCOPE visit, light	1984	156	0.2 ± 1.31	0.5 ± 3.11	0.0148
Total days of vaginal bleeding before 15w SCOPE visit, spotting or light	1984	156	0.7 ± 3.01	1.3 ± 4.10	0.0471

Total days of vaginal bleeding before 15w SCOPE visit, mod or heavy	1984	156	0.0 ± 0.59	0.2 ± 1.83	0.0298
Total days of vaginal bleeding after 12 weeks gestation	1984	156	0.1 ± 0.47	0.2 ± 1.45	0.0036
Duration 1st vaginal bleed (days) before 15w SCOPE visit	1984	156	0.6 ± 2.71	1.1 ± 3.90	0.0425
Duration 3rd vaginal bleed (days) before 15w SCOPE visit	1632	117	0.0 ± 0.46	0.2 ± 1.37	0.0313
Number of episodes of light vaginal bleeding at or before 6w gestation	1984	156	0.0 ± 0.14	0.0 ± 0.21	0.0366
Number of episodes of mod-heavy vaginal bleeding at 7-12w gestation	1984	156	0.0 ± 0.09	0.0 ± 0.16	0.0414
Number of episodes of spotting vaginal bleeding after 12w gestation	1984	156	0.0 ± 0.15	0.1 ± 0.27	0.0025
Number of episodes of mod-heavy vaginal bleeding after 12w gestation	1984	156	0.0 ± 0.03	0.0 ± 0.18	0.0156
Number of episodes of mod-heavy bleeding that last >=5 days before 15w SCOPE visit	1984	156	0.0 ± 0.05	0.0 ± 0.18	0.0254
Total duration of light vag bleeding at or before 6w gestation	1984	156	0.1 ± 1.00	0.3 ± 2.85	0.0428
Total number of spotting or light vaginal bleeding before 15w SCOPE visit	1984	156	0.2 ± 0.58	0.4 ± 0.75	0.0146
Total number of vag bleeding after 6 weeks gestation	1984	156	0.1 ± 0.45	0.3 ± 0.65	0.0055
Total duration of vag bleeding after 6w gestation (days)	1984	156	0.4 ± 1.78	0.8 ± 2.85	0.0111
Number of hospital admissions due to vaginal bleeding before 15w SCOPE visit	1984	156	0.0 ± 0.13	0.0 ± 0.26	0.0220
Number of hospital admissions due to trauma before 15w SCOPE visit	1984	156	0.0 ± 0.02	0.0 ± 0.11	0.0081
Number of hospital admissions due to any other reasons before 15w SCOPE visit	1984	156	0.0 ± 0.04	0.0 ± 0.11	0.0325
Folate dose (?g per day) in 1st trimester	1984	156	730.1 ± 550.73	631.7 ± 454.47	0.0211
number of cigarettes per day in the 3 months pre-pregnancy	1984	156	2.6 ± 6.33	4.2 ± 8.12	0.0040
number of cigarettes per day in the 1st trimester	1984	156	1.9 ± 4.93	3.0 ± 6.17	0.0103
Number of weeks of cigarette exposure in pregnancy prior to 15w SCOPE visit	1980	155	2.0 ± 4.63	3.5 ± 6.14	0.0002

number of cigarettes per day at 15w SCOPE visit	1984	156	0.6 ± 2.51	1.4 ± 4.24	0.0008
Total number of cigarettes a woman was exposed to in the 1st trimester	1980	155	143.9 ± 439.55	265.8 ± 612.00	0.0018
Gestation ceased other recreational drugs (binge drinking ie>=6 units/session or illicit drugs)	1984	156	0.8 ± 2.52	1.8 ± 4.39	0.0000
Number of times marijuana was taken over the 3m pre-pregnancy	1982	156	15.4 ± 179.05	87.9 ± 459.71	0.0013
Number of times marijuana was taken in the 1st trimester	1983	156	9.0 ± 131.45	41.4 ± 239.64	0.0342
Gestation marijuana ceased in pregnancy	1982	155	0.4 ± 1.98	1.2 ± 3.91	0.0000
1st diastolic BP at 15w SCOPE visit measured by mercury or aneroid sphygmomanometer	1984	156	63.9 ± 7.92	65.2 ± 8.08	0.0378
2nd systolic BP at 15w SCOPE visit measured by mercury or aneroid sphygmomanometer	1984	156	106.2 ± 9.96	108.2 ± 10.93	0.0186
2nd diastolic BP at 15w SCOPE visit measured by mercury or aneroid sphygmomanometer	1984	156	63.5 ± 7.71	65.1 ± 8.23	0.0173
2nd MAP (mean arterial pressure) at 15w SCOPE visit measured by mercury or aneroid sphygmomanometer	1984	156	77.8 ± 7.53	79.4 ± 8.30	0.0079
mean diastolic BP from the 1st and 2nd BP recordings at 15w SCOPE visit	1984	156	63.7 ± 7.51	65.1 ± 7.85	0.0204
pulse per minute at 15w SCOPE visit	1982	156	75.9 ± 10.75	77.8 ± 10.87	0.0329
Height at 15w SCOPE visit (cm)	1984	156	165.5 ± 6.55	164.0 ± 6.82	0.0052
Measured Sitting Height at 15w SCOPE visit (cm)	1979	156	131.4 ± 4.98	130.3 ± 5.50	0.0071
Waist Height Ratio at 15w SCOPE visit	1980	156	0.5 ± 0.07	0.5 ± 0.08	0.0342
Head circumference (cm) at 15w SCOPE visit	1983	156	56.0 ± 1.63	55.7 ± 1.50	0.0375

Variable	Category	Uncomplicated pregnancy N	Preterm birth N	Uncomplicated pregnancy %	Preterm birth %	P
Participant's Ethnicity	None	1599	124	80.59	79.49	Reference
	Asian	15	1	0.76	0.64	0.8841
	Caucasian	311	26	15.68	16.67	0.7377
	Other	10	3	0.50	1.92	0.0419
	Maori	22	2	1.11	1.28	0.8309
	African	4	0	0.20	0.00	0.9907
	Indian	3	0	0.15	0.00	0.9919
	Pacific Islander	20	0	1.01	0.00	0.9792
Participant's Other Ethnicity is Other	No (Not other)	1974	153	99.50	98.08	Reference
	Yes (other)	10	3	0.50	1.92	0.0414
Participant's immigration history	Family lived in country >=2 generations	1027	94	51.76	60.26	Reference
	Participant not immigrant and family history unknown	13	2	0.66	1.28	0.4985
	1 parent immigrated	313	24	15.78	15.38	0.4563
	Both parents immigrated	139	9	7.01	5.77	0.3369
	Participant immigrated	492	27	24.80	17.31	0.0231
Participant is an immigrant	No (Not an immigrant)	1492	129	75.20	82.69	Reference
	Yes (Immigrant)	492	27	24.80	17.31	0.0370
Participant's partner is an immigrant. Partner refers to	No	1467	129	74.77	82.69	Reference

biological father of fetus.	Yes	495	27	25.23	17.31	0.0284
Years of schooling in categories	<12	321	41	16.18	26.28	Reference
	12,13	1608	112	81.05	71.79	0.0016
	>13	55	3	2.77	1.92	0.1669
Less than 12 years of schooling	No >12	1051	63	52.97	40.38	Reference
	Yes <=12	933	93	47.03	59.62	0.0027
University education status	No	1010	94	50.91	60.26	Reference
	Dropped out	81	10	4.08	6.41	0.4223
	Still attending	68	2	3.43	1.28	0.1123
	Graduated	825	50	41.58	32.05	0.0179
Any tertiary education at a university or other post school institution	No	390	46	19.66	29.49	Reference
	Yes	1594	110	80.34	70.51	0.0037
Current work situation at 15w SCOPE visit	Full time work	1376	102	69.35	65.38	Reference
	Part time work	315	24	15.88	15.38	0.9072
	Student	60	2	3.02	1.28	0.2710
	Homemaker	70	1	3.53	0.64	0.1037
	Unemployed	145	25	7.31	16.03	0.0004
	Sickness beneficiary	10	1	0.50	0.64	0.7763
	Other (e.g.) voluntary work	8	1	0.40	0.64	0.6239
Unemployed or Sickness Beneficiary at 15w SCOPE visit	No	1829	130	92.19	83.33	Reference
	Yes	155	26	7.81	16.67	0.0002

Codes for maternal occupation using the New Zealand Socioeconomic Index guide (Galbraith C, Jenkin G, Davis P, Coope P, New Zealand Social Economic Index 1996 Users Guide, Statistics New Zealand, Wellington, New Zealand)	Elementary occupations	46	7	2.32	4.49	Reference
	Plant/machine operators	28	1	1.41	0.64	0.1858
	Trade workers	13	0	0.66	0.00	0.9730
	Agriculture/fishery workers	13	3	0.66	1.92	0.5829
	Service/sales workers	434	40	21.88	25.64	0.2524
	Clerks	273	25	13.76	16.03	0.2658
	Armed forces	4	1	0.20	0.64	0.6764
	Associate professional/technical	333	26	16.78	16.67	0.1415
	Professionals	503	29	25.35	18.59	0.0304
	Legislators, administrators, managers	337	24	16.99	15.38	0.0969
Household members	Partner	1526	119	76.92	76.28	Reference
	Parents	86	9	4.33	5.77	0.4178
	Relatives	28	7	1.41	4.49	0.0072
	Friends	28	2	1.41	1.28	0.9053
	Alone	37	3	1.86	1.92	0.9489
	Partner & parents	93	6	4.69	3.85	0.6607
	Partner & relatives	107	7	5.39	4.49	0.6618
	Partner & friends	74	2	3.73	1.28	0.1427
	Other	5	1	0.25	0.64	0.3917
Household members categorised into 3 groups	Partner+/-others	1800	134	90.73	85.90	Reference

	Relatives or Friends	147	19	7.41	12.18	0.0336
	Alone	37	3	1.86	1.92	0.8881
Participant and partner's income grouped into categories	<$25K	152	22	9.33	16.79	Reference
	$25K-$74K	555	45	34.05	34.35	0.0356
	$75-$124K	741	43	45.46	32.82	0.0010
	>$125K	182	21	11.17	16.03	0.4846
Type of Maternity Care Code	Public	607	69	30.59	44.23	Reference
	Public/Private Combination	1	0	0.05	0.00	0.9830
Participant’s birthweight<1500g	>=1500g	1859	137	99.31	97.16	Reference
	<1500g	13	4	0.69	2.84	0.0135
Participant’s birthweight<2500g	=2500g	1775	127	94.82	90.07	Reference
	<2500g	97	14	5.18	9.93	0.0195
Participant’s birthweight (g) in categories	<1500gm	13	4	0.69	2.84	Reference
	1500-2499g	84	10	4.49	7.09	0.1517
	2500-3499g	1061	86	56.68	60.99	0.0221
	>-3500g	714	41	38.14	29.08	0.0047
Participant born preterm (<37 weeks)	No	1876	134	95.67	90.54	Reference
	Yes	85	14	4.33	9.46	0.0057
Participant’s gestation at delivery <34wk	No (includes missing participant's gest at delivery)	1900	140	98.86	96.55	Reference
	Yes	22	5	1.14	3.45	0.0252

Gravidity in categories	gravidity=1 i.e. primigravid	1492	102	75.20	65.38	Reference
	gravidity=2	367	35	18.50	22.44	0.1034
	gravidity=3-6	125	19	6.30	12.18	0.0027
Primigravid	Yes	492	54	24.80	34.62	Reference
	No	1492	102	75.20	65.38	0.0072
Any previous pregnancies	No	1492	102	75.20	65.38	Reference
	Yes	492	54	24.80	34.62	0.0072
Number of previous miscarriages (in categories)	No	1727	127	87.05	81.41	Reference
	=1x	222	20	11.19	12.82	0.4185
	>=2x	35	9	1.76	5.77	0.0011
Number of previous terminations (in categories)	No	1707	124	86.04	79.49	Reference
	=1x	231	28	11.64	17.95	0.0203
	>=2x	46	4	2.32	2.56	0.7341
Any previous pregnancy loss with same man who has fathered the current pregnancy	No	1683	122	84.83	78.21	Reference
	Yes	301	34	15.17	21.79	0.0295
Any previous pregnancy loss with a different man from one who has fathered the current pregnancy	No	1748	129	88.10	82.69	Reference
	Yes	236	27	11.90	17.31	0.0490
Any previous pregnancy loss at <=10 weeks gestation	No	1558	110	78.53	70.51	Reference
	Yes	426	46	21.47	29.49	0.0209
Any previous miscarriage or termination at >10 weeks gestation	No	1878	140	94.66	89.74	Reference
	Yes	106	16	5.34	10.26	0.0124

Any previous termination or miscarriage >10wks gestation with a different man from one who has fathered the current pregnancy	No	1927	147	97.13	94.23	Reference
	Yes	57	9	2.87	5.77	0.0485
Any cervical dilatation	No	1719	118	86.64	75.64	Reference
	Yes	265	38	13.36	24.36	0.0002
Any previous pregnancy according to whether the pregnancy was with the same or a different partner	No previous Preg	1492	102	75.20	65.38	Reference
	previous pregnancy with a different partner	191	20	9.63	12.82	0.0963
	previous pregnancy with a different partner and a same partner	45	7	2.27	4.49	0.0497
	previous pregnancy with a same partner	256	27	12.90	17.31	0.0559
Any previous miscarriage pregnancy	No	1727	127	87.05	81.41	Reference
	yes	257	29	12.95	18.59	0.0479
Two previous miscarriages	No	1949	147	98.24	94.23	Reference
	Yes	35	9	1.76	5.77	0.0014
Any previous termination of pregnancy	No	1707	124	86.04	79.49	Reference
	yes	277	32	13.96	20.51	0.0261
Duration of sex without contraception with father of baby before current pregnancy (in categories)	<=3 month	1296	82	65.45	52.56	Reference
	>3 and <=6 month	282	32	14.24	20.51	0.0075
	>6 and <=12month	197	14	9.95	8.97	0.6977
	>12	205	28	10.35	17.95	0.0009
Duration of sex without contraception with father of baby before current pregnancy <= 3 months	No	684	74	34.55	47.44	Reference
	Yes	1296	82	65.45	52.56	0.0013

Duration of sex without contraception with father of baby before current pregnancy >12 months	No	1775	128	89.65	82.05	Reference
	Yes	205	28	10.35	17.95	0.0039
History of infertility defined as >=12 mths of regular intercourse without contraception and conception has not occurred or if partner is known to be sterile. Includes unknown if participant does not wish to answer question.	No	1700	120	85.69	76.92	Reference
	Yes	283	36	14.26	23.08	0.0033
History of infertility defined as >=12 mths of regular intercourse without contraception and conception has not occurred or if partner is known to be sterile. Binary response 'unknown' combined with NO	No hx infertility	1701	120	85.74	76.92	Reference
	Yes, hx infertility	283	36	14.26	23.08	0.0033
Hormonal treatment to assist conception of current pregnancy	No (either NO fertility tx or NO hormonal tx)	1848	139	93.15	89.10	Reference
	Clomiphene	17	2	0.86	1.28	0.5523
	Other	62	6	3.13	3.85	0.5637
Hormonal treatment, other than clomiphene, to assist conception of current pregnancy	No	1927	147	97.13	94.23	Reference
	Yes	57	9	2.87	5.77	0.0485
Had LLETZ treatment for CIN	No (includes unknown if CIN (n=6) or NO CIN/abnormal smear n=5083)	1909	142	96.22	91.03	Reference
	Yes	75	14	3.78	8.97	0.0025
Had either LLETZ, laser or cryotherapy treatment for CIN/abnormal smear	No	1846	138	93.04	88.46	Reference
	Yes	138	18	6.96	11.54	0.0362
Number of LLETZ where the last colposcopy was 7-12 months before conception	Yes	1976	153	99.60	98.08	Reference
	unknown if CIN (n=6) or NO CIN/abnormal smear (n=5083)=0	8	3	0.40	1.92	0.0207

