The incidence, risk factors and implications of type 1 diabetes: whole-of-population linked-data study of children in South Australia born from 1999-2013

Mumtaz Begum

BSc, MSc, MHSc

A thesis submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy

January 2020

School of Public Health

Faculty of Health and Medical Sciences

The University of Adelaide

Australia

Table of Contents

List of Tables

List of Figures

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

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

Sign Dated: 17/01/2020

Acknowledgements

I thank Almighty Allah for his countless blessings.

I acknowledge and pay my respect to the Kaurna people, the traditional custodians whose ancestral land I stayed on during my PhD candidature at The University of Adelaide. I respect the Kaurna people's deep feeling of attachment, relationship and ongoing connection with the country, and I respect and value their past, present, and their cultural beliefs.

I am grateful to my supervisors Associate Professor Lisa Smithers, Dr Catherine Chittleborough, and Dr Rhiannon Pilkington for their continuous guidance, support and encouragement throughout the PhD candidature. Thank you Lisa, Cathy and Rhiannon for teaching me a lot through your constructive comments, for listening to my challenges, and for your patience all these years. Thank you Professor John Lynch for your insightful comments on the papers. Thank you Dr Murthy Mittinty for your guidance with the statistical analyses. I also thank Dr Megan Penno for her critical input in the research papers.

I am thankful to the South Australian Early Childhood Data Project (SA ECDP) platform, for bringing together various administrative data of children in South Australia that enabled me to have access to de-identified individual-level information, to be able to conduct analyses for this doctoral thesis. I am thankful to the data custodians for providing de-identified data to the SA ECDP database that I used in my PhD project. I also thank SA-NT DataLink for the data linkage.

I am thankful to the BetterStart research group; especially Dr Angela Gialamas, Dr Janet Grant, and Alicia Montgomerie for being supportive throughout my PhD candidature. I am especially thankful to Alicia for answering my numerous questions regarding the datasets and Stata. Thank you Janet for the many communications with the data custodians regarding my PhD project.

To my parents Shamsu Wali khan and Pariyan kai; thank you Nan and Dada for your unconditional love. My mother passed away six months before I commenced my PhD candidature, perhaps that's why the PhD journey was emotionally challenging. Her long struggle with rheumatoid arthritis taught me strength and patience, the two important skills have helped me throughout life. Thank you dada for your constant emotional support, and for your belief in me that kept me motivated throughout PhD studies. Thank you Sanan for keeping me motivated, and for giving me joy each and every day of my life. Thank you Manzoor for always being there for our son, when I couldn't manage during PhD, and for the immense amount of care and support you have always given me, without which this PhD would not be possible. Thanks to Basharat and Haseen for always being my source of strength. Special thanks to Badan khan dada who taught me mathematics during school days, for giving me good foundation for learning. I am greatly thankful to Poqoon bechi (aunt) and Wazir meki (uncle); who both have speech and hearing impairment, for teaching me compassion. Thanks to Dada, Nani, Zulfi lal, Fatah, Wahab, Gule-lala, Zenab, Fatima, and Sajida, for their emotional support. I am thankful to my fellow PhD candidates (Sana, Mahlaka, Blesson, Dinesh, Engida and Magdalena) for the good time we spent together, for the guidance and peer support.

Thank you Adelaide Scholarship International for giving me the wonderful opportunity to study at The University of Adelaide.

I dedicate this doctoral thesis

to my beloved son Sanan and my parents

Shamsu Wali Khan and Pariyan Kai

Publications contributing to this thesis

- Begum M, Chittleborough C, Pilkington R, Mittinty M, Lynch J, Penno M, Smithers L. Incidence of type 1 diabetes by sociodemographic characteristics among South Australian children: whole-of-population study. (accepted for publication in the Journal of Paediatrics and Child Health during examination)

- Begum M, Pilkington R, Chittleborough C, Lynch J, Penno M, Smithers L. Caesarean section and risk of type 1 diabetes: whole-of-population study. Diabetic Medicine, December 2019, Vol.36 (12), pp.1686-1693. (Published)

- Begum M, Pilkington R, Chittleborough C, Lynch J, Penno M, Smithers L. Effect of maternal smoking during pregnancy on childhood type 1 diabetes: Whole-of-population study. *Diabetologia* 2020, Vol.63, pp.1162–1173. (Published)

- Begum M, Chittleborough C, Pilkington R, Mittinty M, Lynch J, Penno M, Smithers L. Educational outcomes among children with type 1 diabetes: whole-of-population linked-data study. Pediatric Diabetes: https://doi.org/10.1111/pedi.13107 (Published during examination, in press)

Presentations arising from this thesis

- Begum M, Chittleborough C, Pilkington R, Mittinty M, Lynch J, Penno M, Smithers L. Incidence of type 1 diabetes by sociodemographic characteristics among South Australian children: whole-of-population study. South Australian Population Health Conference, Adelaide, October, 2017.

- Begum M, Pilkington R, Chittleborough C, Lynch J, Penno M, Smithers L. Caesarean birth and risk of type 1 diabetes: whole-of-population study. Australasian Epidemiological Conference, Perth, October 2018.

- Begum M, Pilkington R, Chittleborough C, Lynch J, Penno M, Smithers L. Caesarean type and risk of type 1 diabetes: whole-of-population study. South Australian Population Health Conference, Adelaide, December, 2018.

- Begum M, Pilkington R, Chittleborough C, Lynch J, Penno M, Smithers L. Maternal smoking during pregnancy and risk of childhood type 1 diabetes: Whole-of-population study. Joint congress of SSM & European Congress of Epidemiology, Cork, Ireland, September 2019. Abstract Published, Journal of Epidemiology and Community Health, September 2019, Vol 73, and Supplement 1.

- Begum M, Chittleborough C, Pilkington R, Mittinty M, Lynch J, Penno M, Smithers L. Educational outcomes among children with type 1 diabetes: whole-of-population linked-data study. Joint congress of SSM & European Congress of Epidemiology, Cork, Ireland, September 2019. Abstract Published, Journal of Epidemiology and Community Health, September 2019, Vol 73, and Supplement 1.

Poster Presentations arising from this thesis

- Begum M, Chittleborough C, Pilkington R, Mittinty M, Lynch J, Penno M, Smithers L. Socioeconomic characteristics of South Australian children with type 1 diabetes, born from 2002-2013, Postgraduate Florey Conference, Adelaide, September, 2018.

- Begum M, Pilkington R, Chittleborough C, Lynch J, Penno M, Smithers L. Maternal smoking during pregnancy and risk of childhood type 1 diabetes: Whole-of-population study, Postgraduate Florey Conference, Adelaide, September, 2019.

Awards

The Walter and Dorothy Duncan Trust Conference Travel Grants, 2019

Abbreviations

AIPW	Augmented inverse probability weighting
ATE	Average treatment effect
ARIA	Australian Remoteness Index for Areas
BMI	Body Mass Index
CI	Confidence Interval
DKA	Diabetic ketoacidosis
HLA	The human leukocyte antigen
HR	Hazard ratio
ICD-10-AM	International Classification of Disease, Australian-modification, 10th edition
IRSD	Index of Relative Socioeconomic Disadvantage
IRSAD	The Index of Relative Socioeconomic Advantage and Disadvantage
MAR	Missing at random
MCAR	Missing completely at random
MNAR	Missing not at random

MICE	Multiple imputation by chained equations
NAPLAN	National Assessment Program-Literacy and Numeracy
SD	Standard deviation
SA	South Australia
SA ECDP	South Australian Early Childhood Data Project
T1D	Type 1 diabetes
WA	Western Australia

Abstract

The aim of this doctoral thesis was to study the incidence, risk factors and outcomes of type 1 diabetes for children in South Australia, born from 1999-2013. The incidence of type 1 diabetes has doubled in the last four decades in many countries including Australia, and has substantial individual and economic consequences. Evidence from studies on type 1 diabetes aetiology and its implications is mixed. In this thesis, the linkage of multiple population-wide administrative data over 15 years, and use of rigorous epidemiological approaches has resulted in a better understanding of the risk factors and implications of type 1 diabetes.

There are four studies in this doctoral thesis. In the first descriptive study, the incidence of type 1 diabetes was estimated by individual and area-level socioeconomic characteristics among children (aged ≤11 years) in South Australia, born from 2002-2013. Findings of the study showed that type 1 diabetes incidence rates differed depending on the measures of socioeconomic characteristics. Individual-level indicators showed higher type 1 diabetes incidence among more advantaged children, however, there was no clear area-level socioeconomic patterning of type 1 diabetes. Area-level measures of socioeconomic position are likely to have a greater risk of misclassification from true socioeconomic position, which suggests that the use of area-level measures may be misleading. Socioeconomic position is a major determinant of health and can modify the risk factors of type 1 diabetes. For example, as per hygiene hypothesis, the socioeconomically dis-advantaged children are less likely to have type 1 diabetes, which is supported by the findings of individual-level socioeconomic patterning of type1 diabetes in the first study. In addition, socioeconomically disadvantaged

women are less likely to have a caesarean birth and more likely to smoke in pregnancy. I chose to study these two risk factors of type 1 diabetes because the evidence was inconsistent, and some studies had methodical limitations.

Evidence about the effect of caesarean section on childhood type 1 diabetes is mixed; ranging from very small or no risk to 20-30% increased risk. A prevailing theory is that exposure to the gut and vaginal microbiota during a vaginal birth protects against type 1 diabetes. Therefore, in the second study, the impact of caesarean birth on childhood type 1 diabetes (aged ≤15 years) was estimated. This involved linking multiple administrative datasets of children in South Australia, born from 1999-2013. The question was extended to whether type 1 diabetes risk differed for children born by prelabour or intrapartum caesarean to further test the idea of microbiota exposure on type 1 diabetes. That is because children born by prelabour caesarean do not get exposure to maternal vaginal microbiota, and intrapartum caesarean births may have some exposure. Findings of the study obtained from Cox proportional hazard regression analysis showed a negligible 5% higher incidence (HR = 1.05, 95% CI 0.86-1.28) for caesarean births compared with normal vaginal delivery, with wide confidence intervals including the null. Contrary to the hypothesis of a higher type 1 diabetes risk for prelabor caesarean (because of non-exposure to maternal vaginal microbiota) type 1 diabetes risk for intrapartum caesarean was slightly higher (HR = 1.08, 95% CI 0.82-1.41) than prelabor caesarean (HR = 1.02, 95% CI 0.79-1.32). This negligible risk of type 1 diabetes for children who had caesarean birth, either prelabor or intrapartum, and the potential for unmeasured confounding suggested that birth method induced variation in neonatal microbiota might not be involved in modifying type 1 diabetes risk.

Like caesarean section, maternal smoking in pregnancy is also a debated risk factor for childhood type 1 diabetes. Evidence about maternal smoking on childhood type 1 diabetes is inconsistent; studies have been small, and many did not adjust for important confounders or address missing data. In the third study of this doctoral thesis, the effect of maternal smoking in pregnancy on childhood type 1 diabetes was estimated using Cox proportional hazard regression analysis, once again by linking multiple administrative datasets of children in South Australia, born from 1999-2013. The analytical approach for this study ranged; from Cox proportional hazard analysis with adjustment for wide range of confounders using the SA ECDP linked data, involving multiple imputation for missing data; to conducting meta-analysis in order to get more precise estimate. But smoking is notoriously residually confounded, therefore, I made special efforts to investigate the possibility of residual confounding by using a negative control and E-value. The findings demonstrated that maternal smoking in pregnancy was associated with a 16% (HR 0.84, 95% CI 0.67, 1.08) lower childhood type 1 diabetes incidence, compared with unexposed children, which was also supported by the meta-analytic estimates of population-based cohort studies (HR 0.72, 95% CI 0.62, 0.82) and case-control studies (OR 0.71, 95% CI 0.55, 0.86). The negative control outcome and E-value analyses indicated the potential for residual confounding in the effect of maternal smoking on childhood type 1 diabetes. Triangulation of evidence from this study along with the results of similar population-based studies, suggested a small reduced risk of childhood type 1 diabetes for children exposed to maternal smoking in pregnancy. However, the mechanisms linking maternal smoking in pregnancy with childhood type 1 diabetes require further investigation.

In the fourth study of this thesis, the impact of childhood type 1 diabetes on children's educational outcomes in year/grade 5 at age ~10 were estimated, linking population-wide data of children in South Australia, born from 1999-2005. In this study, a doubly-robust analytical method called augmented inverse probability weighting (AIPW) was used to compute the average treatment effect of type 1 diabetes on children's educational outcomes. AIPW gives an unbiased estimate if either the outcome model or the treatment model is correctly specified. The findings of this study demonstrated that children with type 1 diabetes are not disadvantaged in terms of educational outcomes in year 5, potentially reflecting improvement in type 1 diabetes management in Australia.

In summary, the work in this doctoral thesis has demonstrated that type 1 diabetes incidence differed depending on the measure of socioeconomic position. The hygiene hypothesis was only supported by the individual-level socioeconomic pattering of type 1 diabetes incidence in South Australia. The involvement of birth method induced variation in neonatal microbiota in type 1 diabetes was not supported by the caesarean and childhood type 1 diabetes study. Despite the evidence of residual confounding in the estimate of maternal smoking in pregnancy on childhood type 1 diabetes, triangulation of the evidence suggested small reduced risk for children exposed to maternal smoking in pregnancy, but further research will be needed to understand the mechanism. The findings of similar educational outcomes for children with and without type 1 diabetes, highlighted the importance of improvements in diabetes management.

CHAPTER 1 INTRODUCTION

Type 1 diabetes is a chronic childhood disease of an autoimmune origin, caused by the destruction of insulin producing pancreatic β-cells.[1] The resulting severe insulin deficiency impairs the cellular uptake and utilization of glucose, and leads to an increased blood glucose concentration and hyperglycaemia. Uncontrolled hyperglycaemia increases the risk of microvascular (diabetic retinopathy, renal disease, peripheral neuropathy and amputation) and macrovascular complications (stroke, myocardial infarction, cardiac failure and peripheral vascular diseases).[2, 3, 4] When the body is unable to metabolize glucose for energy, lipolysis occurs, leading to the production and accumulation of ketone bodies in the blood, a condition called ketoacidosis.[5] Both hyperglycaemia and ketoacidosis have consequences for children's health and development.

In the last four decades the incidence of type 1 diabetes has doubled in many countries in the world including Australia. Genetics alone cannot explain the increasing incidence. The role of the environment in type 1 diabetes pathogenesis is suggested by rising type 1 diabetes incidence[6, 7] and global variation in type 1 diabetes incidence (1.01-60.9 per 100,000 person-years).[3, 6-9] In addition, the disparity in type 1 diabetes incidence in genetically similar European populations with different socio-economic characteristics,[10] and discordance in type 1 diabetes incidence among monozygotic twins,[11] also support the involvement of environmental in type 1 diabetes pathogenies.

Evidence suggests the environment is playing an important role in driving the increasing type 1 diabetes incidence. However, identifying the exact cause(s) that initiate the autoimmune process leading to type 1 diabetes remains elusive.

1.1 Why study the risk factors and implications of type 1 diabetes?

In my doctoral thesis, I studied the incidence, selected risk factors (caesarean birth and maternal smoking in pregnancy) and implications of type 1 diabetes for children's educational outcomes in South Australia. There were four reasons to study these risk factors and implication of type 1 diabetes; firstly the increasing incidence of type 1 diabetes and its impact for children's health and development, the psychosocial burden and economic implications associated with type 1 diabetes care and management. Children with type 1 diabetes are three times more likely to be hospitalized per person-year,[12] have 3% lower school attendance per-year,[13] and 19% higher mental health referral rates than children without type 1 diabetes.[14] People with type 1 diabetes have an average of 12 years lower life expectancy than the general population.[15] The annual estimated health care cost of type 1 diabetes in Australia was $570 million for 2008 and hospitalization accounted for 47% of this cost.[16, 17]

Secondly at population level apart from rising type 1 diabetes incidence,[6] caesareans birth rates are also on the rise,[18] while smoking rates are declining.[19] However, the evidence about the effect of caesarean birth[20-25] and maternal smoking in pregnancy on childhood type 1 diabetes,[26-34] and the implications of type 1 diabetes on educational outcomes are inconsistent;[13, 35-37] it is

not clear if either of these are causal. Women cannot be randomised to have caesarean birth or to smoke in pregnancy; these are such confounded variables, that it is difficult to study these risk factors as *causes.* Therefore, there is a very opportunity to apply some innovative causal methods to try to understand these risk factors and implications of type 1 diabetes on educational outcomes. In addition, large studies are needed for power since type 1 diabetes is rare disease.

Thirdly; most of the findings came from case-control studies, expect few population-based cohort studies, and the whole-of-population linked data platform granted the opportunity to look at these risk factors in a whole-of-population sample rather than a case-control study. Lastly; there were methodological limitations in most previous studies; there is inconsistency in evidence about these risk factors and the implication of type 1 diabetes, due to variability in study designs, small numbers, and unavailability of individual level-socioeconomic information or not adjusting for important confounders putting some estimates at risk of confounding bias. In the South Australian Early Childhood Data Project (SA ECDP) that I have used in my doctoral dissertation, information on individual and areal-level socioeconomic factors, caesarean birth and maternal smoking in pregnancy along with a wide range of other variables (potential confounders) has been collected using validated forms as part of the perinatal statistics (detail in chapter 3). Multiple population-wide datasets (hospitalization, perinatal and births, and school assessment) have been linked in the SA ECDP database, making it feasible to study the risk factors and implications of type 1 diabetes, in an efficient way using cutting-edge epidemiological methods.

1.2 Thesis aim and research questions

The overarching aim of my doctoral project was to study the incidence, risk factors and implications of type 1 diabetes. The four specific question addressed in this thesis included;

1. What are the socioeconomic characteristics of children with type 1 diabetes in South Australia born from 2002-2013?
2. What is the effect of caesarean birth on childhood type 1 diabetes? Does the risk of type 1 diabetes differ by prelabour and intrapartum caesarean?
3. What is the effect of maternal smoking in pregnancy on the risk of childhood type 1 diabetes? What is the risk of bias due to residual confounding in the effect estimate?
4. What is the impact of type 1 diabetes on children's educational outcomes?