Had last LLETZ Rx 7-12 months before conception current pregnancy	No	1976	153	99.60	98.08	Reference
	Yes, CIN or abnormal smear leading to LLETZ 7-12 months of conception	8	3	0.40	1.92	0.0207
Had last LLETZ Rx>12 months before conception current pregnancy	No	1921	146	96.82	93.59	Reference
	Yes, CIN or abnormal smear leading to LLETZ>12 months before conception	63	10	3.18	6.41	0.0359
Number of LLETZ treatments	No Rx or Unknown	1909	142	96.22	91.03	Reference
	1 Rx	71	12	3.58	7.69	0.0113
	>=2 Rx	4	2	0.20	1.28	0.0286
On treatment for PCOS preceding/at conception; other Rx includes ovarian drilling	No (NO PCOS or PCOS + no Rx)	1940	149	97.78	95.51	Reference
	"Yes, fertility drugs"	20	1	1.01	0.64	0.6763
	"Yes, metformin"	7	3	0.35	1.92	0.0134
	"Yes, fertility drugs and metformin"	14	2	0.71	1.28	0.4146
	Other Rx	3	1	0.15	0.64	0.2049
On metformin for PCOS prior to/at conception	No	1963	151	98.94	96.79	Reference
	Yes	21	5	1.06	3.21	0.0252
Self reported hypertension (on more than 1 occasion) while on oral contraception	No	1963	150	98.94	96.15	Reference
	Yes	21	6	1.06	3.85	0.0051
Mild hypertension prior to pregnancy but never on antihypertensive medication (self reported) or at antenatal booking systolic BP 140-159 or diastolic 90-99 mmHg;	No	1966	151	99.09	96.79	Reference
	Yes	18	5	0.91	3.21	0.0121

Some of these women will be white coat hypertension						
Self reported history of depression, may or may not have taken anti-depressants	No	125	9	70.62	42.86	Reference
	Yes	52	12	29.38	57.14	0.0134
Number of sisters who had had a miscarriage	no sister	1801	132	90.78	84.62	Reference
	1 sister had any miscarriage	171	24	8.62	15.38	0.0059
	>=2 sisters had any miscarriage	12	0	0.60	0.00	0.9754
Any sister who had a history of miscarriage	No (also includes no sisters who've had pregnancies or sisters with unknown obstetric history)	1801	132	90.78	84.62	Reference
	Yes	183	24	9.22	15.38	0.0133
Participant's biological mother had a history of pregnancy induced hypertension (gestational hypertension or preeclampsia, but unable to be confident which condition)	No	1752	132	88.31	84.62	Reference
	participant's mother had PIH 1x	121	9	6.10	5.77	0.9713
	participant's mother had PIH >=2x	32	7	1.61	4.49	0.0125
Participant's mother had a history of pregnancy induced hypertension	No	1752	132	88.31	84.62	Reference
	participant's mother had PIH 1x	121	9	6.10	5.77	0.9713
	participant's mother had PIH >=2x	32	7	1.61	4.49	0.0125
Participant's mother had a history of pregnancy induced hypertension	No/unknown	1831	140	92.29	89.74	Reference
	participant's mother had PIH 1x	121	9	6.10	5.77	0.9383
	participant's mother had PIH >=2x	32	7	1.61	4.49	0.0137

Participant's mother had any history of recurrent pregnancy induced hypertension	NO, participant's mother has no history of recurrent PIH	1952	149	98.39	95.51	Reference
	Yes, participant's mother had recurrent PIH	32	7	1.61	4.49	0.0134
Participant's mother had any history of pregnancy induced hypertension (GH or preeclampsia, but unsure which condition), gestational hypertension or preeclampsia	No Participant's mother did not have PIH or GH or PET	1686	123	84.98	78.85	Reference
	Yes, Participant's mother had PIH or GH or PET	298	33	15.02	21.15	0.0426
Any family history of recurrent PIH	No	1950	149	98.29	95.51	Reference
	Yes	34	7	1.71	4.49	0.0193
Family history of PIH (i.e. GH or preeclampsia, but unsure which condition), gestational hypertension or preeclampsia; family members are participant's mother and/or sisters	No Family History of PIH/PET/GH	1626	115	81.96	73.72	Reference
	Yes, Family History of PIH/PET/GH	358	41	18.04	26.28	0.0116
Participant's biological mother had preeclampsia (PET)	No	1754	129	88.41	82.69	Reference
	participant's mother had PET 1x	116	13	5.85	8.33	0.1692
	participant's mother had PET >=2x	29	7	1.46	4.49	0.0058
Participant's mother had a history of PET (3 groups)	No	1754	129	88.41	82.69	Reference
	participant's mother had PET 1x	116	13	5.85	8.33	0.1692
	participant's mother had PET >=2x	29	7	1.46	4.49	0.0058
Participant's mother had any history of PET	No, participant's mother never had PET	1839	136	92.69	87.18	Reference
	Yes, participant's mother had PET	145	20	7.31	12.82	0.0143

Participant's mother had any history of recurrent PET	No, participant's mother never had recurrent PET	1955	149	98.54	95.51	Reference
	Yes, participant's mother had recurrent PET	29	7	1.46	4.49	0.0073
Family history of PET, i.e. participant's mother or sister had had PET	No	1803	131	90.88	83.97	Reference
	Yes	181	25	9.12	16.03	0.0056
Family history of recurrent PET, i.e. participant's mother or sister had had recurrent PET	No	1946	148	98.08	94.87	Reference
	Yes	38	8	1.92	5.13	0.0106
Family history (mother or sisters) of GH or PET	No	1706	122	85.99	78.21	Reference
	Yes	278	34	14.01	21.79	0.0087
Family history (mother or sisters) of recurrent GH or recurrent PET	No Strong Family History of rec_GH or PET or Ecl	1929	145	97.23	92.95	Reference
	Yes, Strong Family' History of rec_GH or PET or Ecl	55	11	2.77	7.05	0.0041
Participant's biological mother had LBW baby	No	1680	116	84.68	74.36	Reference
	participant's mother had LBW baby 1x	189	25	9.53	16.03	0.0054
	participant's mother had LBW baby >=2x	58	8	2.92	5.13	0.0754
	participant's mother had LBW baby 1x	189	25	9.53	16.03	0.0072
	participant's mother had LBW baby >=2x	58	8	2.92	5.13	0.0861
Participant's mother had any history of LBW baby	No, participant's mother never had LBW baby	1737	123	87.55	78.85	Reference
	Yes, participant's mother had LBW baby	247	33	12.45	21.15	0.0022
Number of sisters who had a LBW baby	No sister had LBW	1927	145	97.13	92.95	Reference

	baby					
	1 Sister had LBW baby	54	10	2.72	6.41	0.0112
	>= 2 Sisters had LBW baby	3	1	0.15	0.64	0.1987
Any sister had a history of LBW baby	No	1927	145	97.13	92.95	Reference
	Yes	57	11	2.87	7.05	0.0057
Family history of LBW baby, i.e. participant's mother or sister had had LBW baby	No	1695	117	85.43	75.00	Reference
	Yes	289	39	14.57	25.00	0.0006
Family history of recurrent LBW baby, i.e. participant's mother or sister had had recurrent LBW baby	No	1912	145	96.37	92.95	Reference
	Yes	72	11	3.63	7.05	0.0365
Strong family history (2 or more members of immediate family -mother or sisters) of LBW baby	No	1968	150	99.19	96.15	Reference
	Yes	16	6	0.81	3.85	0.0010
Strong family history (2 or more members of immediate family -mother or sisters) of recurrent LBW baby	No Strong Family History of Recurrent rec_LBW baby	1984	156	100.00	100.00	Reference
	Yes, Strong Family' History of Recurrent rec_LBW baby	0	0	0.00	0.00	0.0010
participant's mother delivered a baby preterm (all PTB-spontaneous or iatrogenic)	No	1697	120	85.53	76.92	Reference
	participant's mother had a PTB (spontaneous or iatrogenic) 1x	187	20	9.43	12.82	0.1027
	participant's mother had a PTB (spontaneous or iatrogenic) >=2x	47	10	2.37	6.41	0.0023
participant's mother delivered a baby preterm (all PTB-	No	1697	120	85.53	76.92	Reference

spontaneous or iatrogenic)	participant's mother had a PTB (spontaneous or iatrogenic) 1x	187	20	9.43	12.82	0.1027
	participant's mother had a PTB (spontaneous or iatrogenic) >=2x	47	10	2.37	6.41	0.0023
Participant's mother had any PTB (spontaneous or iatrogenic)	No, participant's mother never had a PTB (spont or iatrogenic)	1750	126	88.21	80.77	Reference
	Yes, participant's mother had any PTB (spont or iatrogenic)	234	30	11.79	19.23	0.0072
Participant's mother had history of recurrent PTB (all-spontaneous or iatrogenic)	No, participant's mother never had recurrent PTB (spont or iatrogenic)	1937	146	97.63	93.59	Reference
	Yes, participant's mother had recurrent PTB (spont or iatrogenic)	47	10	2.37	6.41	0.0038
Number of sisters with a history of PTB (all-spontaneous or iatrogenic)	No sister had a PTB (spont or iatrogenic)	1915	145	96.52	92.95	Reference
	1 Sister had PTB (spont or iatrogenic)	66	9	3.33	5.77	0.1076
	>= 2 Sisters had PTB (spont or iatrogenic)	3	2	0.15	1.28	0.0177
Any sister with a history of PTB (all-spontaneous or iatrogenic)	No	1915	145	96.52	92.95	Reference
	Yes	69	11	3.48	7.05	0.0267
More than one sister with a history of PTB (all-spontaneous or iatrogenic)	No	1981	154	99.85	98.72	Reference
	Yes	3	2	0.15	1.28	0.0191
FH all PTB (spontaneous or iatrogenic) i.e. participant's mother or sister(s) delivered >=2 babies preterm (spont or iatrogenic)	No	1692	119	85.28	76.28	Reference
	Yes	292	37	14.72	23.72	0.0030

FH recurrent all PTB (spontaneous or iatrogenic) i.e. participant's mother or sister(s) delivered >=2 babies preterm	No	1926	144	97.08	92.31	Reference
	Yes	58	12	2.92	7.69	0.0020
Strong FH all PTB i.e. >=2 family members (participant's mother or sisters) delivered a baby preterm (all PTB-spontaneous or iatrogenic)	No	1972	152	99.40	97.44	Reference
	Yes	12	4	0.60	2.56	0.0121
participant's mother had spontaneous PTB	No	1749	127	88.16	81.41	Reference
	participant's mother had 1x spontaneous PTB	137	20	6.91	12.82	0.0065
	participant's mother had >=2x spontaneous PTB	43	3	2.17	1.92	0.9472
participant's mother had spontaneous PTB, unknown and NA categories fused	No	1749	127	88.16	81.41	Reference
	participant's mother had 1x spontaneous PTB	137	20	6.91	12.82	0.0065
	participant's mother had >=2x spontaneous PTB	43	3	2.17	1.92	0.9472
Participant's mother had any spontaneous PTB	No, participant's mother never had any spont PTB	1804	133	90.93	85.26	Reference
	Yes, participant's mother had any spontaneous PTB	180	23	9.07	14.74	0.0214
Any sister with a history of spont PTB	No	1930	147	97.28	94.23	Reference
	Yes	54	9	2.72	5.77	0.0343
FH spontaneous PTB i.e. participant's mother or sister(s) had a spontaneous PTB	No	1757	125	88.56	80.13	Reference
	Yes	227	31	11.44	19.87	0.0022
participant's mother had GDM	No	1877	140	94.61	89.74	Reference
	participant's mother had 1x GDM	34	7	1.71	4.49	0.0167

	participant's mother had >=2x GDM	11	2	0.55	1.28	0.2495
Participant's mother had any GDM	No, participant's mother never had any GDM	1939	147	97.73	94.23	Reference
	Yes, participant's mother had any GDM	45	9	2.27	5.77	0.0097
Number of sisters with a history of recurrent GDM	No sister had recurrent GDM	1983	154	99.95	98.72	Reference
	1 Sister had recurrent GDM	1	2	0.05	1.28	0.0081
Any sister with a history of GDM	No sister had GDM	1968	152	99.19	97.44	Reference
	Any Sister had GDM	16	4	0.81	2.56	0.0377
Any sister with a history of recurrent GDM	No sister had recurrent GDM	1983	154	99.95	98.72	Reference
	Any sister had recurrent GDM	1	2	0.05	1.28	0.0081
More than 1 sister had a history of GDM	No	1984	156	100.00	100.00	Reference
	Yes	0	0	0.00	0.00	0.0081
FH GDM i.e. participant's mother or sister(s) had GDM	No	1923	143	96.93	91.67	Reference
	Yes	61	13	3.07	8.33	0.0009
FH recurrent GDM i.e. participant's mother or sister(s) had recurrent GDM	No	1972	152	99.40	97.44	Reference
	Yes	12	4	0.60	2.56	0.0121
Strong FH GDM i.e. >=2 family members (participant's mother or sisters) had GDM	No	1984	156	100.00	100.00	Reference
	Yes	0	0	0.00	0.00	0.0121
participant's mother has had a CVA	No	1906	145	96.07	92.95	Reference
	participant's mother has had a CVA	33	7	1.66	4.49	0.0158

participant's mother has any history of type 2 diabetes or diabetes type not specified	No participant's mother does not have type 2 diabetes or diabetes type not specified	1896	143	95.56	91.67	Reference
	participant's mother has type 2 diabetes or diabetes type not specified	88	13	4.44	8.33	0.0299
Participant's mother has one or more of type 2 diabetes, chronic hypertension, CVA and IHD	No participant's mother does not have any of type 2 diabetes, CH, CVA or IHD	1504	107	75.81	68.59	Reference
	participant's mother has one or more of type 2 diabetes, CH, CVA or IHD	480	49	24.19	31.41	0.0452
Any vaginal bleeding in pregnancy before 15w SCOPE visit	NO	1606	113	80.95	72.44	Reference
	YES	378	43	19.05	27.56	0.0106
number of vaginal bleeds before 15w SCOPE visit, categorised	No bleeding	1606	113	80.95	72.44	Reference
	1x bleeding	298	31	15.02	19.87	0.0656
	>=2x bleeding	80	12	4.03	7.69	0.0197
number of vaginal bleeds before 15w SCOPE visit, spotting severity (categorised)	No bleeding	1687	122	85.03	78.21	Reference
	1x bleeding	248	29	12.50	18.59	0.0271
	>=2x bleeding	49	5	2.47	3.21	0.4720
number of vaginal bleeds before 15w SCOPE visit, mod-heavy severity (categorised)	No bleeding	1951	149	98.34	95.51	Reference
	1x bleeding	29	6	1.46	3.85	0.0290
	>=2x bleeding	4	1	0.20	0.64	0.2902

number of vaginal bleeds commencing >12w, categorised	No bleeding	1933	144	97.43	92.31	Reference
	1x bleeding	49	10	2.47	6.41	0.0048
	>=2x bleeding	2	2	0.10	1.28	0.0097
number of vaginal bleeds 5-9 days before 15w SCOPE visit, categorised	No bleeding	1949	150	98.24	96.15	Reference
	1x bleeding	32	6	1.61	3.85	0.0493
	>=2x bleeding	3	0	0.15	0.00	0.9812
Gestational age when 1st vaginal bleed occurred before 15w SCOPE visit, categorised	No Bleeding	1606	113	80.95	72.44	Reference
	1st bleed between 1-6wk	192	21	9.68	13.46	0.0772
	1st bleed between 7-12wk	153	15	7.71	9.62	0.2486
	1st bleed after 12wk	33	7	1.66	4.49	0.0098
Gestational age when 2nd vaginal bleed occurred before 15w SCOPE visit, categorised	No Bleeding	1902	143	95.87	91.67	Reference
	2nd bleed between 1-6wk	14	2	0.71	1.28	0.3989
	2nd bleed between 7-12wk	56	7	2.82	4.49	0.2152
	2nd bleed after 12wk	12	4	0.60	2.56	0.0107
Gestational age when last vaginal bleed occurred before 15w SCOPE visit, categorised	No Bleeding	1606	113	80.95	72.44	Reference
	last bleed between 1-6wk	146	15	7.36	9.62	0.1888
	last bleed between 7-12wk	181	16	9.12	10.26	0.4124
	last bleed after 12wk	51	12	2.57	7.69	0.0003
Number of episodes of light vaginal bleeding at or before 6w gestation (categorised)	No bleeding	1946	149	98.08	95.51	Reference
	1x bleeding	38	7	1.92	4.49	0.0366