1.3 Thesis outline

This doctoral dissertation is organized as follows.

In **Chapter 2** I reviewed the existing literature about the incidence, genetic and environmental risk factors and the hypothesized mechanisms, and burden of type 1 diabetes. The focus then moved to the specific research areas to be explored in my thesis; the socioeconomic patterning of type 1 diabetes; caesarean birth, maternal smoking in pregnancy, and type 1 diabetes; and finally the impact of type 1 diabetes for children's educational outcomes. The purpose of this

review was to find the gaps in the existing literature about the risk factors and implication of type 1 diabetes, and to be able to identify potential confounders for the above mentioned research questions.

In **Chapter 3** I have described; the study population and design, the multiple population-wide administrative datasets used in my doctoral research; including birth registration data, perinatal statistics, hospitalisation data, school enrolment census and school assessment data. It also describes the data linkage process; the epidemiological methods used, which includes directed acyclic graphs (DAGs) for identifying confounding; augmented inverse probability weighting; methods to estimate unmeasured confounding such as negative control outcome analysis and E-value; and the technique used to deal with missing data (multiple imputation). The analytical approaches used in each of the four research questions are also discussed in greater depth than able to be included in each of the papers.

Chapter 4 is a submitted paper in the Journal of Paediatrics and Child Health (accepted for publication during examination) and addressed the first research question of this doctoral dissertation. In this descriptive paper, I estimated the incidence rate of type 1 diabetes by area-level and individual-level socioeconomic characteristics, for children in South Australia born from 2002-2013. I used linked hospitalization, birth registration and perinatal statistics datasets to explore the association between socioeconomic position and the type 1 diabetes incidence in childhood.

Chapter 5 is a published academic paper in Diabetic Medicine. In this study I explored the effect of caesarean birth on childhood type 1 diabetes using Cox proportional hazard regression,

extending the question to whether type 1 diabetes risk differed by prelabour or intrapartum caesarean. Multiple linked population-wide datasets (hospital, births, and perinatal characteristics) of children in South Australia, born from 1999-2013 were used for this study.

Chapter 6 is an academic paper, accepted for publication in Diabetologia (published during examination). This was the most extensive work from my doctoral dissertation. Firstly, I estimated the association between maternal smoking in pregnancy and childhood type 1 diabetes using Cox proportional hazard regression using multiple linked population-wide datasets of children in South Australia, born from 1999-2013. Secondly, I assessed the risk of bias due to residual confounding with a negative-control outcome design and an E-value analysis; and finally in order to get more precise estimate of the association between maternal smoking in pregnancy and type 1 diabetes, I conducted a meta-analysis of published population-based and case-control studies, along with my own findings.

Chapter 7 is an academic paper published in Pediatric Diabetes during examination. In this paper I estimated the effect of type 1 diabetes on children's educational outcomes, compared to children without type 1 diabetes, using multiple population-wide administrative datasets of children in South Australia. Here I apply the potential outcomes approach to compute the average treatment effect of type 1 diabetes on children's educational outcomes in year/grade 5 (age ~ 10 years), by using augmented inverse probability weighting.

Chapter 8 summarizes the key findings; and discusses the contribution of this doctoral research, the variability in effect estimates, the limitations and potential areas for future research.

CHAPTER 2 BACKGROUND

2.1 Preface

In this chapter I have reviewed the concepts of type 1 diabetes prevalence and incidence; the genetics and common environmental risk factors; the proposed hypotheses for the increasing type 1 diabetes incidence; and the burden of type 1 diabetes. In addition, I also reviewed the existing literature around the main areas I have explored in my doctoral dissertation, including the risk factors (socioeconomic position, caesarean birth, maternal smoking in pregnancy) and implications of type 1 diabetes for children's educational outcomes.

2.2 What is type 1 diabetes?

The autoimmune-mediated destruction of pancreatic beta (β) cells in type 1 diabetes results in impaired insulin production. The progressive loss of β-cell function causes absolute insulin deficiency, making people dependent on the administration of exogenous insulin for survival.[2, 3] Insulin is an essential hormone for the human body to regulate cellular uptake and utilization of glucose.[2, 38] Without insulin to allow glucose to enter cells, blood glucose builds up in circulation resulting in hyperglycaemia. Due to severe insulin deficiency the body is unable to metabolize glucose for energy and consequently the breakdown of fatty acids (lipolysis) occurs, leading to ketogenesis (production of ketone bodies).[5] The accumulation of these acidic ketone

bodies results in ketoacidosis.[5] Persistent hyperglycaemia can cause microvascular (affecting capillaries and small blood vessels) and macrovascular (affecting large arteries) complications. Microvascular complications include diabetic retinopathy, renal disease, peripheral neuropathy and amputation. Macrovascular complications are stroke, myocardial infarction, cardiac failure and peripheral vascular diseases.[2, 3 4]

The risk of type 1 diabetes can be identified in the asymptomatic stage by genetic susceptibility, appearance of autoantibodies, and dysglycemia.[38, 39] While clinically asymptomatic (stage 0 to stage 2) type 1 diabetes autoantibodies start appearing,[40] but the β-cell mass remains intact.[1, 41] Long-term risks associated with type 1 diabetes are higher for children who are positive for one autoantibody than people without autoantibodies, however less than 10% of people who are positive for a single autoantibody develop clinical type 1 diabetes over 10 years.[1, 42] The risk of type 1 diabetes is over 90% in first-degree relatives of patients who are positive for at least two autoantibodies, whereas risk is less than 20% in relatives who are positive for one autoantibody.[2]

The rate of progression from autoantibody seroconversion to overt type 1 diabetes varies from a few months to many years.[40, 43] Islet autoimmunity starts many months and years before type 1 diabetes diagnosis, and 90% of children who are diagnosed with type 1 diabetes before puberty have islet autoantibodies by age 5 years.[41] Longitudinal follow-up studies among genetically at risk children have shown that islet autoantibodies appear in the second half of infancy[43] peaking at around 9-24 months,[43, 44 45] but are rarely detected before the age of 6 months.[45] Remittance and relapse of autoantibodies can occur before the clinical onset of type 1 diabetes.[43-45] The

accelerated progression from islet autoimmunity to clinical diabetes is proposed to have caused the escalation in type 1 diabetes incidence among children.[46]

There are three stages in the development of type 1 diabetes. Appearance of at least two islet autoantibodies marks stage 1 of type 1 diabetes.[1, 41, 42] Stage 2 is characterized by two positive autoantibodies, dysglycaemia, the start of decline in pancreatic β-cell mass, and this is accompanied by a five year risk of 75% for developing type 1 diabetes.[1, 42] Stage 3 is the clinically symptomatic phase of type 1 diabetes and is characterized by the presence of at least two autoantibodies, hyperglycaemia, and 10-20% loss of the pancreatic β-cell mass.[1, 42] In addition to the appearance of autoantibodies, widely accepted criteria for diagnosis of clinical type 1 diabetes is fasting blood glucose levels ≥126 mg/dl or ≥7 mmol/L and oral glucose tolerance test ≥200 mg/dl or ≥11.1 mmol/L.[38, 47] Type 1 diabetes stages are important for screening and identifying at risk individuals for potential interventions in order to delay the onset of clinical type 1 diabetes. Most importantly, identification of individuals, particularly children, at the asymptomatic stage, will help prevent severe hyperglycaemia and diabetic ketoacidosis and the long-term consequences for children's health and development.

Type 1 diabetes can be diagnosed at any age, however, most people with type 1 diabetes are diagnosed in childhood.[48, 49] Type 1 diabetes accounts for ≥85% of all diabetes in children aged <20 years in the world.[50] In Australia, 61% of newly diagnosed type 1 diabetes cases were made for people less than 25 years of age in 2016.[51]

2.3 Prevalence of type 1 diabetes

According to the 2017 International Diabetes Federation report, globally around 1.1 million children aged <19 years have type 1 diabetes and 132,600 new cases are diagnosed every year.[52] In developed countries approximately 7-9% of all diabetes cases are type 1 diabetes, and 87-91% are type 2 diabetes cases.[52] The proportion of diabetes cases that are type 1 and type 2 is not clear in middle and low income countries, due to a lack of data. In Australia, the National Diabetes Register data demonstrated that 6,091 children aged 0-14 had type 1 diabetes in 2013, equating to a prevalence of 139 cases per 100,000.[53] Within Australian jurisdictions, South Australia had the second highest prevalence of type 1 diabetes (159 per 100,000), with the highest being in Tasmania (166 per 100,000) and the lowest in the Northern Territory (50 per 100,000).[53]

2.4 Incidence of type 1 diabetes

2.4.1 Trends and variation in global type 1 diabetes incidence

There are regional and country-level variations in type 1 diabetes incidence.[3] Current evidence suggests that type 1 diabetes incidence ranges from 1.01 per 100,000 person-years in China[8] to 60.9 per 100,000 person-years in Finland.[9] Europe and North America have the highest type 1

diabetes incidence, and the Western Pacific region and Africa have the lowest incidence.[54] Type 1 diabetes incidence has almost doubled in most European countries in the last three decades.[7] For example, in Finland, type 1 diabetes incidence was 38.2 per 100,000 person-years from 1989-1993, and has reached 60.9 per 100,000 person-years in 2009-2013.[7] The regional and country-level variations in type 1 diabetes incidence depict the wide distribution of type 1 diabetes related genes, and different environmental factors in each country that could be playing a vital role in influencing type 1 diabetes pathogenesis.