Number of episodes of light vaginal bleeding at 7-12w gestation (categorised)	No bleeding	1942	151	97.88	96.79	Reference
	1x bleeding	39	3	1.97	1.92	0.9858
	>=2x bleeding	3	2	0.15	1.28	0.0191
Number of episodes of mod-heavy vaginal bleeding at 7-12w gestation (categorised)	No bleeding	1970	152	99.29	97.44	Reference
	1x bleeding	13	4	0.66	2.56	0.0167
	>=2x bleeding	1	0	0.05	0.00	0.9836
Number of episodes of spotting vaginal bleeding after 12w gestation (categorised)	No bleeding	1945	147	98.03	94.23	Reference
	1x bleeding	37	8	1.86	5.13	0.0085
	>=2x bleeding	2	1	0.10	0.64	0.1238
Total number of spotting or light vaginal bleeding before 15w SCOPE visit, categorised	No bleeding	1622	116	81.75	74.36	Reference
	1x bleeding	293	30	14.77	19.23	0.0942
	>=2x bleeding	69	10	3.48	6.41	0.0446
Total number of vag bleeding after 6 weeks gestation, categorised	No bleeding	1752	128	88.31	82.05	Reference
	1x bleeding	190	21	9.58	13.46	0.0944
	>=2x bleeding	42	7	2.12	4.49	0.0487
Two or more episodes of vaginal bleeding before 15w SCOPE visit	No	1904	144	95.97	92.31	Reference
	Yes	80	12	4.03	7.69	0.0331
Any spotting or light vaginal bleeding before 15w SCOPE visit	No	1622	116	81.75	74.36	Reference
	Yes	362	40	18.25	25.64	0.0237
Any mod-heavy vaginal bleeding before 15w SCOPE visit	No	1951	149	98.34	95.51	Reference
	Yes	33	7	1.66	4.49	0.0162

Any vaginal bleeding after 12w gestation	No	1933	144	97.43	92.31	Reference
	Yes	51	12	2.57	7.69	0.0005
Any mod or heavy vaginal bleeding continuing for at least 5 days before 15w SCOPE visit	No	1981	154	99.85	98.72	Reference
	Yes	3	2	0.15	1.28	0.0191
Any light vag bleeding at or before 6w gestation	No	1946	149	98.08	95.51	Reference
	Yes	38	7	1.92	4.49	0.0366
Any mod-heavy vag bleeding at 7-12w gestation	No	1970	152	99.29	97.44	Reference
	Yes	14	4	0.71	2.56	0.0224
Any spotting vag bleeding after 12w gestation	No	1945	147	98.03	94.23	Reference
	Yes	39	9	1.97	5.77	0.0033
Any mod-heavy vag bleeding after 12w gestation	No	1982	154	99.90	98.72	Reference
	Yes	2	2	0.10	1.28	0.0109
Bleeding gums when brushing teeth prior to pregnancy	NO	925	66	75.39	65.35	Reference
	YES	302	35	24.61	34.65	0.0270
Bleeding gums when brushing teeth at time of 15w SCOPE visit	NO	732	50	59.76	49.50	Reference
	YES	493	51	40.24	50.50	0.0453
Number of hospital admissions due to hyperemesis before 15w SCOPE visit, categorised	No admission due to hyperemesis (includes women with no hospital admission for any reason)	1949	149	98.24	95.51	Reference
	1x admission due to Hyperemesis	30	4	1.51	2.56	0.3021
	>=2x admission due to	5	3	0.25	1.92	0.0051

Hyperemesis						
Number of hospital admissions due to vaginal bleeding before 15w SCOPE visit, categorised	No admission due to vaginal bleeding (includes women with no hospital admission for any reason)	1959	151	98.74	96.79	Reference
	1x admission due to Vaginal Bleeding	23	3	1.16	1.92	0.3959
	>=2x admission due to Vaginal Bleeding	2	2	0.10	1.28	0.0107
Number of hospital admissions due to trauma before 15w SCOPE visit, categorised	No admission due to trauma (includes women with no hospital admission for any reason)	1983	154	99.95	98.72	Reference
	1x admission due to Trauma	1	2	0.05	1.28	0.0081
Number of hospital admissions due to any other reasons before 15w SCOPE visit, categorised	No admission due to any other reasons (includes women with no hospital admission for any reason)	1980	154	99.80	98.72	Reference
	1x admission due to any other reasons	4	2	0.20	1.28	0.0325
Any hospital admissions due to hyperemesis before 15w SCOPE visit	No admission due to hyperemesis (includes women with no hospital admission for any reason)	1949	149	98.24	95.51	Reference
	Admission due to Hyperemesis	35	7	1.76	4.49	0.0229
Any hospital admissions due to trauma before 15w SCOPE visit	No admission due to trauma (includes women	1983	154	99.95	98.72	Reference

	with no hospital admission for any reason)					
	Admission due to Trauma	1	2	0.05	1.28	0.0081
Any hospital admissions due to any other reasons before the 15w SCOPE visit	No admission due to any other reasons (includes women with no hospital admission for any reason)	1980	154	99.80	98.72	Reference
	Admission due to any other reasons	4	2	0.20	1.28	0.0325
Did NOT consume oily fish (which is high in omega 3 long chain fatty acids) in the month prior to conception	Consumed oily fish	1427	100	71.93	64.10	Reference
	No consumption of oily fish	557	56	28.07	35.90	0.0383
	3-6x per week	262	34	13.21	21.79	0.0003
	1-2x per week	269	27	13.56	17.31	0.0227
	1-3x per month	96	12	4.84	7.69	0.0219
	Never	67	7	3.38	4.49	0.1668
Frequency consumed fruit in the month prior to conception, compressed categories (4 gps)	>=1x per day	1290	76	65.02	48.72	Reference
	3-6x per week	262	34	13.21	21.79	0.0003
	1-2x per week	269	27	13.56	17.31	0.0227
	1-3x per month or less	163	19	8.22	12.18	0.0114
high (>=3 times per day) fruit consumption in the month prior to conception	No	1472	130	74.19	83.33	Reference
	Consumed fruit >=3x per day	512	26	25.81	16.67	0.0122
Frequency consumed fruit in pregnancy prior to 15w SCOPE visit, compressed categories (5 gps)	>=1x per day	1537	100	77.47	64.10	Reference

Frequency consumed fruit in pregnancy prior to 15w SCOPE visit, compressed categories (4 gps)	>=1x per day	1537	100	77.47	64.10	Reference
	3-6x per week	253	32	12.75	20.51	0.0019
	1-2x per week	121	17	6.10	10.90	0.0058
	1-3x per month or less	73	7	3.68	4.49	0.3429
high (>=3 times per day) fruit consumption in pregnancy prior to 15w SCOPE visit	No	1105	101	55.70	64.74	Reference
	Consumed fruit >=3x per day	879	55	44.30	35.26	0.0290
	1-3/mth	486	39	39.84	38.61	0.9052
	1 or 2/wk	340	22	27.87	21.78	0.4112
	3 or 4/wk	59	11	4.84	10.89	0.0358
	5 or 6/wk	8	2	0.66	1.98	0.1759
	1-2/day	10	1	0.82	0.99	0.8598
	3-4/day	3	0	0.25	0.00	0.9811
Frequency consumed burger in the month prior to conception	Never	314	26	25.74	25.74	Reference
	1-3x per month	486	39	39.84	38.61	0.9052
	1-2x per week	340	22	27.87	21.78	0.4112
	3-6x per week	67	13	5.49	12.87	0.0198
	>=1x per day	13	1	1.07	0.99	0.9445
Frequency consumed burger in pregnancy prior to 15w SCOPE visit	Never	369	29	30.20	28.71	Reference
	1-3x per month	503	45	41.16	44.55	0.6010
	1-2x per week	303	18	24.80	17.82	0.3665
	3-6x per week	41	9	3.36	8.91	0.0134

	>=1x per day	6	0	0.49	0.00	0.9825
Frequency consumed fried chicken in pregnancy prior to 15w SCOPE visit	Never	873	69	71.50	68.32	Reference
	1-3x per month	288	21	23.59	20.79	0.7550
	1-2x per week	54	9	4.42	8.91	0.0503
	3-6x per week	4	2	0.33	1.98	0.0350
	>=1x per day	2	0	0.16	0.00	0.9846
Frequency consumed pizza in the month prior to conception	Never	231	32	18.92	31.68	Reference
	1-3x per month	771	53	63.14	52.48	0.0030
	1-2x per week	213	16	17.44	15.84	0.0563
	3-6x per week	6	0	0.49	0.00	0.9818
Frequency consumed hot chips/french fries in the month prior to conception	Never	228	9	18.72	8.91	Reference
	1-3x per month	550	41	45.16	40.59	0.0911
	1-2x per week	349	36	28.65	35.64	0.0120
	3-6x per week	76	11	6.24	10.89	0.0055
	>=1x per day	15	4	1.23	3.96	0.0037
Frequency consumed hot chips/french fries in pregnancy prior to 15w SCOPE visit	Never	206	11	16.87	10.89	Reference
	1-3x per month	534	40	43.73	39.60	0.3338
	1-2x per week	370	40	30.30	39.60	0.0447
	3-6x per week	98	7	8.03	6.93	0.5597
	>=1x per day	13	3	1.06	2.97	0.0396
Any folate intake in 1st trimester	No	107	17	5.39	10.90	Reference

	Yes	1877	139	94.61	89.10	0.0056
Folate dose by <=800µg and >800µg per day in 1st trimester	Not taking folate	107	17	5.39	10.90	Reference
	Yes <=800	1677	134	84.53	85.90	0.0128
	Yes >800	200	5	10.08	3.21	0.0004
Folate dose by <=800µg and >800µg per day at 15w SCOPE visit	Not taking folate	489	44	24.65	28.21	Reference
	Yes <=800	1382	109	69.66	69.87	0.4792
	Yes >800	113	3	5.70	1.92	0.0439
number of cigarettes per day in the 3 months pre-pregnancy (categories)	No	1549	108	78.07	69.23	Reference
	1-5 cigs	134	10	6.75	6.41	0.8427
	6-10 cigs	120	15	6.05	9.62	0.0451
	>10 cigs	181	23	9.12	14.74	0.0134
number of cigarettes per day in the 1st trimester (categories)	No	1588	111	80.04	71.15	Reference
	1-5 cigs	168	15	8.47	9.62	0.3934
	6-10 cigs	119	15	6.00	9.62	0.0428
	>10 cigs	109	15	5.49	9.62	0.0205
Smoked during the 1st trimester	No	1588	111	80.04	71.15	Reference
	Yes	396	45	19.96	28.85	0.0088
Consumed/inhaled/injected other recreational drugs or binge alcohol consumption (>=6 units/session) in the 1st trimester	No	1734	126	87.40	80.77	Reference
	Yes	250	30	12.60	19.23	0.0191
High binge alcohol consumption in 1st trimester (defined as >1 binge per week)	No	1975	153	99.55	98.08	Reference
	Yes	9	3	0.45	1.92	0.0299

Gestation binge alcohol ceased in pregnancy (categories)	No binge Alcohol	1817	141	91.58	90.38	Reference
	1-6 weeks	144	9	7.26	5.77	0.5416
	>6 weeks	23	6	1.16	3.85	0.0094
Pulse <=60/min at 15w SCOPE visit	No	1798	149	90.72	95.51	Reference
	Yes	184	7	9.28	4.49	0.0483
Short Height (<161 cm)	>=161 cm	1531	106	77.17	67.95	Reference
	<161 cm	453	50	22.83	32.05	0.0095
If you do paid work, what activity best describes the main activities at work evaluated at 15w SCOPE visit	"Administrative, sitting activities"	558	51	32.69	40.48	Reference
	Sitting and some walking	541	27	31.69	21.43	0.0137
	Standing	38	3	2.23	2.38	0.8125
	Standing/walking	360	28	21.09	22.22	0.5097
	Standing/walking/intermittent exercise	191	16	11.19	12.70	0.7704
	Regular exercise	19	1	1.11	0.79	0.5943
PSS: On top of things evaluated at 15w SCOPE visit	Never	267	22	13.55	14.10	Reference
	Almost never	844	56	42.82	35.90	0.4070
	Sometimes	663	58	33.64	37.18	0.8183
	Fairly often	178	15	9.03	9.62	0.9486
	Very often	19	5	0.96	3.21	0.0345
In last month, number of episodes of waking during a night's sleep, evaluated at 15w SCOPE visit	Don’t wake up	127	18	6.44	11.54	Reference
	Once a night	644	40	32.67	25.64	0.0060
	2-3 times	1004	70	50.94	44.87	0.0114

	>=4 times	196	28	9.94	17.95	0.9805
Snored most nights, evaluated at 15w SCOPE visit	No	1338	90	67.44	57.69	Reference
	Yes	215	23	10.84	14.74	0.0582
Engaged in vigorous exercise (which made the woman breathe harder or puff or pant) in the last month, evaluated at 15w SCOPE visit	Never	1160	98	58.85	62.82	Reference
	Once a week	503	37	25.52	23.72	0.4892
	2-3 / wk	236	11	11.97	7.05	0.0680
	4-6 x /wk	54	5	2.74	3.21	0.8483
	Daily	16	3	0.81	1.92	0.2114
	More than once a day	2	2	0.10	1.28	0.0140
Engaged in less vigorous exercise (the woman did not breathe harder or puff or pant) in the last month, evaluated at 15w SCOPE visit	Never	453	50	22.98	32.05	Reference
	Once a week	557	40	28.26	25.64	0.0522
	2-3 / wk	621	40	31.51	25.64	0.0148
	4-6 x /wk	171	8	8.68	5.13	0.0282
	Daily	150	17	7.61	10.90	0.9288
	More than once a day	19	1	0.96	0.64	0.4750
Engaged in less vigorous exercise (the woman did not breathe harder or puff or pant) in the last month, evaluated at 15w SCOPE visit, compressed into 3 categories	Never	453	50	22.98	32.05	Reference
	1-3 times/ week	1178	80	59.77	51.28	0.0100
	>=4 times / wk	340	26	17.25	16.67	0.1457
Extreme exercise in pregnancy (undertook vigorous exercise at least once a day) evaluated at 15w SCOPE visit	No	1953	151	99.09	96.79	Reference
	YES	18	5	0.91	3.21	0.0126
Watching >=5h TV per day evaluated at 15w SCOPE visit	NO	1748	129	88.60	82.69	Reference