There is consensus in evidence that childhood type 1 diabetes incidence has been rising over the last four decades in many countries.[3, 6, 7, 55] However, some recent studies have reported stabilization in type 1 diabetes incidence,[9, 49, 56] while others depicted a sinusoidal pattern in type 1 diabetes incidence rates.[57] Until now the most comprehensive global evidence (data from 112 centres around the world) of type 1 diabetes incidence comes from the World Health Organization's DIAMOND study that demonstrated a 2.8% average annual increase in type 1 diabetes incidence from 1990-1999, among children aged ≤14 years.[6] More recent evidence from 26 European centres demonstrated a 3.4% average annual increase in type 1 diabetes incidence from 1989-2013, with a range of 0.5% in Spain−Catalonia to 6.6% average annual increase in Poland-Katowice, among 0-14 year olds.[7] Some studies suggest that previously type 1 diabetes incidence was rising rapidly among younger children (0-4, 5-9 years),[58] although this has recently shifted to older age groups (10-14 years).[9, 59]

Most type 1 diabetes incidence data comes from industrialized countries with a high Human Development Index. However, it cannot be concluded that type 1 diabetes incidence is only

increasing in high income countries, due to the lack of population-based data from middle- and low-income countries.

2.4.2 Type 1 diabetes incidence in Australia

According to the Australian Institute of Health and Welfare, in 2017 about 2,742 people were diagnosed with type 1 diabetes (all age groups), which equates to an incidence of 12 cases per 100,000 population, and 9% of all diabetes cases.[51] About 1,114 or 41% of new type 1 diabetes cases diagnosed in 2017 were amongst children 0-14 years old, equating to an incidence of 24 per 100,000 population for this age group.[51] Australia ranks ninth in the world for published type 1 diabetes incidence among children aged under 15 years (Figure 2-1).[54]

Table 4 – Top 10 countries for published type 1 diabetes incidence rate (a) and estimated prevalent cases (b) in the under 15-year age-group.

Rank	Country	(a) Incidence rate (per 100,000 population aged under 15 year)	Rank	Country	(b) Estimated new cases (1000s)
1	Finland	57.6	1	United States of America	13.0
2	Sweden	43.1	2	India	10.9
3	Norway	32.8	3	Brazil	5.0
4	Saudi Arabia	31.4	4	United Kingdom	3.1
5	United Kingdom	28.2	5	Russian Federation	2.6
6	Canada	25.9	6	Saudi Arabia	2.6
7	Denmark	25.1	7	Germany	2.4
8	United States of America	23.7	8	Nigeria	2.2
9	Australia	22.5	9	Mexico	2.2
10	Kuwait	22.3	10	Egypt	2.0

Figure 2-1: Top 10 countries with published type 1 diabetes incidence

Source: Patterson *et al.* Diabetes research and clinical practice (2014)[54]

Australia experienced increasing trends in type 1 diabetes incidence a decade ago, but more recent evidence has depicted some stabilization in type 1 diabetes incidence. A population-based register study reported that type 1 diabetes incidence was 13.6 per 100,000 population per year, among 0-14 year olds in 1984, in Sydney.[60] Similar type 1 diabetes incidence (13.2 per 100,000 person-year) was observed in Western Australia from 1985-1989.[61] Since then type 1 diabetes incidence has almost doubled among 0-14 year olds in Australia (24 per 100,000 population, 2000-2013).[57, 62] Further evidence of increasing incidence is seen in jurisdiction specific studies. An average 3.3% annual increase in type 1 diabetes incidence was observed from 1985-2003 in Western Australia (WA),[63] and a 2.8% average annual increase was reported in New South Wales from 1990-2002.[64]

As mentioned following the increase in incidence, a plateauing trend in type 1 diabetes incidence has been observed over the last decade in many countries[7] including Norway,[56] Finland,[9] Sweden[59] and Australia.[63] A study in Western Australia (WA) showed an average 3.3% per year increase in type 1 diabetes incidence from 1985 to 2003, however, no significant change in the temporal trend were observed (−0.6% per year) from 2003 to 2016.[63] At a national level, the Australian National Diabetes Register data showed a sinusoidal cyclic pattern (five-year cyclic variation) in childhood type 1 diabetes incidence from 2000-2011, and an average annual increase of 1.2% was observed only among 10-14 year old children.[57, 65] Thus, the increasing trend in type 1 diabetes incidence is showing signs of slowing down, or plateauing in many countries including Australia.

2.4.3 Ethnic variation in type 1 diabetes incidence

Variation in childhood type 1 diabetes incidence has been observed across different ethnic groups within Australia,[62] Norway,[66] Sweden,[67] and the US.[32] In Australia type 1 diabetes incidence is lower among Aboriginal and Torres Strait Islander people than non-Indigenous Australians.[62] In Norway, lower type 1 diabetes incidence was observed among children from other backgrounds (based on mother's country of birth) than the Norwegian children.[66] Similar ethnic differences in type 1 diabetes incidence were observed in the US (higher type 1 diabetes incidence in white Americans than other ethnicities),[32] and Sweden (higher type 1 diabetes incidence in native Swedes than second-generation immigrants, except Finnish immigrants).[67] These findings suggest that Caucasians have higher type 1 diabetes incidence than other ethnic groups in Europe, US and Australia. Potential reasons for these ethnic differences in type 1 diabetes incidence could be genetic variation, differences in socioeconomic conditions, lifestyle, eating and dietary behaviours among these groups.

2.4.4 Type 1 diabetes incidence by gender and age

There is some suggestion of disparities in type 1 diabetes incidence by gender and age-group, however the evidence is mixed. In most countries, including Italy,[68] Finland,[9] and Sweden,[59] the incidence of type 1 diabetes is higher among males than females, but no gender difference has

been reported in Australia[63] and Poland.[69] For example, in Finland from 2006 to 2011, the incidence of type 1 diabetes was 68 per 100,000 person-years for boys and 55 per 100,000 person-years for girls.[9] Countries appear to differ at the age when the type 1 diabetes incidence peaks. In most countries, including Sweden,[59] Norway,[66] US,[70] Northern Ireland,[71] and Australia,[57, 63, 64] the highest type 1 diabetes incidence has been observed among 10-14 year olds. Gender differences in peak age of diagnosis have also been reported. For example, in Sweden[72] and Australia[63] type 1 diabetes incidence rate has been observed to be highest for males aged 10-14 years and females aged 5-9 years. This suggest that peak age of diagnosis happens earlier for girls than boys, at least in some countries.

2.4.5 Urban-rural variation in type 1 diabetes incidence

Urban-rural disparities have also been observed in type 1 diabetes incidence. Based on the remoteness index in Australia, high type 1 diabetes incidence was observed in more accessible areas[53] or urban[73] areas than remote and very remote areas, but no urban-rural difference was found in Poland.[69] On the contrary, a higher type 1 diabetes incidence was observed in remote areas in Northern Ireland,[71] and rural areas in Finland[74] compared to more accessible and urban areas. The differences in individual socioeconomic conditions, ethnicity of the residents, and household conditions in remote and accessible areas could play a role in the variation in type 1 diabetes incidence across urban-rural areas.

2.5 Genetic and familial risk of type 1 diabetes

In terms of genetic susceptibility, more than 61 gene variants have been linked with the risk of type 1 diabetes.[75] In particular, the human leukocyte antigen (HLA) class II genes are the most important loci [76-80] and account for up to 50% of genetic risk of type 1 diabetes.[77, 80] HLA genes have many different alleles that modify the adaptive immune system, and help the body in distinguishing between the body's own protein and foreign proteins or pathogens. HLA genes are categorized into class I, class II, and class III. The HLA class II genes are the most important alleles for type 1 diabetes pathogenesis, and has three major (HLA-DP, HLA-DQ and HLA-DR) and two minor histocompatibility complex proteins (MHC class II DM and DO).[80, 81] The HLA DR-DQ haplotypes confer the highest risk for type 1 diabetes and are known as DR3 and DR4.[80, 81 82, 83] Some haplotypes such as DR2 confer protection for diabetes.[81] In addition, some HLA class 1 and class III haplotypes are also reported to be involved in type 1 diabetes pathogenesis.[81] The frequencies of HLA allele and haplotype differs across populations in different countries and this may partly explain the variation in type 1 diabetes incidence globally. Only less than 10% of people with HLA-conferred susceptibility develop type 1 diabetes. Non-HLA genetic variants also contribute to type 1 diabetes risk.[75, 77, 80, 84] For example, the insulin gene on chromosome 11p15 has shown to confer about 10% of the genetic susceptibility to type 1 diabetes.[84] The combination of more than two different haplotypes further escalates the risk.[85] Genetic risk scores consisting of multiple loci can predict more than 10% risk of pre-symptomatic type 1 diabetes for children without family history.[86] Some environmental factors such as maternal smoking in pregnancy are thought to impact genetic expression through

epigenetic modification or DNA methylation,[87] and might have a role in type 1 diabetes pathogenesis. However, this is a very new area of investigation and further work is needed to understand the epigenetic processes.