	YES	225	27	11.40	17.31	0.0294
Number of hours spent using a computer per day in the month prior to the 15±1 week SCOPE interview	None	194	26	9.83	16.67	Reference
	<2h	557	35	28.23	22.44	0.0054
	2-4h	290	21	14.70	13.46	0.0454
	5-6h	329	26	16.68	16.67	0.0702
	>6h	603	48	30.56	30.77	0.0427
Any computer usage in last month evaluated at 15w SCOPE visit	NO	194	26	9.83	16.67	Reference
	YES	1779	130	90.17	83.33	0.0077
Never used a computer in the last month evaluated at 15w SCOPE visit	NO	1779	130	90.17	83.33	Reference
	YES	194	26	9.83	16.67	0.0077
Behaviour Responses to Pregnancy': Put parts of life on hold since pregnant evaluated at 15w SCOPE visit	Not at all	510	51	25.88	32.69	Reference
	Rarely	586	33	29.73	21.15	0.0131
	Some days	572	45	29.02	28.85	0.2610
	Most days	235	18	11.92	11.54	0.3500
	Every day	68	9	3.45	5.77	0.4653
Behaviour Responses to Pregnancy': Pushed myself until I cannot push anymore evaluated at 15w SCOPE visit	Not at all	1055	73	53.50	46.79	Reference
	Rarely	605	61	30.68	39.10	0.0373
	Some days	257	19	13.03	12.18	0.8040
	Most days	45	3	2.28	1.92	0.9512
	Everyday	10	0	0.51	0.00	0.9776
Behaviour Responses to Pregnancy': Carried on until	Not at all	601	35	30.48	22.44	Reference

body unable to cope any longer evaluated at 15w SCOPE visit	Rarely	622	59	31.54	37.82	0.0272
	Some days	465	37	23.58	23.72	0.2004
	Most days	235	19	11.92	12.18	0.2663
	Everyday	49	6	2.48	3.85	0.1109
I have felt better than ever in pregnancy evaluated at 15w SCOPE visit. This is not part of the questionnaire and is an additional question	Not at all	454	25	23.05	16.03	Reference
	Rarely	567	42	28.78	26.92	0.2547
	Some days	657	59	33.35	37.82	0.0471
	Most days	255	29	12.94	18.59	0.0106
	Every day	37	1	1.88	0.64	0.4913
I have felt better than ever in pregnancy evaluated at 15w SCOPE visit, compressed categories	Not at all	454	25	23.05	16.03	Reference
	Sometimes	1224	101	62.13	64.74	0.0787
	At least most days	292	30	14.82	19.23	0.0265
Score for 'Short form State-Trait Anxiety Inventory evaluated at 15w SCOPE visit >90th centile	NO	1838	137	93.25	88.39	Reference
	YES	133	18	6.75	11.61	0.0251
Depression Scale: Looked forward to things with enjoyment evaluated at 15w SCOPE visit	As much as always	1340	101	67.92	64.74	Reference
	Not quite so much	519	45	26.31	28.85	0.4527
	Definitely not so much	102	5	5.17	3.21	0.3595
	Not at all	12	5	0.61	3.21	0.0016
Support people around to provide emotional support evaluated at 15w SCOPE visit	All the time	1339	96	67.94	61.54	Reference
	Most of the time	465	48	23.59	30.77	0.0486
	Sometimes	136	10	6.90	6.41	0.9415

	Seldom	27	2	1.37	1.28	0.9648
	Never	4	0	0.20	0.00	0.9784
Support people around to provide emotional support evaluated at 15w SCOPE visit, compressed categories	All the time	1339	96	67.94	61.54	Reference
	Most of the time	465	48	23.59	30.77	0.0486
	Sometimes	136	10	6.90	6.41	0.9415
	Seldom /Never	31	2	1.57	1.28	0.8862
Social support (listening ears and practical support scores added) categorised	2	1046	84	53.07	53.85	Reference
	3	349	16	17.71	10.26	0.0450
	4	349	40	17.71	25.64	0.0780
	5	115	10	5.83	6.41	0.8195
	>5	112	6	5.68	3.85	0.3510
Current work situation at 20w SCOPE visit	Full time work	1315	92	67.82	61.74	Reference
	Part time work	316	24	16.30	16.11	0.7297
	Student	53	2	2.73	1.34	0.3964
	Homemaker	87	6	4.49	4.03	0.9737
	Unemployed	150	22	7.74	14.77	0.0034
	Sickness beneficiary	6	2	0.31	1.34	0.0580
	Other (e.g.) voluntary work	12	1	0.62	0.67	0.8673
Unemployed or Sickness Beneficiary at 20w SCOPE visit	No	1783	125	91.95	83.89	Reference
	Yes	156	24	8.05	16.11	0.0010
Work status at 20w SCOPE visit (3 groups)	No paid work	308	33	15.88	22.15	Reference

	Part time work	316	24	16.30	16.11	0.2190
	Full time work	1315	92	67.82	61.74	0.0449
Any hyperemesis between SCOPE 15w and 20w visit	NO	1886	141	97.22	93.38	Reference
	YES	54	10	2.78	6.62	0.0107
Hyperemesis continuing at SCOPE 20w visit	No Hyperemesis	1816	136	93.61	90.07	Reference
	Hyperemesis at or before 15w visit, no vomiting afterwards	70	5	3.61	3.31	0.9200
	Hyperemesis at or before 15w vst, ongoing between 15w and 20w visit	16	5	0.82	3.31	0.0060
	New onset vomiting between 15w and 20w visit	38	5	1.96	3.31	0.2442
Two episodes of vaginal bleeding between 15w and 20w SCOPE visits	No	1937	149	99.85	98.68	Reference
	Yes	3	2	0.15	1.32	0.0185
Any vaginal bleeding recorded at either 15w or 20w SCOPE visit	No	1579	109	79.59	71.24	Reference
	Yes	405	44	20.41	28.76	0.0154
Any vaginal bleeding between 13 weeks' gestation and 20w SCOPE visit (could be recorded at either 15w or 20w visit)	No	1844	135	95.05	89.40	Reference
	Yes	96	16	4.95	10.60	0.0038
Any hospital admissions due to trauma between the 15w and 20w SCOPE visits	No admission due to trauma (includes women with no hospital admission for any reason)	1937	149	99.85	98.68	Reference
	Admission due to Trauma	3	2	0.15	1.32	0.0185

Frequency consumed oily fish (which is high in omega 3 long chain fatty acids) between 15w and 20w SCOPE visits	Never	793	77	40.96	51.33	Reference
	1-3/mth	493	38	25.46	25.33	0.2632
	1 or 2/wk	538	28	27.79	18.67	0.0062
	3 or 4/wk	88	6	4.55	4.00	0.4201
	5 or 6/wk	17	0	0.88	0.00	0.9805
	1-2/day	5	1	0.26	0.67	0.5120
	3-4/day	2	0	0.10	0.00	0.9933
Did NOT consume oily fish between the 15w and 20w SCOPE visits	No	1143	73	59.04	48.67	Reference
	Yes	793	77	40.96	51.33	0.0136
	1-3/mth	37	1	1.91	0.67	0.0637
	1 or 2/wk	265	22	13.66	14.67	0.0536
	3 or 4/wk	382	40	19.69	26.67	0.1305
	5 or 6/wk	143	12	7.37	8.00	0.0799
	1-2/day	921	53	47.47	35.33	0.0036
	3-4/day	143	14	7.37	9.33	0.1376
	>=5/day	15	1	0.77	0.67	0.3110
Frequency consumed green leafy vegetables between 15w and 20w SCOPE visits	>=1x per day	1079	68	55.62	45.33	Reference
	3-6x per week	525	52	27.06	34.67	0.0184
	1-2x per week	265	22	13.66	14.67	0.2792
	1-3x per month	37	1	1.91	0.67	0.4070
	Never	34	7	1.75	4.67	0.0063

	1-3/mth	521	36	43.71	37.89	0.6766
	1 or 2/wk	366	34	30.70	35.79	0.1890
	3 or 4/wk	59	10	4.95	10.53	0.0186
	5 or 6/wk	4	0	0.34	0.00	0.9787
	1-2/day	10	1	0.84	1.05	0.6413
Frequency consumed hot chips/french fries between 15w and 20w SCOPE visits	Never	232	14	19.46	14.74	Reference
	1-3x per month	521	36	43.71	37.89	0.6766
	1-2x per week	366	34	30.70	35.79	0.1890
	3-6x per week	63	10	5.29	10.53	0.0271
	>=1x per day	10	1	0.84	1.05	0.6413
	Yes	10	3	0.52	1.99	0.0399
PSS: Upset because of something that happened unexpectedly evaluated at 20w SCOPE visit	Never	299	31	15.46	20.67	Reference
	Almost never	776	40	40.12	26.67	0.0050
	Sometimes	679	66	35.11	44.00	0.7777
	Fairly often	153	10	7.91	6.67	0.2210
	Very often	27	3	1.40	2.00	0.9135
PSS: Unable to control the important things in life evaluated at 20w SCOPE visit	Never	552	43	28.54	28.67	Reference
	Almost never	748	65	38.68	43.33	0.5928
	Sometimes	462	36	23.89	24.00	0.9990
	Fairly often	136	3	7.03	2.00	0.0370
	Very often	36	3	1.86	2.00	0.9136

PSS: Confident about ability to handle personal problems evaluated at 20w SCOPE visit	Never	622	38	32.18	25.33	Reference
	Almost never	863	76	44.65	50.67	0.0752
	Sometimes	335	23	17.33	15.33	0.6687
	Fairly often	73	5	3.78	3.33	0.8161
	Very often	40	8	2.07	5.33	0.0049
PSS: Felt things going your way evaluated at 20w SCOPE visit	Never	440	25	22.75	16.67	Reference
	Almost never	965	97	49.90	64.67	0.0137
	Sometimes	446	21	23.06	14.00	0.5359
	Fairly often	57	3	2.95	2.00	0.9028
	Very often	26	4	1.34	2.67	0.0833
PSS: Could not cope with all the things had to do evaluated at 20w SCOPE visit	Never	329	35	17.02	23.33	Reference
	Almost never	846	53	43.77	35.33	0.0198
	Sometimes	583	54	30.16	36.00	0.5430
	Fairly often	143	6	7.40	4.00	0.0400
	Very often	32	2	1.66	1.33	0.4784
Change in exercise level in pregnancy prior to the 20w SCOPE visit evaluated	Decreased	317	33	16.39	21.85	Reference
	Unchanged	1176	94	60.81	62.25	0.2127
	Increased	441	24	22.80	15.89	0.0197
In last month, number of episodes of waking during a night's sleep, evaluated at 20w SCOPE visit	Don’t wake up	129	19	6.67	12.58	Reference
	Once a night	675	47	34.90	31.13	0.0094
	2-3 times	937	64	48.45	42.38	0.0056

	>=4 times	193	21	9.98	13.91	0.3681
Snores most night, evaluated at 20w SCOPE visit	No	1214	87	62.84	57.62	Reference
	Yes	259	29	13.41	19.21	0.0474
Watching >=5h TV per day in the last month, evaluated at 20w SCOPE visit	NO	1742	121	90.07	80.13	Reference
	YES	192	30	9.93	19.87	0.0002
Number of hours spent using a computer per day in the last month, evaluated at 20w SCOPE visit	None	183	29	9.46	19.21	Reference
	<2h	564	30	29.16	19.87	0.0001
	2-4h	295	24	15.25	15.89	0.0222
	5-6h	371	26	19.18	17.22	0.0042
	>6h	521	42	26.94	27.81	0.0084
Any computer usage in last month evaluated at 20w SCOPE visit	NO	183	29	9.46	19.21	Reference
	YES	1751	122	90.54	80.79	0.0002
'Behaviour Responses to Pregnancy' :Not able to carry on with usual activities in pregnant, evaluated at 20w SCOPE visit	Not at all	358	36	18.54	23.84	Reference
	Rarely	484	34	25.06	22.52	0.1499
	Some days	664	54	34.39	35.76	0.3453
	Most days	334	18	17.30	11.92	0.0367
	Everyday	91	9	4.71	5.96	0.9661
'Behaviour Responses to Pregnancy' :Not slowed down since pregnant, evaluated at 20w SCOPE visit	Not at all	193	19	9.99	12.58	Reference
	Rarely	355	14	18.38	9.27	0.0118
	Some days	523	53	27.08	35.10	0.9177
	Most days	652	39	33.76	25.83	0.0874

	Everyday	208	26	10.77	17.22	0.4526
'Behaviour Responses to Pregnancy' :Avoided usual activities, evaluated at 20w SCOPE visit	Not at all	523	51	27.07	33.77	Reference
	Rarely	785	63	40.63	41.72	0.3219
	Some days	522	27	27.02	17.88	0.0099
	Most days	86	8	4.45	5.30	0.9056
	Everyday	16	2	0.83	1.32	0.7452
Edinburgh Postnatal Depression Score evaluated at 20w SCOPE visit categorised	Unlikely to experience depression <5	843	76	43.70	50.33	Reference
	increased risk of depression in the next year 5-9	669	39	34.68	25.83	0.0323
	Likely depressed >9	417	36	21.62	23.84	0.8373
Support people around to provide emotional support evaluated at 20w SCOPE visit	All the time	1316	94	68.08	62.67	Reference
	Most of the time	479	51	24.78	34.00	0.0282
	Sometimes	125	3	6.47	2.00	0.0663
	Seldom	10	1	0.52	0.67	0.7496
	Never	3	1	0.16	0.67	0.1840
Support people around to provide emotional support evaluated at 20w SCOPE visit, compressed categories	All the time	1316	94	68.08	62.67	Reference
	Most of the time	479	51	24.78	34.00	0.0282
	Sometimes	125	3	6.47	2.00	0.0663
	Seldom /Never	13	2	0.67	1.33	0.3172
Social support (listening ears and practical support scores added) categorised	2	1045	78	54.06	52.00	Reference
	3	340	21	17.59	14.00	0.4553

	4	338	42	17.49	28.00	0.0114
	5	119	6	6.16	4.00	0.3667
	>5	91	3	4.71	2.00	0.1721
Any fetal anomalies on 19-21w scan	None or Not assessed or Not visualized	1954	147	98.49	95.45	Reference
	Yes	30	7	1.51	4.55	0.0082
Head circumference on 19-21w scan transformed to multiple of median (MoM) for gestational age <10th centile	No	1733	128	89.05	83.66	Reference
	Yes	213	25	10.95	16.34	0.0444
Head circumference on 19-21w scan transformed to Z score for gestational age <10th Centile	No	1733	128	89.05	83.66	Reference
	Yes	213	25	10.95	16.34	0.0444
Liquor volume reduced at 19-21w scan	Not reduced LiqVol	1983	150	99.95	96.77	Reference
	Reduced LiqVol	1	5	0.05	3.23	0.0001
Liquor volume increased at 19-21w scan	Not Increased LiqVol	1984	155	100.00	100.00	Reference
	Increased LiqVol	0	0	0.00	0.00	0.0001
Mean uterine artery RI at 19-21w >=75th centile	Mean uterine artery RI<0.63	368	44	19.33	29.53	Reference
	Mean uterine artery RI>=0.63	1536	105	80.67	70.47	0.0031
Results of cervical scan communicated to care provider (protocol was to communicate only if <=15mm)	No	1823	136	99.24	97.14	Reference
	Yes	14	4	0.76	2.86	0.0193
Urinary tract infection (lower) in pregnancy between 20w SCOPE and delivery	No (tick box, therefore NO is default option)	1925	140	97.03	90.91	Reference
	Yes	59	14	2.97	9.09	0.0001
Pyelonephritis infection in pregnancy between 20w SCOPE and delivery	No (tick box, therefore NO is default option)	1984	154	100.00	100.00	Reference