Having a first degree relative with type 1 diabetes is linked with 10-15 fold increased lifetime risk of type 1 diabetes than general population.[28, 41, 88, 89] However, about 78-85% of new cases are being diagnosed without having any family history of type 1 diabetes,[77, 90] suggesting an influence of environment in the pathogenesis of type 1 diabetes. The type 1 diabetes prevalence is ~0.3% in the general population by age 20 years, and ~5% among those who have a first-degree relative with type 1 diabetes.[41] A study using Finnish Diabetes Register data (type 1 diabetes n = 1,488), established to characterize the familial and sporadic cases of newly diagnosed type 1 diabetes, demonstrated that 21.8% of newly diagnosed cases had a first and second degree relative affected with type 1 diabetes, and 78.2% were sporadic cases.[90] About 12.2% of the newly diagnosed cases had a first degree relative with type 1 diabetes, out of which 6.2% had an affected father, 3.2% had an affected mother, and 4.8% cases had a sibling with type 1 diabetes.[90] Among first-degree relatives, siblings of children with type 1 diabetes have a higher risk of type 1 diabetes than parents.[89] Compared with mothers, a father having type 1 diabetes is a stronger predictor of a child having type 1 diabetes.[21, 28, 89] The very high proportion of sporadic cases without having any first-and-second degree relatives with type 1 diabetes or family history points to the environment as playing an important role in increasing the type 1 diabetes incidence. A study that analysed HLA class II genotype frequencies over time (1965-2006) in two large populations with type 1 diabetes (diagnosed at aged ≤18 years) demonstrated a stepwise decrease in the highest risk HLA genotypes (HLA-DR3/4-DQB1*0302) in new onset

type 1 diabetes cases (Figure 2-2).[91] This reduction in high risk genotypes over four decades suggests an increasing influence of environmental factors in type 1 diabetes pathogenesis.

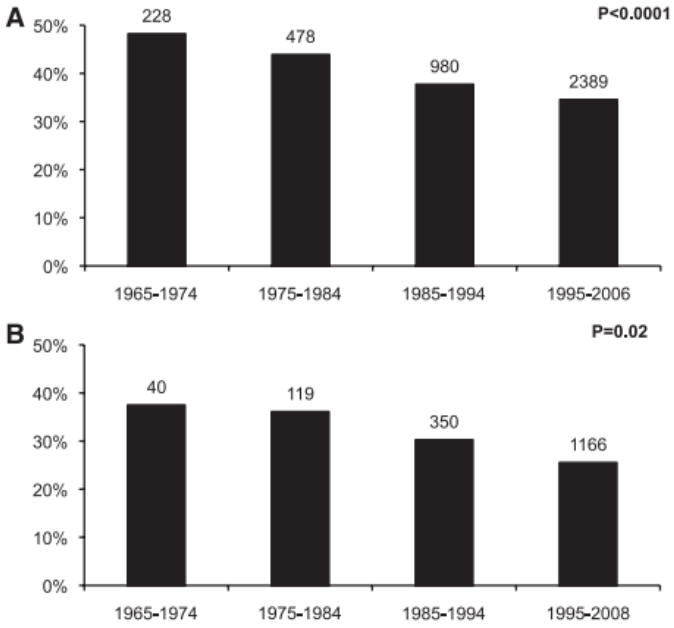


Figure 2-2: Percent HLA-DR3/4-DQB1*0302 in new onsets type 1 diabetes over time

(A stepwise decrease in the highest risk HLA genotypes frequencies in new onset type 1 diabetes cases over time from 1965-2006)

Source: Steck et al. Diabetes (2011)[91]

2.6 Potential mechanisms explaining environmental determinants

Although research suggests that environmental factors might be interacting with genetic predisposition in altering type 1 diabetes risk, the cause of the initiation of the autoimmune process of type 1 diabetes remains unknown. Many hypotheses have been proposed as potential explanations for the rising type 1 diabetes incidence. I will discuss the most commonly proposed hypotheses here; the hygiene hypothesis and β-cell stress hypothesis and microbiota composition.

2.6.1 Hygiene hypothesis and type 1 diabetes

The hygiene hypothesis[92-94] evolved from epidemiological observations that reported low type 1 diabetes incidence in areas with overcrowded households[95, 96] and high population density;[97] although the evidence about area-level variation in type 1 diabetes incidence is inconsistent.[98, 99] In the early 20th century type 1 diabetes was a rare but fatal disease, particularly in the pre-insulin era, and insulin discovery (1922) made prolonged survival possible for people with type 1 diabetes.[55] Studies from different European countries (Denmark, Sardinia area of Italy, and Norway) reported increasing trends in type 1 diabetes incidence from the mid-20th century.[55, 100-102] The 'Epidemiological transition' is a well-known phenomenon, one aspect of which is a decline in infectious diseases in the last century (as a cause of mortality) and a substantial increase in non-communicable diseases.[103 104] The decline in infections over time is attributed

in part to improvements in living standards, socioeconomic conditions, sanitation, and hygiene. With the decline in infectious diseases over time there has been a concomitant rise in the incidence of immune-mediated diseases such as type 1 diabetes,[105] theorized to be due to under-stimulation of the immune system.[94, 106] David P. Strachan, a professor of epidemiology, is known as one of the early researchers to point to the hygiene concept as an explanation for increasing incidence in hay fever,[107] and he wrote that, "*over the past century declining family size, improvements in household amenities, and higher standards of personal cleanliness have reduced the opportunity for cross infection in young families.*"[107] Researchers in type 1 diabetes epidemiology have been suggesting similar concepts termed as the hygiene hypothesis as one of the potential explanations for the rise in type 1 diabetes incidence.[32, 92, 93] The hygiene hypothesis proposes that the increasing type 1 diabetes incidence is attributed to reduced stimulation of the immune system by lack of exposure to microbial antigens postnatally and in early childhood and decreased frequency of childhood infections, perhaps due to improved living conditions and change in human environment and hygienic measures.[92, 108] Gale writes "*something protective might have been lost from the childhood environment over recent decades, a concept known as the hygiene hypothesis.*"[93]

In line with the hygiene hypothesis, regions with the lowest frequency of tuberculosis and diarrhoeal diseases, such as Northern Europe, North America and Australia, have the highest incidence rates of type 1 diabetes.[92] Improved sanitation and hygiene may also have caused a reduction in herd immunity for enterovirus infection in pregnancy,[109] thus increasing the risk of prenatal enterovirus infections. For example, over 20 years in Finland and Sweden the enterovirus autoantibodies have declined in pregnant women and childhood type 1 diabetes

incidence has substantially increased.[109] Most studies suggest that maternal enterovirus infections are linked with higher risk of childhood type 1 diabetes, but the evidence is inconsistent.[94, 110, 111] The Environmental Determinants of Diabetes in the Young (TEDDY) study reported that other gestational infections were not linked with islet autoimmunity but maternal respiratory infections were associated with reduced islet autoimmunity in genetically at risk children.[112]

Epidemiological evidence for the hygiene hypothesis of type 1 diabetes is inconsistent. For example, a large population-based study from Germany showed that recurrent viral respiratory tract infections in the first six months of life were associated with 17% increased risk of type 1 diabetes at the age of 8 years, compared with unexposed children.[113] On the other hand, a matched case-control study from the UK reported no association between early life infections and risk of childhood type 1 diabetes, hence not supporting the hygiene hypothesis.[114] Most studies that reported increased risk of type 1 diabetes for children exposed to respiratory infections were based on parental reporting about the frequency of childhood respiratory infections.[82, 115] There is lack of consensus about the link between early life infections and childhood type 1 diabetes.[113, 114, 116] It is hard to determine whether the associations between early infections and type 1 diabetes were due to increased exposure to viruses or underlying susceptibility to viral infections due to immune response dysregulation.[82] Despite the mixed evidence, the hygiene hypothesis continues to be discussed as an environmental cause and potential explanation for the increasing type 1 diabetes incidence.