	Yes	0	0	0.00	0.00	0.0001
Antibiotic/antifungal treatment in pregnancy between 20w SCOPE and delivery; if no infections in list = yes then this must =NO	No	296	16	43.59	23.53	Reference
	Yes	383	52	56.41	76.47	0.0019
Highest pre-labour proteinuria measured by dipstick	Neg/trace	1340	82	97.17	92.13	Reference
	1+ or 0.3 g/L	33	4	2.39	4.49	0.2069
	2+ or 1 g/L	5	2	0.36	2.25	0.0262
	3+ or >=3 g/L	1	1	0.07	1.12	0.0489
Baby Length measured in neonatometer	No	933	87	47.03	56.49	Reference
	Yes	1051	67	52.97	43.51	0.0241
LGA >90th percentile for customized birthweight centiles adjusted for mothers height, weight at 15w visit, ethnicity, sex and weight and gestation at delivery of baby; all mothers were 0 parity	No	1790	129	90.22	83.77	Reference
	Yes LGA (n=531, 9.4%)	194	25	9.78	16.23	0.0120

VI. Significant Cochran-Mantel-Haenszel (CMH) Test Results for PTB

Variable	P	Variable	P	Variable	P
Gravidity	0.0115	Umbilical artery Doppler RI at 19-21w, transformed to MoM by gestation	0.0229	FH all PTB (spontaneous or iatrogenic) i.e. participant's mother or sister(s) delivered >=2 babies preterm (spont or iatrogenic)	0.0108
Number of D&C or surgical terminations of pregnancy i.e. Number of cervical dilatations	0.0012	Left uterine RI at 19-21w	0.0235	FH recurrent all PTB (spontaneous or iatrogenic) i.e. participant's mother or sister(s) delivered >=2 babies preterm	0.0040
number of vaginal bleeds before 15w SCOPE visit	0.0105	Unemployed or Sickness Beneficiary at 15w SCOPE visit	0.0154	Strong FH all PTB i.e. >=2 family members (participant's mother or sisters) delivered a baby preterm (all PTB- spontaneous or iatrogenic)	0.0483
number of vaginal bleeds before 15w SCOPE visit, spotting severity	0.0285	Participant??s birthweight<1500g	0.0309	Participant's mother had any spontaneous PTB	0.0491
number of vaginal bleeds before 15w SCOPE visit, mod-heavy severity	0.0230	Participant??s birthweight<2500g	0.0399	FH spontaneous PTB i.e. participant's mother or sister(s) had a spontaneous PTB	0.0048
number of vaginal bleeds commencing >12 weeks	0.0014	Participant born preterm (<37 weeks)	0.0166	Any sister with a history of recurrent GDM	0.0046
Total days of vaginal bleeding before 15w SCOPE visit	0.0105	Primigravid	0.0115	FH GDM i.e. participant's mother or sister(s) had GDM	0.0069
Total days of vaginal bleeding before 15w SCOPE visit, spotting	0.0285	Any previous pregnancies	0.0115	FH recurrent GDM i.e. participant's mother or sister(s) had recurrent GDM	0.0288
Total days of vaginal bleeding before 15w SCOPE visit, spotting or light	0.0238	Any previous pregnancy loss with same man who has fathered the current pregnancy	0.0373	participant's mother has had a CVA	0.0206
Total days of vaginal bleeding before 15w SCOPE visit, mod or heavy	0.0230	Any previous pregnancy loss at <=10 weeks gestation	0.0347	participant's father has IHD	0.0354

Total days of vaginal bleeding after 12 weeks gestation	0.0014	Any previous miscarriage or termination at >10 weeks gestation	0.0189	Family history of ischaemic heart disease (participant's mother, father, sibling)	0.0445
Duration 1st vaginal bleed (days) before 15w SCOPE visit	0.0105	Any cervical dilatation	0.0012	Any vaginal bleeding in pregnancy before 15w SCOPE visit	0.0105
Duration 2nd vaginal bleed (days) before 15w SCOPE visit	0.0203	Two previous miscarriages	0.0019	Two or more episodes of vaginal bleeding before 15w SCOPE visit	0.0346
Duration last vaginal bleed (days) before 15w SCOPE visit	0.0105	Any previous termination of pregnancy	0.0420	Any spotting or light vaginal bleeding before 15w SCOPE visit	0.0238
Number of episodes of light vaginal bleeding at or before 6w gestation	0.0390	Duration of sex without contraception with father of baby before current pregnancy <= 3 months	0.0017	Any mod-heavy vaginal bleeding before 15w SCOPE visit	0.0230
Number of episodes of mod-heavy vaginal bleeding at 7-12w gestation	0.0343	Duration of sex without contraception with father of baby before current pregnancy >12 months	0.0047	Any vaginal bleeding after 12w gestation	0.0014
Number of episodes of spotting vaginal bleeding after 12w gestation	0.0104	History of infertility defined as >=12 mths of regular intercourse without contraception and conception has not occurred or if partner is known to be sterile. Binary response 'unknown' combined with NO	0.0037	Any mod or heavy vaginal bleeding continuing for at least 5 days before 15w SCOPE visit	0.0387
Number of episodes of mod-heavy vaginal bleeding after 12w gestation	0.0301	Fertility treatment to conceive current pregnancy	0.0250	Any light vag bleeding at or before 6w gestation	0.0390
Number of episodes of mod-heavy bleeding that last >=5 days before 15w SCOPE visit	0.0387	Hormonal treatment, other than clomiphene, to assist conception of current pregnancy	0.0388	Any mod-heavy vag bleeding at 7-12w gestation	0.0343
Total duration of light vag bleeding at or before 6w gestation	0.0390	Had LLETZ treatment for CIN	0.0015	Any spotting vag bleeding after 12w gestation	0.0104
Total duration of mod-heavy vag bleeding at 7-12w gestation	0.0343	Had either LLETZ, laser or cryotherapy treatment for CIN/abnormal smear	0.0190	Any mod-heavy vag bleeding after 12w gestation	0.0301
Total duration of spotting vag bleeding after 12w gestation	0.0104	Had last LLETZ Rx 7-12 months before conception current pregnancy	0.0332	Any hospital admissions due to trauma before 15w SCOPE visit	0.0225

Total duration of spotting or light vag bleeding after 12w gestation	0.0038	Had last LLETZ Rx>12 months before conception current pregnancy	0.0323	Acupuncture used as alternative therapy at 15w SCOPE visit	0.0350
Total duration of mod-heavy vag bleeding after 12w gestation	0.0301	On metformin for PCOS prior to/at conception	0.0162	Smoked at 15w SCOPE visit	0.0017
Total number of spotting or light vaginal bleeding before 15w SCOPE visit	0.0238	Self reported hypertension (on more than 1 occasion) while on oral contraception	0.0018	Used any marijuana in the 1st trimester	0.0009
Total number of vag bleeding after 6 weeks gestation	0.0309	Mild hypertension prior to pregnancy but never on antihypertensive medication (self reported) or at antenatal booking systolic BP 140-159 or diastolic 90-99 mmHg; Some of these women will be white coat hypertension	0.0215	Using marijuana at 15w SCOPE visit	0.0001
Total duration of vag bleeding after 6w gestation (days)	0.0334	Any sister who had a history of miscarriage	0.0142	Any use of marijuana during pregnancy	0.0009
Number of hospital admissions due to trauma before 15w SCOPE visit	0.0225	Participant's mother had any history of recurrent pregnancy induced hypertension	0.0219	Extreme exercise in pregnancy (undertook vigorous exercise at least once a day) evaluated at 15w SCOPE visit	0.0261
number of cigarettes per day at 15w SCOPE visit	0.0009	Any family history of recurrent PIH	0.0329	Any vaginal bleeding recorded at either 15w or 20w SCOPE visit	0.0162
Total number of cigarettes a woman was exposed to in the 1st trimester	0.0394	Family history of PIH (i.e. GH or preeclampsia, but unsure which condition), gestational hypertension or preeclampsia; family members are participant's mother and/or sisters	0.0458	Any vaginal bleeding between 13 weeks' gestation and 20w SCOPE visit (could be recorded at either 15w or 20w visit)	0.0074
Gestation marijuana ceased in pregnancy	0.0023	Participant's mother had any history of recurrent PET	0.0147	Smoking at 20w SCOPE visit (week prior to interview)	0.0039
2nd diastolic BP at 15w SCOPE visit measured by mercury or aneroid sphygmomanometer	0.0091	Family history of PET, i.e. participant's mother or sister had had PET	0.0292	Serum taken at 20w SCOPE visit	0.0005
2nd MAP (mean arterial pressure) at 15w	0.0265	Family history of recurrent PET, i.e.	0.0199	EDTA plasma taken at 20w SCOPE visit	0.0080

SCOPE visit measured by mercury or aneroid sphygmomanometer		participant's mother or sister had had recurrent PET			
How many hours of sleeping during day on weekend day on average at 15w SCOPE visit	0.0199	Family history (mother or sisters) of GH or PET	0.0422	Watching >=5h TV per day in the last month, evaluated at 20w SCOPE visit	0.0060
Participant's ferritin (ug/L) before 20 weeks	0.0010	Family history (mother or sisters) of recurrent GH or recurrent PET	0.0051	Any computer usage in last month evaluated at 20w SCOPE visit	0.0169
bb_trig	0.0016	Participant's mother had any history of LBW baby	0.0080	Any fetal anomalies on 19-21w scan	0.0096
Total days of vaginal bleeding between 13 weeks' gestation and 20w SCOPE visit (could be recorded at either 15w or 20w visit)	0.0074	Any sister had a history of LBW baby	0.0134	Mean uterine artery RI at 19-21w >=75th centile	0.0263
number of cigarettes per day in the week prior to 20w SCOPE visit	0.0052	Family history of LBW baby, i.e. participant's mother or sister had had LBW baby	0.0024	Results of cervical scan communicated to care provider (protocol was to communicate only if <=15mm)	0.0358
Number of times marijuana was taken between 15w and 20w SCOPE visit	0.0001	Strong family history (2 or more members of immediate family - mother or sisters) of LBW baby	0.0033	If cervical scan results communicated then antibiotics or cerclage or nifedipine or NSAID or Betamimimetics or MgSO4 Rx commenced	0.0005
Random glucose measured by glucometer at 20w SCOPE visit (mmol/L)	0.0224	Participant's mother had any PTB (spontaneous or iatrogenic)	0.0285	Flu/Respiratory tract infection between 20w SCOPE and delivery	0.0236
How many hours of sleeping at night on weekend day on average at 20w SCOPE visit	0.0478	Participant's mother had history of recurrent PTB (all-spontaneous or iatrogenic)	0.0093	Urinary tract infection (lower) in pregnancy between 20w SCOPE and delivery	0.0003
Umbilical artery Resistance Index (RI) measured using Doppler ultrasound at 19-21w	0.0291	Any sister with a history of PTB (all-spontaneous or iatrogenic)	0.0420	Antibiotic/antifungal treatment in pregnancy between 20w SCOPE and delivery; if no infections in list = yes then this must =NO	0.0010

VII.Paper I: Tiered Prediction System for Preeclampsia

Leemaqz, S.Y., Dekker, G.A. and Roberts, C.T. (2013) Tiered Prediction System for Preeclampsia: an integrative application of multiple models. In Piantadosi, J., Anderssen, R.S. and Boland J. (eds) *MODSIM2013, 20th International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand*, December 2013, pp. 2041–2046. ISBN: 978-0-9872143-3-1. www.mssanz.org.au/modsim2013/L5/leemaqz.pdf

VIII. Paper II: Risk Factors for Preterm Birth

OPEN ACCESS Freely available online

PLoS one

Risk Factors for Preterm Birth in an International Prospective Cohort of Nulliparous Women

Gustaaf Albert Dekker[1,2]*, Shalem Y. Lee[2], Robyn A. North[3], Lesley M. McCowan[4], Nigel A. B. Simpson[5], Claire T. Roberts[2]

1 Lyell McEwin Hospital, Adelaide, South Australia, Australia, **2** Robinson Institute, University of Adelaide, Adelaide, South Australia, Australia, **3** Division of Women's Health, Women's Health Academic Centre, King's College London and King's Health Partners, London, United Kingdom, **4** Department of Obstetrics and Gynaecology, University of Auckland, Auckland, New Zealand, **5** Academic Department of Obstetrics and Gynaecology, University of Leeds, Leeds, United Kingdom

Abstract

Objectives: To identify risk factors for spontaneous preterm birth (birth <37 weeks gestation) with intact membranes (SPTB-IM) and SPTB after prelabour rupture of the membranes (SPTB-PPROM) for nulliparous pregnant women.

Design: Prospective international multicentre cohort.

Participants: 3234 healthy nulliparous women with a singleton pregnancy, follow up was complete in 3184 of participants (98.5%).

Results: Of the 3184 women, 156 (4.9%) had their pregnancy complicated by SPTB; 96 (3.0%) and 60 (1.9%) in the SPTB-IM and SPTB-PPROM categories, respectively. Independent risk factors for SPTB-IM were shorter cervical length, abnormal uterine Doppler flow, use of marijuana pre-pregnancy, lack of overall feeling of well being, being of Caucasian ethnicity, having a mother with diabetes and/or a history of preeclampsia, and a family history of low birth weight babies. Independent risk factors for SPTB-PPROM were shorter cervical length, short stature, participant's not being the first born in the family, longer time to conceive, not waking up at night, hormonal fertility treatment (excluding clomiphene), mild hypertension, family history of recurrent gestational diabetes, and maternal family history of any miscarriage (risk reduction). Low BMI (<20) nearly doubled the risk for SPTB-PPROM (odds ratio 2.64; 95% CI 1.07–6.51). The area under the receiver operating characteristics curve (AUC), after internal validation, was 0.69 for SPTB-IM and 0.79 for SPTB-PPROM.

Conclusion: The ability to predict PTB in healthy nulliparous women using clinical characteristics is modest. The dissimilarity of risk factors for SPTB-IM compared with SPTB-PPROM indicates different pathophysiological pathways underlie these distinct phenotypes.

Trial Registration: ACTR.org.au ACTRN12607000551493

Citation: Dekker GA, Lee SY, North RA, McCowan LM, Simpson NAB, et al. (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

Editor: Hamid Reza Baradaran, Tehran University of Medical Sciences, (Islamic Republic of Iran)

Received April 13, 2012; **Accepted** May 16, 2012; **Published** July 16, 2012

Funding: This study was funded by New Enterprise Research Fund, Foundation for Research Science and Technology; Health Research Council (04/198); Evelyn Bond Fund, Auckland District Health Board Charitable Trust; and the Premier's Science and Research Fund, South Australian Government. The study sponsors had no role in study design, data analysis or writing this report.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: gustaaf.dekker@adelaide.edu.au

Introduction

In the developed world, spontaneous preterm birth (SPTB) is without doubt a major problem in modern obstetrics; its prevalence is still rising in many industrialised countries. According to the USA National Vital Statistics Reports, 11–12% of the 4 million neonates born each year are delivered before 37 weeks and 3.6% are delivered before 34 weeks [1–3]. Early PTB (before 34 weeks) is particularly associated with high rates of mortality and morbidity, including intraventricular hemorrhage, necrotizing enterocolitis, respiratory distress syndrome and neurological deficit [2]. PTB has long-term medical and social sequelae; the risk of medical and social disabilities in adulthood increases with decreasing gestational age at birth [4,5].

To identify women at risk of SPTB, clinicians use prior preterm birth, multiple pregnancy and prior cervical surgery as major risk factors. Useful clinical risk factors in predicting SPTB in nulliparous women with a singleton pregnancy are scant, except for a history of prior cervical surgery. In low risk women, maternal history alone misses more than half of the women at risk for SPTB [6]. The use of vaginal posterior fornix testing for fetal fibronectin only yields meaningful positive tests after 22 weeks gestation and may be only a few weeks prior to the actual preterm birth. Measuring cervical length is the only screening test for SPTB that has been shown to have potential for effective intervention. Fonseca et al. [7] demonstrated, in a cohort of seemingly low risk women with cervical length ≤1.5 cm at 20 weeks gestation (n = 413), that vaginal progesterone reduced the risk of SPTB by

45%. While most countries have not introduced routine screening for cervical shortening in asymptomatic patients, a recent cost-effectiveness analysis concluded that on the basis of the published efficacy of vaginal progesterone treatment, cervical length measurement should become part of routine antenatal care [8].

It is important to note that 'preterm birth' is in itself not a diagnosis. The term describes the clinically easily identifiable end-result of various different major pathophysiological pathways. Preterm labour leading to SPTB may present with intact membranes (SPTB-IM) or following spontaneous rupture of membranes (SPTB-PPROM); the pathways leading to these different clinical phenotypes are likely to be different [9].

The SCOPE (Screening for Pregnancy Endpoints) study is a prospective, multi-centre cohort study of 'healthy' nulliparous women, with the primary aim of developing screening tests to predict preeclampsia, small for gestational age (SGA) infants and SPTB. The study design incorporated prospective collection of information on all known clinical risk factors for preterm birth.

The objectives for this part of SCOPE were to identify risk factors for SPTB-IM and SPTB-PPROM and to develop multivariable predictive models based on clinical risk factors present in early pregnancy (15±1 weeks), together with cervical length measurements and routine sonographic findings obtained during the 'morphology scan' at 20±1 weeks' gestation.

Methods

The STROBE checklist for this trial is available as supporting information; see Checklist S1.

Nulliparous women with singleton pregnancies were recruited to the SCOPE study between November 2004 and August 2008 in Auckland, New Zealand, and Adelaide, Australia. Ethical approval was obtained from local ethics committees [New Zealand AKX/02/00/364, Australia REC 1712/5/2008] [10] and all women provided written informed consent.

Women attending hospital antenatal clinics, obstetricians, general practitioners or community midwives prior to 15 weeks' gestation were invited to participate in the SCOPE study. Women were excluded if (1) they were judged to be at a particularly high risk of pre-eclampsia, SGA or SPTB due to underlying medical conditions (chronic hypertension requiring antihypertensive medication, diabetes, renal disease, systemic lupus erythematosus, antiphospholipid syndrome, sickle cell disease, human immunodeficiency virus), previous cervical knife cone biopsy, ≥3 terminations or ≥3 miscarriages, current ruptured membranes; 2) their pregnancy was complicated by a known major fetal anomaly or abnormal karyotype or 3) they had received interventions that might have modified pregnancy outcome (e.g., aspirin, cervical suture) [10]. Participants were interviewed and examined by a research midwife at 15±1 and 20±1 weeks of gestation and underwent an ultrasound scan at 20±1 weeks. At the time of interview, data were entered into an internet accessed, central database with a complete audit trail (MedSciNet^{AB}).

At time of recruitment the following data were collected [10]: demographic information including age, ethnicity, immigration details, education, work, socioeconomic index, income level, living situation; the woman's birthweight and gestation at delivery, and whether it was a singleton or multiple pregnancy; previous miscarriages, terminations or ectopic pregnancies and whether these pregnancies were with the same partner as the current pregnancy or not; history of infertility, use of assisted reproductive technologies, duration of sexual relationship and exposure to partner's sperm; gynaecological (number of cervical dilatations, abnormal PAP smears, and treatment for cervical intraepithelial neoplasia, polycystic ovarian syndrome) and medical history including hypertension while taking combined oral contraception, asthma, urinary tract infection, inflammatory bowel disease, thyroid disease and thrombo-embolism; family history (mother, sisters) of obstetric complications (miscarriage, preeclampsia, eclampsia, gestational hypertension, spontaneous preterm birth,

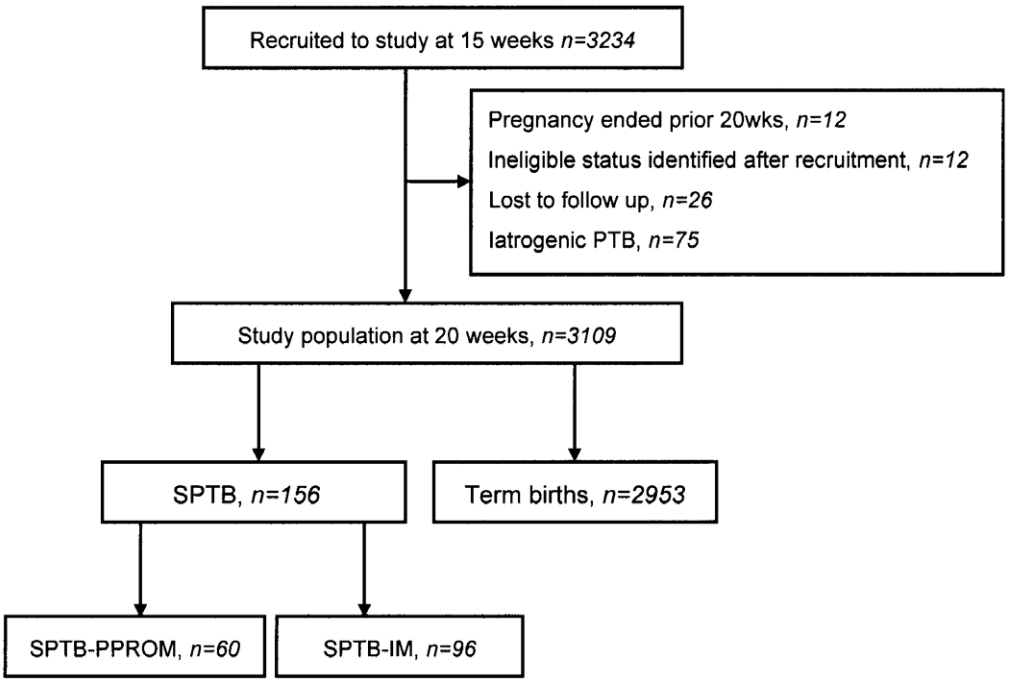

Figure 1. Participants recruited and study population.
doi:10.1371/journal.pone.0039154.g001

Table 1. Clinical characteristics.

	Term births	SPTB-IM	P	SPTB-PPROM	P
Maternal Characteristics	**2953**	**96**		**60**	
Age	28.0 (5.8)	27.6 (6.5)	0.50	28.0 (5.8)	0.90
BMI	25.6 (5.3)	26.1 (5.5)	0.35	25.2 (6.0)	0.58
Height (cm)	165.2 (6.6)	164.5 (6.9)	0.26	163.3 (6.7)	***0.023***
Head circumference (cm)	56.0 (1.7)	55.9 (1.4)	0.47	55.5 (1.6)	***0.019***
Systolic BP (mmHg)	108 (11)	108 (10)	0.95	107 (11)	0.55
Diastolic BP (mmHg)	65 (8)	66 (8)	0.31	65 (8)	0.86
Caucasian	2558 (86.6)	90 (93.6)	***0.048***	52 (86.7)	0.99
First born	1708 (58.1)	42 (44.2)	0.66	15 (25.0)	***0.01***
Social Characteristics					
SEI	40.674 (16.5)	39.5 (17.3)	0.51	40.3 (15.1)	0.87
Full-time employment	1972 (66.8)	58 (60.4)	0.19	44 (73.3)	0.29
Diet Characteristics					
Smoking (15 weeks)	313 (10.6)	22 (22.9)	***0.000***	9 (15.0)	0.28
Marijuana (pre-preg)	191 (6.5)	17 (17.7)	***0.000***	5 (8.3)	0.57
Marijuana (1st trimester)	31 (1.0)	8 (8.3)	***0.000***	2 (3.3)	0.11
Psychological Characteristics					
Anxiety Index >90%	211 (7.2)	12 (12.6)	***0.049***	6 (10.0)	0.41
Not feeling better than ever	2275 (77.5)	83 (86.5)	***0.04***	48 (80.0)	0.64
Obstetric Characteristics					
Gravidity	1.3 (0.6)	1.6 (0.8)	***0.000***	1.4 (0.6)	0.54
Months to conceive	5.9 (11.6)	7.4 (11.9)	0.23	11.9 (22.1)	***0.000***
<=3 months to conceive	1871 (63.6)	51 (53.1)	***0.038***	31 (51.7)	0.06
Donor sperm	141 (4.8)	5 (5.2)	0.84	8 (13.3)	***0.004***
Hormonal treatment	90 (3.0)	2 (2.1)	0.59	7 (11.7)	***0.001***
Mild Hypertension (not on treatment)	29 (1.0)	2 (2.1)	0.30	3 (5.0)	***0.007***
LLETZ	107 (3.6)	7 (7.3)	0.07	7 (11.7)	***0.002***
>1 Vaginal bleeding	145 (4.9)	9 (9.4)	0.05	4 (6.7)	0.54
APH	162 (5.5)	23 (24.0)	***0.000***	5 (8.6)	0.31
Waking at night					
Once	918 (31.2)	27 (28.1)	0.10	13 (21.7)	***0.014***
≥2 times	1837 (62.5)	59 (61.5)	0.13	39 (65.0)	0.07
Cervical length (mm)	41.0 (7.4)	38.7 (7.9)	***0.006***	38.9 (6.9)	***0.047***
Average UTRI >90%	240 (7.5)	17 (18.1)	***0.002***	7 (12.7)	0.27
Average UTRI	0.56 (0.09)	0.59 (0.12)	***0.002***	0.57 (0.09)	0.29
Family History					
Gestational diabetes	106 (3.6)	8 (8.3)	***0.020***	5 (8.3)	0.062
Recurrent GDM	19 (0.6)	2 (2.1)	0.11	2 (3.3)	***0.027***
Preeclampsia	284 (9.6)	20 (20.8)	***0.000***	5 (8.3)	0.74
Mother had preeclampsia	233 (7.9)	16 (16.7)	***0.003***	4 (6.7)	0.73
Gestational Hypertension	6 (0.2)	0 (0.0)	0.98	1 (1.7)	***0.051***
Miscarriage (mother)	888 (30.1)	28 (29.2)	0.85	10 (16.7)	***0.028***
Diabetes Type 2 (mother)	137 (4.6)	9 (9.4)	***0.037***	2 (3.3)	0.63
Low birthweight baby*	27 (0.9)	5 (5.2)	***0.000***	1 (1.7)	0.55
Birth Outcomes					
Gestational age 40 (1)		34 (4)	0.97	33 (5)	0.97
Birthweight (g)	3481 (472)	2378 (736)	***0.000***	2379 (761)	***0.000***
Customized centile	49 (29)	49 (31)	0.85	51 (32)	0.50

Table 1. Cont.

	Term births	SPTB-IM	P	SPTB-PPROM	P
SGA	285 (10)	11 (11.5)	0.56	6 (10)	0.93

Characteristics as mean (SD) or n (%); head circumference and height in centimetres; [x] mother/sister with low birth weight baby; APH = antepartum haemorrhage; BP = blood pressure; UTRI = uterine artery resistance index.
doi:10.1371/journal.pone.0039154.t001

any preterm birth, gestational diabetes, stillbirth and neonatal death) and family history (mother, father, sibling) of medical conditions (hypertension, coronary artery heart disease, cerebrovascular accident, type 1 and 2 diabetes and venous thromboembolism).

Information was collected on early pregnancy vaginal bleeding (gestation, severity and duration of bleeding and number of bleeding episodes), hyperemesis and infections during pregnancy. Vegetarian status was recorded and other dietary information preconception and during pregnancy was obtained using food frequency questions for fruit, green leafy vegetables, oily and other fish and fast foods. Use of folate and multivitamins, cigarettes, alcohol (including binge drinking) and recreational drugs (including marijuana, amphetamine, cocaine, heroin, ecstasy, lysergic acid diethylamide) was recorded for preconception, 1st trimester and at 15 weeks. A lifestyle questionnaire was completed on work, exercise and sedentary activities, snoring, domestic violence and social support. Psychological scales were completed measuring perceived stress, depression, anxiety, and behavioural responses to pregnancy (adapted from the Behavioural Responses to Illness Questionnaire [11]). Two consecutive manual blood pressure measurements were recorded. Other maternal measurements included maternal height, weight and waist, hip, arm and head circumference.

Ultrasound examination at 20±1 weeks' gestation included measurements of the fetus (biparietal diameter, head circumference, abdominal circumference and femur length), Doppler studies of the umbilical and uterine arteries, and transvaginal cervical length measurements [12]. Notching of each uterine artery was recorded. An abnormal uterine artery.

Doppler result was defined as a mean resistance index >90th centile (>0.695) [12].

The technique used to measure the cervical length was that modified from Berghella et al. [13].

As described by Gomez et al [14] no fundal or suprapubic pressure was applied during the examinations. All fetal measurements were adjusted for gestational age by calculating the multiple of the median for each gestational week.

Participants were followed prospectively, with pregnancy outcome data and baby measurements collected by research midwives. Data monitoring included 1) individually checking all data for each participant, including any transcription errors of the lifestyle questionnaire, and 2) detection and correction of illogical or inconsistent data and outliers using customised software.

Primary outcome: The primary outcome was SPTB (birth <37 weeks' gestation) as per the two main phenotypes, i.e. SPTB-IM

Table 2. Clinical risk factors at 15 weeks, and ultrasound scan variables at 20 weeks in logistic regression model for SPTB-IM.

SPTB-IM			
	OR	Lower 95%	Upper 95%
BMI <20	1.46	0.62	3.42
BMI 25–30	1.63	0.96	2.79
BMI >30	1.21	0.63	2.32
Caucasian ethnicity	2.73	0.98	7.60
Marijuana pre-pregnancy	2.34	1.22	4.52
Not feeling better than ever	1.78	0.90	3.51
Having a history of >1 vaginal bleed	2.33	1.08	5.04
Mother with diabetes type 1 or 2	2.19	0.99	4.86
Mother with a history of preeclampsia	2.34	1.30	4.21
Strong family history of low birth weight babies	5.64	1.79	17.80
Abnormal uterine artery Doppler 20 wks	2.18	1.20	3.94
Shortest transvaginal cervical length in mm	1.05	1.01	1.08

Reference BMI 20–<25.
doi:10.1371/journal.pone.0039154.t002

Table 3. Clinical risk factors at 15 weeks, and ultrasound scan variables at 20 weeks in logistic regression model for SPTB-PPROM.

SPTB-PPROM			
	OR	Lower 95%	Upper 95%
BMI <20	2.64	1.07	6.51
BMI 25–30	1.20	0.57	2.51
BMI >30	0.94	0.39	2.26
Height (per cm)	0.93	0.89	0.97
Participant position in family	1.91	0.97	3.76
Waking once a night	0.32	0.12	0.89
Waking more than once a night	0.45	0.19	1.05
Months to conceive (per month)	1.02	1.00	1.03
Other hormonal fertility treatment[1]	3.67	1.24	10.83
Mild hypertension not requiring treatment	9.65	2.51	37.14
Family history of recurrent GDM[2]	8.01	1.51	42.45
Maternal family history of any miscarriage	0.43	0.19	0.94
Shortest transvaginal cervical length per mm	1.05	1.01	1.09

1 = hormonal fertility treatment other than clomiphene; GDM = gestational diabetes mellitus; participant's position in family = index mother not being the first-born); Reference BMI 20–<25.
doi:10.1371/journal.pone.0039154.t003

and SPTB-PPROM. SPTB-PPROM was defined as SPTB where the women presented with confirmed rupture of the membranes in the absence of labour and the time between the rupture of the membranes to delivery was at least 6 hours greater than the combined time for established labour (i.e. duration of first stage + duration of second stage [10]).