2.6.2 Accelerator and β-cell stress hypothesis

Another potential explanation for increasing type 1 diabetes incidence is the accelerator and β-cell stress hypothesis. This hypothesis suggests that many environmental factors leading to fast growth and overweight in early life could cause insulin resistance and consequently damage pancreatic β-cells.[117] When insulin resistance is present, blood glucose rises leading to glucotoxic effects, which in turn may impact insulin producing pancreatic β-cells (β-cell apoptosis),[118] or may induce neo-autoantigens and initiate autoimmune-mediated destruction of the pancreatic β-cells. According to the accelerator hypothesis three processes accelerate the loss of pancreatic β-cell through apoptosis; 1) constitution (intrinsic), 2) insulin resistance, and 3) autoimmunity.[119] The first accelerator is the intrinsically high rate of apoptosis of the pancreatic β-cells, this step is necessary for developing diabetes, however is not sufficient to cause diabetes alone. The second accelerator is insulin resistance, which results from lack of physical activity and weight gain, and it further escalates the β-cell apoptosis. This second accelerator is driven by environmental factors and thought to be involved in the increasing diabetes incidence. In the last process a small proportion of people with genetic component, with an intrinsically high β-cell apoptosis rate and insulin resistance, develop islet autoimmunity.[119]

According to the accelerator and β-cell stress hypothesis visceral fat gain is as crucial for developing type 1 as it is for type 2 diabetes.[119] Supporting this hypothesis, many studies of various designs including a prospective cohort study in a high risk population, case-control

studies, and population-based linked data studies have demonstrated that higher weight gain in infancy is positively associated with type 1 diabetes risk.[120-123] For example, a population-based study that used Norwegian and Danish cohorts showed that higher weight gain from birth to six months was associated with 21% increased risk of childhood type 1 diabetes.[121] Similarly, the Australian BABY DIAB study demonstrated that weight gain in early infancy was associated with an increased risk of islet autoimmunity.[122]

Higher birthweight,[24, 124, 125] and gestational age (both preterm and post-term births)[124] have also been linked with higher childhood type 1 diabetes risk, although the evidence is divergent.[125, 126] For example, a meta-analysis of 12 studies showed that heavier birthweight (>4000 g) was associated with 17% increased risk of childhood type 1 diabetes, and lower birthweight (<2500g) was associated with 18% reduced type 1 diabetes risk.[127] The small increased risk of type 1 diabetes for children born with a higher birthweight is supported by a large meta-analysis consisting of 30 studies by Cardwell et al.[128] On the contrary, a large population-based study from Sweden that used a sibling design depicted no excess type 1 diabetes risk for children who were large-for-gestational-age, and similar to the meta-analysis depicted low risk of type 1 diabetes for extremely low birthweight children (<1500g).[126] It is difficult to tease apart why studies arrive at different conclusions, however, in this case the null effect from the sibling study points to the potential for residual confounding to have biased results in some of the observational studies. In addition, most previous studies have only used birthweight as a measure of growth, the sibling study used birthweight-for-gestational age, which extricates birth weight from gestational age.

2.6.3 Microbiota composition and type 1 diabetes

Microbiota composition is one of the hypothesized mechanisms linking various environmental factors with type 1 diabetes. Studies have reported the link between microbiota and type 1 diabetes pathogenesis.[129-132]A child's intestinal microbiota goes through changes during the first three years of life, and by age three it transitions in to an adult like composition.[131 133] Numerous factors can modify a child's intestinal microbiota composition; such as birth method,[134] exposure to antibiotics,[135, 136] microbial exposure, hygiene level in the immediate environment, feeding and dietary practices.[133, 137-139] During early life the immune system also develops, and intestinal microbiota can influence immune maturation.[140, 141] Studies have shown differences in microbiota composition of children with and without type 1 diabetes.[131, 142] In addition, alterations in microbiota have been shown to occur before the onset of islet autoimmunity among children who develop type 1 diabetes.[131, 143] The Diabetes Prediction and Prevention study (DIPP) study observed reduced diversity of the gut microbial composition in children who later develop type 1 diabetes, compared to healthy children.[144] This suggests an important role of microbiota composition in type 1 diabetes pathogenies.

2.7 Environmental determinants

Studies have suggested that environmental factors are interacting with genetic predisposition leading to an increase in type 1 diabetes incidence.[82, 106, 111, 145] Some authors even highlight an

increasingly important role of the changing environment in type 1 diabetes pathogenesis.[82] There could be factors changing in the environment that are potentially driving the increase in type 1 diabetes incidence. The role of the environment is supported by a study that followed type 1 diabetes progression rate in monozygotic twins in Britain and the US.[11] It demonstrated that the probability of concordance for type 1 diabetes was 39% at 40 years of follow-up from type 1 diabetes diagnosis of the index twin.[11] Most of the twins in that study remained discordant for type 1 diabetes at the end of follow-up, highlighting the role of non-genetic factors in the disease development. The role of environment in type 1 diabetes pathogenesis is supported by further evidence of a large disparity in type 1 diabetes incidence between the adjacent regions of Karelia, Russia (7.4 per 100,000) and Finland (41.4 per 100,000) from 1990-1999, although HLA DQ genotypes for type 1 diabetes were similarly prevalent in these areas.[146] In addition, migrants have been shown to acquire the same risk of type 1 diabetes as the population in the newly adopted country of residence.[147, 148] For example, a Swedish study (n = 1.9 million) demonstrated that residents born in Sweden with foreign-born parents had 60–70% higher type 1 diabetes prevalence compared with new immigrants and adoptees with the same origin.[147]

In the following sections I will describe the existing literature revolving around my doctoral project including; socioeconomic position and type 1 diabetes incidence; caesarean birth and maternal smoking in pregnancy and risk of type 1 diabetes; and implications of type 1 diabetes for children's educational outcome. I will also briefly mention the potential mechanisms (discussed above) linking each specific exposure with the outcome, and also how my research fits in the gap of existing literature.

2.7.1 Socioeconomic patterning of type 1 diabetes incidence

Socioeconomic circumstances are a major determinant of health, and can influence the living environment, household crowding, and health behaviours (smoking, nutritional intake etc) and hence can play role in altering the risk of type 1 diabetes. The hygiene hypothesis is one of the hypothesized link between socioeconomic conditions and type 1 diabetes incidence; potentially through differential microbial exposures and immune stimulation in low and high socioeconomic groups.[32] For example people from low socioeconomic conditions are more likely to live in crowded households and use public transport leading to more immune stimulation; and household crowding has been linked with low type 1 diabetes incidence.[95] There is mixed evidence about socioeconomic patterning of type 1 diabetes. However, most studies have relied on area-level measures, which don't account for individual variability in environments, which may expose children to more or less immune stimulants. Most previous studies that looked at the broader ecological picture, reported a positive association between country-level socioeconomic determinants and type 1 diabetes incidence.[105, 149] For example a Polish study reported a positive association between increasing life expectancy, and gross domestic product (GDP) and type 1 diabetes incidence.[105] The lack of data from low income countries on type 1 diabetes incidence makes it harder to draw firm conclusions as to whether socioeconomic circumstances are associated with type 1 diabetes or not. However, the countries ranked in the top 10 for type 1 diabetes incidence are mostly high income countries, such as Finland, Sweden, Norway, UK, Canada, Denmark, USA and Australia.[54] This suggests the economic development driven changes at a country level that flow on to impact individual

environments and lifestyles may play an important role in altering type 1 diabetes risk. Studies from middle income countries reported low type 1 diabetes incidence, such as 1.01 per 100,000 person-years in China (2010-2013),[8] and 12.8/100,000 in Brazil (1986 to 2015).[150]

Most studies have looked at the country-level, or small area level socioeconomic position and type 1 diabetes incidence using different measures. For example, studies from the UK have reported mixed evidence about the association between Townsend deprivation index (area-level measure of disadvantage) and type 1 diabetes incidence.[98, 99, 151] Two studies from Australia reported inconsistent evidence of association between the Index of Relative Socio-economic Disadvantage (IRSD) and type 1 diabetes incidence.[73, 152] Others have reported an association between neighbourhood level measures of affluence such as the proportion of people with higher education, higher income, and car ownership and type 1 diabetes incidence.[96] The indices (Townsend deprivation index and IRSD) used to measure area-level disadvantage, have been created by combining information on the prevalence of different characteristics of individuals within a particular area. For example, the IRSD combines information on the prevalence of people with low income, unemployed, with no internet, people classified as labours and machine operators, one parent families, overcrowded household, families with no cars and with poor English speaking skills,[153] to generate an area-level score of relative disadvantage. It is possible the different types and numbers of variables used to create these various indices may explain variability in the evidence of the association between area-level socioeconomic indicators and type 1 diabetes incidence. However, even when multiple studies used the same index, mixed findings of the association between small area-level socioeconomic conditions and type 1 diabetes incidence have been reported.[73, 98, 99, 151, 152] For example a Western Australian

population-based study that used IRSD as a measure of area-level socioeconomic disadvantage based on maternal residence at the time of child birth during 1980-2002, demonstrated no difference in type 1 diabetes incidence between people living in the least and most disadvantaged areas.[152] Yet another Western Australia study (1985 to 2002) reported a higher type 1 diabetes incidence in the least disadvantaged groups;[73] when used IRSAD based on the maternal residence at the time of type 1 diabetes diagnosis. This may point to that area-level socioeconomic indices are not reliable measures that can be used for studying health disparities because they do not capture individual variation and are prone to ecological fallacy. Work undertaken by the Australia Bureau of Statistics demonstrated individual variation within each area that in the area-level index would all be represented as having the same socioeconomic conditions. This adds to the evidence suggesting that area-level measures are not reflective of individual-level socioeconomic conditions.[154]

As can be seen from the above evidence, that most studies have used area-level measures of socioeconomic position that is probably because individual level-data is not easily available in most studies. Only two studies (Italy, US) have used individual-level socioeconomic information to study the incidence of type 1 diabetes, and both reported higher type 1 diabetes incidence in socioeconomically advantaged children.[32, 155] I took advantage of the linked data to look at type 1 diabetes incidence with both individual and area-level indicators of socioeconomic circumstances, which was an opportunity to understand if previous studies may be subject to the ecological fallacy, and to also explore the hygiene hypothesis. To my knowledge, all previous Australian studies on socioeconomic conditions and type 1 diabetes incidence, have used area-level measures of socioeconomic condition.[73, 152] Therefore in the first

descriptive study, I compared the area-level and individual-level socioeconomic disparity in type 1 diabetes incidence, using South Australian population-wide linked data. Due to the data-linkage, I had access to individual-level socioeconomic information to be able to study the Question-1 of my doctoral thesis (Chapter 4). The individual-level indicators of socioeconomic condition in my study include parent's employment, private or public healthcare, and whether the child was born in a public or private hospital. The area-level socioeconomic position used in my dissertation has been measured by The Index of Relative Socio-economic Advantage and Disadvantage (IRSAD). More detail about this index is given in Chapter 3.