Statistical Methods

Women who had SPTB-IM or SPTB-PPROM were separately analyzed and compared with all women who had term births. Variables with more than 10% missing data were excluded from analyses, with the exception of the dental health variables included in the univariate analysis only (available in 38% of participants as added later to the database) and cervical length in the multivariable analysis. Of the variables selected for modelling, data were complete in >99% of participants for each variable other than cervical length (18.6% missing data), uterine artery Doppler (5% missing) and participant born preterm before 34 weeks' gestation (4% missing). Missing data was handled in the multivariable analysis by omitting participants with any missing data. R version 2.12.1 was used to perform the analyses. Univariate data analyses including Student's *t* test and Chi-square tests were used to compare and test the association of predictors with SPTB-IM and SPTB-PPROM.

A total of 933 variables were tested for association with SPTB-PPROM and SPTB-IM separately using univariate analysis. Variables were then excluded due to P value >0.1 on univariate comparison (797 variables for SPTB-PPROM, 691 variables for SPTB-IM), variables with >10% missing data (5 variables for SPTB-PPROM, 11 variables for SPTB-IM), and variables assessed after 15 weeks of gestation with the exception of uterine artery resistance index and cervical length both measured at 20 weeks of gestation (65 variables for SPTB-PPROM, 87 variables for SPTB-IM). Of the remaining variables, a list of 49 variables for SPTB-PPROM and 30 variables for SPTB-IM were selected based on known predictors and variables of interest. The initial variable lists used to train the multivariate models are available as supporting information (File S1). Two multivariable logistic regression models were then trained for SPTB-PPROM and SPTB-IM based on corresponding selected predictors. A backward stepwise method was used to develop an optimal model. Akaike Information Criteria (AIC) were obtained for each model as a goodness of fit measure and the optimal model was determined based on minimum AIC [15]. The sensitivity and specificity were calculated as measures of goodness of classification. The receiver operating characteristic (ROC) curve and the Area Under Curve (AUC) [16] were also obtained to assess predictive utility. Ten-fold cross validations were performed on all models using 90% of the data randomly chosen for training purposes, and validating on the remaining 10%.

Results

3234 nulliparous pregnant women with singleton pregnancies were recruited to the SCOPE study between November 2004 and

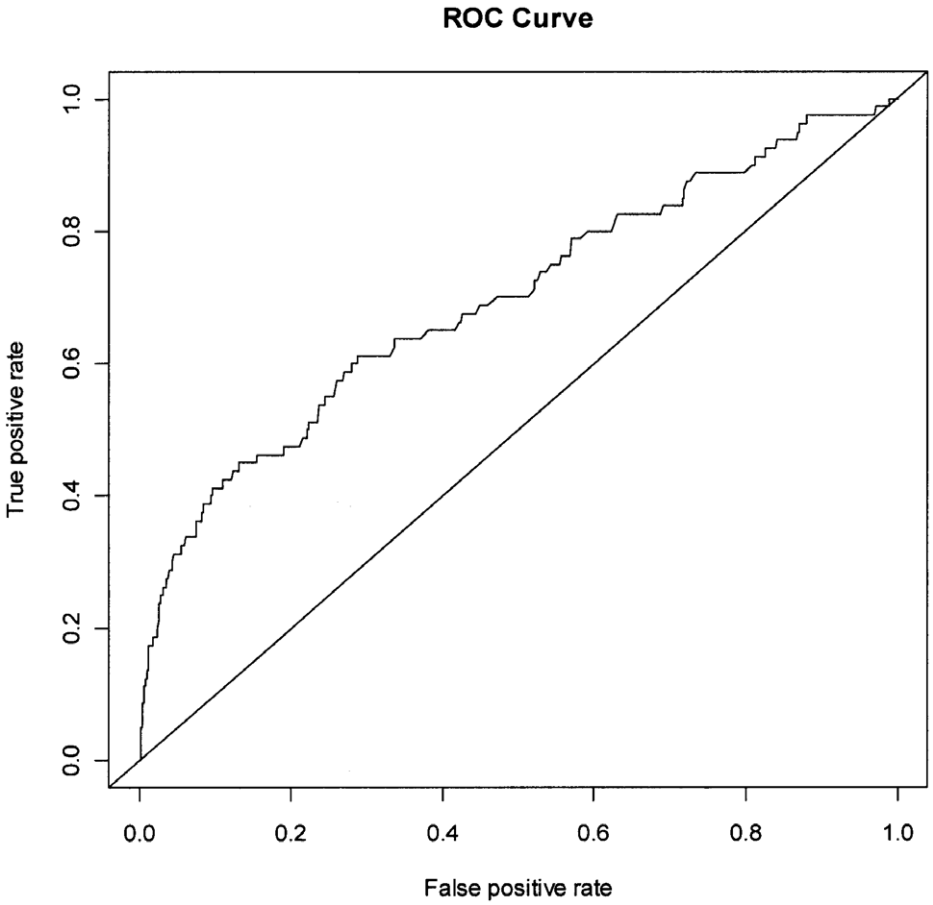

Figure 2. Receiver-operating characteristic curve for SPTB-IM.
doi:10.1371/journal.pone.0039154.g002

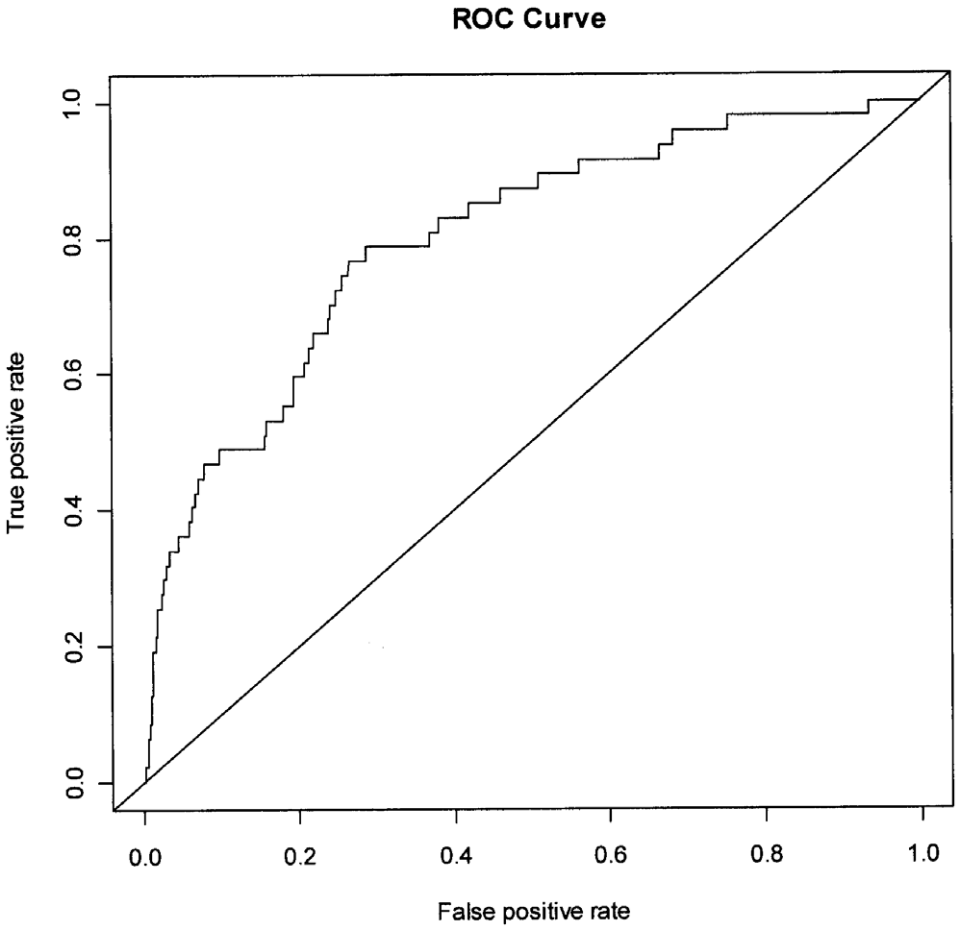

Figure 3. Receiver-operating characteristic curve for SPTB-PPROM.
doi:10.1371/journal.pone.0039154.g003

August 2008 in Auckland, New Zealand and Adelaide, Australia. Follow up was complete in 3184 (98.5%) of participants (Figure 1). Of the total of 156 SPTB, 96 (61.5%) were in the SPTB-IM and 60 (38.5%) in SPTB-PPROM categories. Women with iatrogenic PTB were excluded from the study population.

After omitting participants with any missing data, a total of 2499 (80.4%) patients for SPTB-IM and 2455 (79%) patients for SPTB-PPROM were included in the logistic regression analyses.

The characteristics in this cohort of nulliparous pregnant women with term birth, and the 2 main subtypes of SPTB are shown in Table 1.

In the 1987 participants in whom data on dental health were collected, there was no difference in a history of easily bleeding gums, swollen gums or sore teeth prior to or during the first trimester of pregnancy between the term birth group and either SPTB-IM and SPTB-PPROM.

Clinical risk factors recorded at 15 weeks' gestation and the ultrasound scan results from the 20 weeks' gestation, with significant independent associations for SPTB-IM and SPTB-PPROM, and/or contributing to the model are shown in Table 2 and 3, respectively.

In the logistic regression model for SPTB-IM, a shorter cervical length as a continuum was associated with an increased risk of 1.04 per mm decrease in cervical length. Regular marijuana use up to conception was a significant and strong risk factor. Similar risks were found to be associated with the presence of an abnormal uterine Doppler flow velocity waveform pattern at 20 weeks' gestation and maternal family history of any type of diabetes and/or preeclampsia. A strong family history of low birth weight babies (mother and/or sister with a low birth weight baby) was the strongest risk factor with odds exceeding 5. With regard to a history of vaginal bleeding, only the presence of more than one episode of vaginal bleeding was an independent risk marker. 'Not feeling better than ever' contributed to the model for SPTB-IM, though the odds ratio crossed unity (odds ratio 1.78; 95% CI 0.90–3.51).

Whilst Caucasian ethnicity and a low or elevated BMI were included as independent risk factors in the model, the confidence intervals for each adjusted OR crossed unity.

Except cervical length, the independent variables in the SPTB-PPROM model (table 3) were strikingly different to those in the SPTB-IM model. Having a low BMI had an odds ratio of 2.64. For every cm maternal height increase there was a 7% reduced risk for SPTB-PPROM. Length of sexual cohabitation in months, as a continuum, increased the risk by one percent per additional month. Having a history of hormonal fertility treatment (excluding clomiphene), and having mild hypertension (chronic hypertension requiring treatment was an exclusion criterion for the SCOPE study) were both independent risk factors. Having a family history of recurrent gestational diabetes was strongly associated with SPTB-PPROM, albeit with large confidence intervals. Participant's position in family (index mother not being the first-born) was a significant independent risk factor.

The predictive capability for SPTB-IM is shown in figure 2; AUC 0.69, with a sensitivity of 0.39 based on 90% specificity. Figure 3 shows the ROC curve for SPTB-PPROM; AUC 0.79, with a sensitivity of 0.49 based on 90% specificity.

Discussion

Analysis of data from this large prospective cohort of low-risk nulliparous pregnant women have demonstrated that clinical risk factors, including cervical length and uterine artery Doppler ultrasound measurements at 20 weeks, have only a modest predictive capacity for the two major phenotypes of SPTB. In this particular analysis we selected a case-control approach instead of a case – non case approach because of potential overlap in pathophysiology not only between the 2 major phenotypes but also between iatrogenic preterm birth and SPTB. Most likely, a strict case-non case approach would have further dropped the performance of the models. While it is clear that these risk markers by themselves cannot be translated into a useful clinical tool for daily practice, the data provide further insight into these conditions.

The minimal overlap between risk factors for SPTB-PPROM and SPTB-IM reinforces the increasingly accepted view that SPTB is a heterogeneous entity with different pathological pathways leading to SPTB with or without intact membranes [9] and also differences between patients with SPTB at different gestational ages [17–19]. This heterogeneity is illustrated by the observation that antepartum haemorrhage (APH) is significantly more common in the SPTB-IM group (24%) than the SPTB-PPROM group (8.6%) or term births (5.5%). APH was not entered in the multivariate analysis since it occurs by definition after 20 weeks' gestation.

Regarding variables related to placentation, we found a lengthier sexual relationship (as a continuum) known to exert a protective effect for preeclampsia and intra-uterine growth restriction [20], to be associated with a small but significant increased risk for SPTB-PPROM. It should be noted that in univariate analysis (table 1), conceiving within 3 months (table 1) was also less common in both SPTB phenotypes compared with term birth (SPTB-IM $p = 0.038$; SPTB-PPROM $p = 0.06$). In contrast, donor insemination was significantly ($p = 0.005$) more common in the SPTB-PPROM group (8 out of 60; 13.3%) versus term birth (4.8%). While, the presence of abnormal uterine Doppler flow patterns at the time of the morphology scan nearly doubled the risk for SPTB-IM this was not an independent risk factor for SPTB-PPROM. Also recurrent vaginal bleeding in early pregnancy, a previously described risk factor [21], while doubling the risk for SPTB-IM was not a risk factor for SPTB-PROM.

Decreased cervical length (per mm decrease) was the only variable with a comparable effect in both SPTB phenotypes; 4 and 5% increased risk for SPTB-IM and SPTB-PPROM, respectively. This would mean that for example the risk for SPTB for two comparable nulliparous pregnant women with cervical length of 41 mm versus 28 mm at 20 weeks gestation would be at least 60% higher in the woman with the shorter cervix. Using a cost-effectiveness analysis, Werner et al [8] predicted if there were universal cervical length screening, there would be a net health improvement of 735 quality adjusted life years and net savings to the healthcare system (USA data) of $19 million for every 100 000 women screened. This cost-effectiveness analysis was primarily based on the Fonseca et al [7] study, but the results were analysed and confirmed by including the recent result of the Hassan et al multicentre study [22]. Unfortunately, these 2 large multicentre vaginal progesterone studies do not specifically address the SPTB phenotype.

Most of the independent risk factors for SPTB-IM could, at least in theory, fit in one of the seven major molecular pathways previously described by Romero et al [23]. 'Not feeling as well' could be a marker of stress or lack of support, and as such fits in one of the pathways to preterm birth [23]. In contrast to several epidemiological studies on stress and employment [24,25], the other variables capturing data on employment, income, anxiety and depression were not independent risk factors.

We have shown that marijuana is a strong 'environmental risk factor or SPTB-IM in this population. We are unable to determine whether this association is due to a toxic effect of marijuana or is a marker of a suite of lifestyle factors that contribute to the risk. Pre-pregnancy marijuana use may be a more reliable marker since one can anticipate that women would be more likely to disclose it than persistent marijuana use during pregnancy. In contrast to the results of this large prospective cohort study, large American population studies [26–28], did not find an association between maternal marijuana use and preterm birth.

In this cohort of 3234 low risk nulliparous women, with 156 cases of SPTB, we do find the highest rate of smokers amongst the SPTB-IM group (22.9% versus 10.6% in term births; p 0.00), with an intermediate rate in the SPTB-PPROM group (15%; p 0.291). However, smoking was not an independent risk factor for either phenotype. Because of our very rich data it is possible that the effect of smoking is now explained by other variables in the models such as abnormal uterine artery Doppler [29]. Maternal tobacco smoking has typically been described as a risk factor for SPTB in many studies; however the mechanism for this effect remains unclear [30]. In a retrospective cohort study covering all preterm births in the major tertiary referral centre in Western Australia during the period 2004–2008, Henderson et al [31] found a significant association of smoking in only one SPTB subtype: SPTB-PPROM between 27 and 33 weeks' gestation, and suggested that these data indicate that tobacco smoking may have a specific effect on the fetal membranes while not influencing spontaneous labour. Furthermore, an analysis based on a large Swedish population cohort [30] demonstrated that smoking ($\geq$10 cigarettes per day; odds ratio 1.7) was primarily associated with increased risks of very preterm birth and there were small numbers of very preterm births in this cohort.

Ethnic differences in the prevalence of various adverse pregnancy outcomes, including SPTB, have been previously described [32,33]. Although specific high risk genetic polymorphisms may partially explain those ethnic differences, most studies appear to point to socio-economic deprivation, smoking, obesity, poverty-induced stress and the associated poor nutrition as the key mediators. It should be noted that the non-Caucasian pregnant women in this SCOPE cohort consisted primarily of women of Asian descent and to a lesser degree also Maori and Pacific Island women. The low total number of non-Caucasian ethnicities did not permit further sub-analysis. Surprisingly (on univariate comparison) Caucasian ethnicity was significantly more common in the SPTB-IM group. Being of Caucasian ethnicity, as an independent variable in the regression model, more than doubled the risk for SPTB-IM, although the 95% CI just crossed 1. Although this was not captured by our socio-economic variables, these findings might be explained by the fact that women in the Australian part of the SCOPE study come from one of the most underprivileged urban areas in Australia with a primarily Caucasian population [34,35]. Our data demonstrate that taking a full family history can provide potentially important indicators for risk for SPTB, as a strong family history of low birth weight

babies was the strongest risk factor with odds exceeding 5 (albeit present in just over 1% of the whole cohort) for SPTB-IM, while a positive family history in the mother for preeclampsia and any type of diabetes more than doubled the risk.

In addition to decreased cervical length, BMI was the only variable present in both models. Conventional wisdom indicates that women with low BMI are at increased risk for SPTB, while the association between maternal overweight or obesity and SPTB remains controversial. Heterogeneity in the definitions of pregnancy outcomes (spontaneous vs. medically indicated PTB) and the inclusion of different gestational ages in delivery categories in various studies are probably only a partial explanation for these controversies [36]. In this prospective cohort low BMI, doubled the risk for SPTB-PPROM with the odds ratio just crossing 1 (odds ratio 2.1; 95% CI 0.93–4.54). It is not surprising that the contemporary literature regarding BMI and the risk for preterm birth, and indeed any adverse pregnancy outcome, is often conflicting. In the past low BMI was associated with undernutrition. However, more recently obesity has become a marker of socio-economic deprivation with overconsumption of calorie-dense but nutrient-poor food [34,35].

In contrast to the independent risk factors associated with SPTB-IM, those associated with SPTB-PPROM are largely difficult to explain, and considering the number of variables in the final analysis for SPTB-PROM (49 variables) could well represent false discoveries for some of these findings.

To our knowledge, these data are the first to suggest that greater maternal height only provides protection from SPTB-PPROM but not SPTB-IM. Chan et al [37] previously reported that Asian women of shorter stature were at a higher risk of preterm birth. Transgenerational reproductive adaptation, i.e. earlier birth to allow safe passage through a smaller pelvis has been suggested [38], while other explanations like women of shorter stature having a shorter cervix have been rejected [39].

While being born preterm has received recent recognition as a risk factor for developing hypertension as an adult [40], this is to our knowledge the first time that having mild hypertension (patients with severe hypertension requiring medication were excluded) has been identified as an independent risk factor for SPTB-PPROM with an odds ratio of 9.65 (95% CI 2.5–37.1). Interestingly a family history of recurrent gestational diabetes was associated with SPTB-PPROM, albeit with wide confidence intervals. It is tempting to speculate that the presence of the insulin resistance syndrome would explain these associations [41,42]. This may also explain the risk associated with hormonal fertility treatment, but again one would typically expect a clear association with the use of clomiphene; an association not demonstrable in this dataset.

It is difficult to explain why waking up during the night would be protective against SPTB-PPROM. Future studies on the full international SCOPE cohort of 5600 women may finally reveal whether this 'protective' factor represents a true finding. Similarly, inexplicable at this moment in time, appears to be the risk reduction associated with having a mother who had a miscarriage. Just as surprising was the finding of a doubling of risk associated with the index mother not being the first-born. Thinking of possible suboptimal placentation in the first pregnancy, one would anticipate the opposite.

Variables relating to dental health were only available in just over 30% of recruited women. In these women dental health, as assessed by several specific questions on easily bleeding gums, swollen gums, and sore teeth was no different between women with term birth and women with SPTB-IM or SPTB-PPROM. It should be noted that a recent systematic review [43] on periodontal disease came to an estimated odds ratio of 1.78 (CI 95%: 1.58, 2.01) for SPTB. Our negative findings regarding periodontal health and preterm labour could also be explained by the fact that self-assessed dental health by pregnant women is poorly associated with more objective markers as identified by a professional oral and dental examination [44].

A major strength of this study was its large multicentre prospective design with excellent follow-up. It should be noted that although the current study reports on a large very well defined prospective cohort of more than 3000 healthy nulliparous women, identification of risk factors in the current study risk factor was based on only 156 women with their pregnancies complicated by SPTB. To identify risk factors for very-early preterm birth, much larger prospective cohorts will be required.

Conclusion

The dissimilarity of clinical risk factors for SPTB-IM compared with SPTB-PPROM indicates different pathophysiological pathways underlie these distinct sub-phenotypes of spontaneous preterm birth. The ability to predict SPTB in healthy nulliparous women using clinical characteristics is modest. Given no reliable biomarkers have emerged as risk predictors of SPTB [45], the development of a clinically useful test will probably require SPTB phenotype-specific combinations of clinical risk factors and the discovery and evaluation of novel biomarkers.

Supporting Information

Checklist S1 STROBE Checklist.
(DOC)

File S1 Initial variable lists used to train multivariate models.
(DOC)

Author Contributions

Provided statistical analysis: SYL GAD CTR LMM RAN. Designed SCOPE database: RAN GAD LMM CTR. Conceived and designed the experiments: GAD SYL CTR. Performed the experiments: GAD SYL RAN LMM NABS CTR. Analyzed the data: GAD SYL RAN LMM NABS CTR. Contributed reagents/materials/analysis tools: GAD RAN LMM CTR. Wrote the paper: GAD SYL CTR.

References

1. Martin JA, Hamilton BE, Sutton PD, Ventura SJ, Mathews TJ, et al. (2010) Births: final data for 2007. Natl Vital Stat Rep. 58: 1–85.
2. Mathews TJ, MacDorman MF (2011) Infant mortality statistics from the 2007 period linked birth/infant death data set. Natl Vital Stat Rep. 59: 1–30.
3. Green N, Damus K, Simpson J, Iams J, Reece E, et al. (2005) Research agenda for preterm birth: recommendations from the March of Dimes. Am J Obstet Gynecol 193: 626–635.
4. Bardin C, Piuze G, Papageorgiou A (2004) Outcome at 5 years of age of SGA and AGA infants born less than 28 weeks of gestation. Semin Perinatol 28: 288–294.
5. Moster D, Lie R, Markestad T (2008) Long-term medical and social consequences of preterm birth. N Engl J Med 359: 262–273.
6. Iams J, Goldenberg R, Mercer B, Moawad A, Meis P, et al. (2001) The preterm prediction study: can low-risk women destined for spontaneous preterm birth be identified? Am J Obstet Gynecol 184: 652–655.
7. Fonseca E, Celik E, Parra M, Singh M, Nicolaides K (2007)Progesterone and the risk of preterm birth among women with a short cervix. N Engl J Med 357: 462–469.
8. Werner EF, Han CS, Pettker CM, Buhimschi CS, Copel JA, et al. (2011) Universal cervical-length screening to prevent preterm birth: a cost-effectiveness analysis. Ultrasound Obstet Gynecol.38(1): Epub 2011 May 24.

9. Gotsch F, Gotsch F, Romero R, Erez O, Vaisbuch E, et al. (2009) The preterm parturition syndrome and its implications for understanding the biology, risk assessment, diagnosis, treatment and prevention of preterm birth. J Matern Fetal Neonatal Med. 22 Suppl 2:5–23.
10. McCowan L, North R, Taylor R (2007) ACTRN12607000551493. Australian New Zealand Clinical Trials Registry. Available: www.anzctr.org.autrialSearch.aspx.
11. Spence M, Moss-Morris R, Chandler T (2005) The behavioural responses to illness questionnaire (BRIQ): a new predictive measure of medically unexplained symptoms following acute infection. Psychol Med 35: 583–93.
12. Groom KM, North RA, Stone PR, Chan EH, Taylor RS, et al. (2009) Patterns of change in uterine artery Doppler studies between 20 and 24 weeks of gestation and pregnancy outcomes. Obstet Gynecol 113: 332–8.
13. Berghella V, Bega G, Tolosa JE, Berghella M (2003) Ultrasound assessment of the cervix. Clin Obstet Gynecol 46: 947–962.
14. Gomez R, Galasso M, Romero R, Mazor M, Sorokin Y, et al. (1994) Ultrasonographic examination of the uterine cervix is better than cervical digital examination as a predictor of the likelihood of premature delivery in patients with preterm labor and intact membranes. Am J Obstet Gynecol 171: 956–964.
15. Akaike H (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6): 716–723.
16. Zweig MH, Campbell G (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem. 39: 561–77.
17. Kovo M, Schreiber L, Ben-Haroush A, Asalee L, Seadia S, et al. (2011) The placental factor in spontaneous preterm labor with and without premature rupture of membranes. J Perinat Med. 39: 423–9.
18. Romero R, Kusanovic JP, Chaiworapongsa T, Hassan SS (2011) Placental bed disorders in preterm labor, preterm PROM, spontaneous abortion and abruptio placentae. Best Pract Res Clin Obstet Gynaecol. 25: 313–27.
19. Arias F, Victoria A, Cho K, Kraus F (1997) Placental histology and clinical characteristics of patients with preterm premature rupture of membranes. Obstet Gynecol 89: 265–271.
20. Kho EM, McCowan LM, North RA, Roberts CT, Chan E, et al. (2009) SCOPE Consortium. Duration of sexual relationship and its effect on preeclampsia and small for gestational age perinatal outcome. J Reprod Immunol. 82: 66–73.
21. Wijesiriwardana A, Bhattacharya S, Shetty A, Smith N, Bhattacharya S (2006)Obstetric outcome in women with threatened miscarriage in the first trimester. Obstet Gynecol. 107: 557–62.
22. Hassan S, Romero R, Vidyadhari D, Fusey S, Baxter J, et al. (2011)Vaginal progesterone reduces the rate of preterm birth in women with a sonographic short cervix: a multicenter, randomized, double-blind, placebo controlled trial. 2011 Ultrasound Obstet Gynecol DOI:10.1002/uog.9017.
23. Romero R, Espinoza J, Kusanovic JP, Gotsch F, Hassan S, et al. (2006) The preterm parturition syndrome. BJOG.Suppl 3: 113 17–42.
24. Roy-Matton N, Moutquin JM, Brown C, Carrier N, Bell L (2011) The impact of perceived maternal stress and other psychosocial risk factors on pregnancy complications. J Obstet Gynaecol Can. 33: 344–52.
25. Moutquin JM (2003) Socio-economic and psychosocial factors in the management and prevention of preterm labour. BJOG. 110 Suppl 20: 56–60.
26. van Gelder MM, Reefhuis J, Caton AR, Werler MM, Druschel CM, et al. (2010) National Birth Defects Prevention Study. Characteristics of pregnant illicit drug users and associations between cannabis use and perinatal outcome in a population-based study. Drug Alcohol Depend. 109: 243–7.
27. Bada HS, Das A, Bauer CR, Shankaran S, Lester BM, et al. (2005) Low birth weight and preterm births: etiologic fraction attributable to prenatal drug exposure. J Perinatol 25: 631–7.
28. Shiono PH, Klebanoff MA, Nugent RP, Cotch MF, Wilkins DG, et al. (1995) The impact of cocaine and marijuana use on low birth weight and preterm birth: a multicenter study. Am J Obstet Gynecol. 172: 19–27.
29. Kho EM, North RA, Chan E, Stone PR, Dekker GA, et al. (2009) Changes in Doppler flow velocity waveforms and fetal size at 20 weeks gestation among cigarette smokers. BJOG. 116: 1300–1306.
30. Kyrklund-Blomberg NB, Cnattingius S (1998) Preterm birth and maternal smoking: risks related to gestational age and onset of delivery. Am J Obstet Gynecol. 179(4): 1051–5.
31. Henderson JJ, McWilliam OA, Newnham JP, Pennell CE (2012) Preterm birth aetiology 2004–2008. Maternal factors associated with three phenotypes: spontaneous preterm labour, preterm pre-labour rupture of membranes and medically indicated preterm birth. J Matern Fetal Neonatal Med. 25: 587–594.
32. Mason SM, Kaufman JS, Emch ME, Hogan VK, Savitz DA (2010) Ethnic density and preterm birth in African-, Caribbean-, and US-born non-Hispanic black populations in New York City. Am J Epidemiol. 172: 800–8.
33. Craig ED, Mitchell EA, Stewart AW, Mantell CD, Ekeroma AJ (2004) Ethnicity and birth outcome: New Zealand trends 1980–2001: Part 4. Pregnancy outcomes for European/other women. Aust N Z J Obstet Gynaecol. 44: 545–8.
34. Glover J, Hetzel D, Glover L, Page A, Leahy K (2005)Central Northern Adelaide Health Service: A social health atlas.Adelaide: The University of Adelaide.
35. Population health in South Australia (2005) Burden of disease and injury estimates, 1999–2001.South Australian Health.
36. Torloni MR, Betrán AP, Daher S, Widmer M, Dolan SM, et al. (2009) Maternal BMI and preterm birth: a systematic review of the literature with meta-analysis. J Matern Fetal Neonatal Med 22: 957–70.
37. Chan BC, Lao TT (2009) Maternal height and length of gestation: does this impact on preterm labour in Asian women? Aust N Z J Obstet Gynaecol. 49: 388–92.
38. Patel RR, Steer P, Doyle P, Little MP, Elliott P (2003) Does gestation vary by ethnic group? A London-based study of over 122 000 pregnancies with spontaneous onset of labour. Int J Epidemiol 33: 107–113.
39. Gagel CK, Rafael TJ, Berghella V (2010) Is short stature associated with short cervical length? Am J Perinatol 27: 691–5.
40. Alexander BT, Intapad S (2012) Preterm Birth: A Novel Risk Factor for Higher Blood Pressure in Later Life. Hypertension. 59: 189–190.
41. Palomba S, Falbo A, Russo T, Battista L, Tolino A, et al. (2010) Uterine blood flow in pregnant patients with polycystic ovary syndrome: relationships with clinical outcomes.BJOG. 117: 711–21.
42. Skarzinski G, Khamaisi M, Bursztyn M, Mekler J, Lan D, et al. (2009) Intrauterine growth restriction and shallower implantation site in rats with maternal hyperinsulinemia are associated with altered NOS expression. Placenta 30: 898–906.
43. Corbella S, Taschieri S, Francetti L, De Siena F, Del Fabbro M (2011) Periodontal disease as a risk factor for adverse pregnancy outcomes: a systematic review and meta-analysis of case-control studies. Odontology [Epub ahead of print].
44. Christensen LB, Jeppe-Jensen D, Petersen PE (2003) Self-reported gingival conditions and self-care in the oral health of Danish women during pregnancy. J Clin Periodontol. 30: 949–53.
45. Menon R, Torloni MR, Voltolini C, Torricelli M, Merialdi M, et al. (2011) Biomarkers of spontaneous preterm birth: an overview of the literature in the last four decades. Reprod Sci. 18: 1046–70.

IX. Paper III: Multivariate Visual Clustering

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. In *Proceedings of the 5th Annual Conference in Applied Statistics Education and Research Collaboration (ASEARC 2012)*. pp 56-59.