2.7.2 Caesarean birth and risk of childhood type 1 diabetes

A hypothesized link between caesarean birth and type 1 diabetes is thought to be birth mode-induced variation in neonatal microbiota composition,[156] which could impact immune development.[140] Gut bacterial colonisation patterns differ among neonates born by emergency caesarean, elective caesarean and natural birth.[156] For example, children born by emergency and elective caesarean have different profiles than those born by normal vaginal delivery in colonization of *E coli, K oxytoca,* and *K pneumoniae*.[156] Gut microbial community composition and the products of bacterial metabolism may influence the education and maturation of the immune system in early life that is required for self/non-self regulation and thereby susceptibility to immune-mediated diseases.[140] Indeed, studies have shown differences in the gut microbial composition of children with and without type 1 diabetes.[142, 157]As mentioned

previously - changes in microbiota composition have been observed before the onset of islet autoimmunity among children who developed type 1 diabetes.[143] [131] A recent TEDDY study comparing the composition and functional capacity of the microbiome of children followed from three months up to age 5 years found that, compared with controls, the microbiome of children that developed type 1 diabetes had fewer genes related to fermentation and short-chain fatty acid synthesis.[158]

A further phenomenon that may link caesarean birth with type 1 diabetes is that in the last three decades both caesarean rates and type 1 diabetes incidence have increased in parallel, and this has occurred in many countries. Most of the countries ranked in the top ten for published type 1 diabetes incidence, such as Australia, the UK, Canada, the USA, Saudi Arabia and Denmark also have much higher caesarean rates than the WHO recommendation of 10-15% of all births.[18, 159, 160] However, not all countries with higher caesarean rates have high type 1 diabetes incidence. For example, Brazil[150, 161] and China[8, 162] have the highest rates of caesarean births in the world at 77% and 46% respectively, but have low type 1 diabetes incidence. However, studies that have tested the hypothesized link between the caesarean birth induced changes in neonatal microbiota and childhood type 1 diabetes have reported inconsistent evidence. Many case-control studies,[28, 163] a linked data study,[164] and a meta-analysis consisting of 20 studies (17 case-control)[25] showed 23-39% higher type 1 diabetes risk for children born by caesarean section. In contrast, large population-based cohort studies with better adjustment for potential confounding depicted a very small effect or no association between caesarean and childhood type 1 diabetes.[20-24] For example, a large population-based cohort study of 2.6 million Swedish children, when used sibling-design, the effect of caesarean birth on childhood type 1 diabetes

attenuated closer to the null.[20] This suggests previously reported higher effects of caesarean on childhood type 1 diabetes might be due to residual confounding.

Of the studies on caesarean birth and risk of type 1 diabetes, [20-24 28, 163, 164] only one[20] focused on caesarean type (emergency and elective caesarean) and risk of type 1 diabetes. However, emergency and elective caesarean birth do not distinguish whether birth occurred before or after the onset of labour. Therefore, given the proposed mechanism is through microbiome exposure I was able to take advantage of the whole-of-population administrative data in the SA ECDP and I explored the association between prelabour and intrapartum caesarean and risk of childhood type 1 diabetes in study 2. That's because children born by caesarean before labour (prelabour) have no exposure to maternal vaginal microbiota, and they might have high risk of type 1 diabetes compared with children born after the onset of labour (intrapartum caesarean). Therefore, in study 2, I estimated the risk of type 1 diabetes for children born by caesarean; extending the question to whether the risk differs by prelabour or intrapartum caesarean (detail in Chapter 5) to better understand the role of exposure to the vaginal microbiome.

2.7.3 Maternal smoking during pregnancy and risk of childhood type 1 diabetes

The mechanism linking maternal smoking during pregnancy and childhood type 1 diabetes is not fully understood, although it has been hypothesized to be through multiple pathways such as epigenetic modification (altered DNA methylation), gene expression or immune

suppression.[87, 165-169] A genome-wide meta-analysis of 13 birth cohort studies reported differential DNA methylation CpG sites among children born to mothers who smoked during pregnancy, compared to unexposed children.[87] Similarly, another study found differences in mean DNA methylation level between children with type 1 diabetes and their healthy co-twin.[170] There is also evidence about transfer of nicotine from mother to the foetus,[168, 171] and nicotine may impact both innate and adaptive immunity.[169, 172]

In many countries, the incidence of childhood type 1 diabetes has been increasing.[3, 54, 173]At the same time maternal smoking in pregnancy rates have been declining,[19, 174] raising a question about the causal effect of maternal smoking on childhood type 1 diabetes. For example, in Australia, type 1 diabetes incidence has increased from 21.5 per 100,000 population in 2000, to 24.7 per 100,000 population in 2015; [175] while prevalence of smoking during late pregnancy reduced from 17.3% in 2006 [176] to 9.9% in 2016.[177] The link between maternal smoking in pregnancy and childhood type 1 diabetes has been investigated using various study design (high risk population cohort studies, case-control studies, population-based studies); however the evidence is mixed; ranging from increased, [26, 27] decreased [28-33] to no risk.[34] Studies that showed higher risk of childhood type 1 diabetes for children exposed to maternal smoking were generally smaller studies conducted in populations at risk of developing type 1 diabetes.[26, 27] Studies that have shown decreased risk of childhood type 1 diabetes following exposure to maternal smoking in pregnancy have effect estimates that vary from 13%-60% reduced risk in case-control studies,[28, 32, 33] and 25%-35% lower risk in population-based cohort studies.[29, 30] The varying estimates could be due to small numbers of children exposed to maternal smoking, and could also be due to differences in adjustment for confounders. The variability in evidence

led me to extensively explore the effect of maternal smoking in pregnancy on childhood type 1 diabetes in study 3. To study the effect of maternal smoking in pregnancy on childhood type 1 diabetes, I used multiple population-wide administrative linked datasets. I deployed advanced analytical methods to assess the risk of bias due to unmeasured confounding such as negative-control outcome analysis and E-value. I had the problem of having a large study with an imprecise effect estimate, so did all the others studies, rather than simply adding another study to the mix, combining relevant studies in a meta-analysis would bring new knowledge. Therefore, I summarized the adjusted effect estimates from previous case-control and population-based cohort studies, along with my own study in meta-analyses in study 3 (detail in Chapter 6).

In the following section, I will focus on some other perinatal factors relevant to the studies undertaken in my doctoral thesis. Some of the factors I am going to discuss are potential confounders in my studies; including birthweight and gestational age, parental age, and maternal obesity. Some of the factors discussed below are the antecedents of the confounders, and are associated with the discussed hypothesized mechanisms of type 1 diabetes. For example pre-natal nutrition affects maternal pregnancy weight and child's weight , which are linked with the β-cell stress and accelerators hypothesis, and also influences microbiota composition impacting children's immune development. Although there are plethora of other risk factors that have been studied but I do not focus on them here.

2.7.4 Other Parental characteristics and risk of childhood type 1 diabetes

As mentioned, parental characteristics such as age, obesity, and diabetes are important confounders in my studies (Chapter 5, 6, and 7). Parental diabetes and child risk of type 1 diabetes has previously been discussed (section 2.5). Thus, here I briefly explore existing research on the effect of parental age, obesity, and some dietary factors on type 1 diabetes.

2.7.4.1 Parental age and childhood type 1 diabetes

Whether parental characteristics such as age at child birth are associated with childhood type 1 diabetes is uncertain.[24, 31, 163] Mostly older parental age (both mother's and father's) has been associated with a small increased risk of childhood type 1 diabetes.[31, 163, 178, 179] However, large population-based cohort studies with more comprehensive adjustment for confounding did not find clear parental age patterning for childhood type 1 diabetes.[21, 24, 163] For example, a meta-analysis of 30 observational studies, 25 of which were case-control studies, demonstrated that a five-year increase in maternal age was associated with a 5% increased risk of childhood type 1 diabetes.[180] In contrast, a large Swedish study consisting of 14,949 children with type 1 diabetes with individual level socioeconomic information of over 36 years, demonstrated no effect of mother's age.[24] It is possible the small observed effect of maternal age in previous case-control studies could be due to unmeasured confounding, as matching in case-control studies does not remove confounding and introduces bias if not included as an adjustment variables.[181, 182]

2.7.4.2 Maternal obesity and childhood type 1 diabetes

Most studies support the idea that maternal obesity is associated with increased risk of type 1 diabetes. Multiple case-control studies, population-based cohort studies and a meta-analysis showed that maternal obesity is associated with 18%-41% higher risk of childhood type 1 diabetes,[24, 31, 183-186] except a Finnish birth cohort study.[187] For example, a prospective population-based Finnish birth cohort study (type 1 diabetes n = 175) that monitored children with increased HLA-conferred susceptibility for 3-12 months from birth (1997-2002) showed no association between maternal BMI measured at the beginning of pregnancy and β-cell autoimmunity.[187] However, the largest population-based studies have shown positive associations between high maternal BMI and type 1 diabetes. Data from the Danish National Birth Cohort (children's mean age 15.5 years) and the Norwegian Mother and Child Cohort (children's mean age 11 years) showed that obesity was associated with a 41% higher incidence of type 1 diabetes (HR =1.41, 95% CI 1.06-1.89).[185] The balance of evidence suggests there is some increased risk of type 1 diabetes as a result of higher maternal BMI.

2.7.4.3 Maternal nutrition and childhood type diabetes

The mechanisms that potentially link maternal diet in pregnancy to children's risk of type 1 diabetes could be through the impact of diet on maternal weight, nutritional deficiencies, DNA methylation and the gut microbiota that children acquire at birth. There is divergent evidence about the effect of maternal nutrition in pregnancy such as gluten,[188-190] and vitamin D,[82 191-194] on childhood type 1 diabetes. Although there are not many studies that have explored the effect of maternal gluten intake on childhood type 1 diabetes; of those, a study in high-risk population

(based on HLA genotype and family history) showed a negligible protective effect;[189] whereas large population-based cohort studies demonstrated increased type 1 diabetes risk.[188] A small protective effect of higher prenatal vitamin D concentration on childhood type 1 diabetes is reported by many small studies.[82 191-194] However, a recent large population-based Scandinavian cohort study that estimated the effect of average maternal vitamin D concentration during pregnancy on type 1 diabetes risk and showed no effect.[195] No clear pattern has been observed between maternal polyunsaturated fatty acid intake or serum concentration during pregnancy[196, 197] or lactation[198] and risk of childhood type 1 diabetes. Maternal iron supplement intake during pregnancy has been associated with 5% to 33% higher type 1 diabetes risk.[199-201] However maternal iron intake from food, and anaemia or low haemoglobin levels has not been linked with high risk for childhood type 1 diabetes.[200] Some possibilities for these inconsistencies in evidence about the findings about maternal nutritional intake and type 1 diabetes incidence could be due to differences in study samples (high risk populations vs population-based), short duration of follow-up time and small number of children with type 1 diabetes which affected the power and precision, and perhaps measuring the blood concentration of nutrients at different points in pregnancy.

2.7.5 Child feeding and early life characteristics and type 1 diabetes

2.7.5.1 Breastfeeding and childhood type 1 diabetes

Breastfeeding could be linked with type 1 diabetes through its potential impact on gut microbiota composition[137] and immune development.[202, 203] There is consensus in evidence about a small protective effect of breastfeeding for childhood type 1 diabetes, however, there is heterogeneity in the evidence about the effect of exclusive breastfeeding duration.[27, 204-208] For example, a 2019 systematic review concluded there was little to moderate evidence that never feeding human milk compared with ever feeding human milk (both breastfed or human milk fed by other methods), and shorter duration of exclusive human milk feeding were associated with higher type 1 diabetes risk.[205] Similarly a large population-based study that used data from two large Scandinavian birth cohorts showed twofold increased risk of type 1 diabetes for children who were never breastfed (HR = 2.29, 95% CI 1.14-4.61) compared with breastfed children.[204]

2.7.5.2 Infant feeding, weaning age and childhood type 1 diabetes

Early feeding practices induce changes in children's gut microbiota,[138, 209] and thus may influence their immune maturation. Infant feeding also influences children's body weight,[210] and weight gain in early infancy has been linked with increased risk of type 1 diabetes (has been discussed above).[119-123] Many studies have highlighted age at cereal or gluten introduction (both early and delayed gluten introduction) as a potential risk factor for type 1 diabetes, however, the evidence is mixed.[189, 207, 208, 211-214] For example, an American prospective birth cohort study (DAISY) conducted in a high risk population either HLA genotype or have a first-degree relative with type 1 diabetes (n=1183, 1994-2002) demonstrated that both early (0 to 3 months, HR = 2.65; 95% CI 0.76-9.33) and late (≥7 months, HR = 1.70, 95% CI 0.79-3.7) initiation of gluten

containing cereals, compared to 4-6 months were associated with an increased incidence of islet autoimmunity in children.[211] Early gluten introduction and high type 1 diabetes risk is supported by other studies.[208 27] In contrast, the TEDDY[212] study that prospectively followed genetically at risk children (n = 8,676) in the US, Finland, Germany and Sweden reported that late (>9 months) introduction of gluten increased the rate of islet autoantibody appearance (HR = 1.57, 95% CI 1.07-2.31), compared to introduction between 4-9 months of age, and demonstrated lower incidence of insulin autoantibody appearance (HR = 0.68, 95% CI 0.47- 0.99) for early introduction of gluten (<4 months). A Finnish prospective birth cohort study in a high genetic risk group (n = 5,915) demonstrated that early (<3 months) introduction of solid food was associated with higher type 1 diabetes risk only in short term follow-up (up to age 3 years), but not at longer term follow up (15 years).[207] There is also discrepancy in evidence about weaning age for other solid foods and type 1 diabetes risk.[27, 82, 208] The reasons for variability in the findings about gluten introduction and type 1 diabetes incidence in these cohort studies could be due to; differences in the timings and type of first complementary food; variation in length of follow up; differences in types of infant formulas, and perhaps changing infant feeding habits or use of probiotics overtime in different populations.

2.7.5.3 Other early-life characteristics and type 1 diabetes

Looking broadly across the literature, other child related factors associated with an increased risk of childhood type 1 diabetes include higher neonatal blood iron content,[201] high iron intake in first four months of life,[215] hospitalization for gastroenteritis,[216] exposure to broad spectrum antibiotics in the first two years of life (particularly among caesarean births)[217].[218] Meta-

analyses by Cardwell et al. (2011, 2012)[219, 220] have demonstrated that a birth interval of ≤3 years[219] and increasing birth order[220] are associated with a small lower childhood type 1 diabetes risk, which could be linked with the hygiene hypothesis. Childhood vaccinations[221] have also been investigated and found to not be linked to type 1 diabetes.

2.8 Burden of type 1 diabetes

In recent decades the reduction in incidence rates of severe complications of type 1 diabetes such as diabetic retinopathy, nephropathy, and cardiovascular diseases has been possible through technological advances in glucose monitoring and insulin delivery, and improved management of type 1 diabetes.[3, 222-225] In the pre-insulin era (prior to 1922) type 1 diabetes was a fatal disease, and 50% patients would die within the first 20 months of diagnosis with less than 10% survival over 5 years.[226] Type 1 diabetes-specific mortality was 824 deaths/1,000 person-years in 1898-1914 among ≤10 year old children in the US.[227] After insulin discovery and use for treatment, type 1 diabetes-specific mortality declined sharply to 61 deaths per 1,000 person-years in 1922-1926, to less than 1 death per 1,000 in 1950-1961.[227]Although the type 1 diabetes-specific mortality rate has declined, however, a gap in life expectancy of people with and with and without type 1 diabetes persists.[226 228] Achieving glycaemic control remains challenging for people with type 1 diabetes, and some acute and chronic complications contribute to lower life expectancy.[229] In Australia, people with type 1 diabetes have a life expectancy of 68.6 years, which is 12.2 years less than the general population.[15]

Poor glycaemic control leading to diabetic ketoacidosis and severe hyper-and-hypoglycaemia are major causes of morbidity, hospitalization,[12] mortality[15] and high health-related costs.[17] Children with type 1 diabetes have been reported to have three times more hospital days per person-year than children without type 1 diabetes.[12] Children with type 1 diabetes may be prone to other illnesses and experience more hospitalizations, which is one of the reasons for the high costs to the health system. For example, in a European survey study, people with type 1 diabetes reported poorer health and poor health related quality of life;[230] and reported much higher prevalence of chronic comorbidities including pain (37.8% vs 22.2%), hypertension (28.1% vs 16.3%), and depression (16.9% vs 10.5%) than the general population.[230] Managing a complex medical condition requiring daily adherence to multiple self-care behaviours is challenging and stressful for children with type 1 diabetes. Over 50% of adolescent with type 1 diabetes reported general and diabetes-specific stress, and high level of stress was negatively associated with poor quality of life and fewer self-management activities.[231] In addition, children's diagnosis with type 1 diabetes leads to higher levels of paediatric parenting stress.[232] It suggests that type 1 diabetes not just effects the individual diagnosed with this illness; but has substantial psychological consequences for the family, because of the complex nature of the disease that requires vigilance with glycaemic control for survival and to avoid debilitating complications like stroke, vision loss, and amputations.

For the individual, type 1 diabetes has an impact on patients' quality of life and their life expectancy, and for a country, type 1 diabetes is linked to high costs of health care and increased hospitalisations. The annual estimated health care cost of type 1 diabetes for Australia in 2008 was $570 million and hospitalizations accounted for 47% of this cost.[16, 17] The Australian

government started subsidising insulin pumps in 2008 for low income families with children ≤18 years old, with full subsidization of continuous glucose monitoring systems for all children aged <21 in 2017,[233-235] which is an additional cost to the health system. Although type 1 diabetes only affects ~0.3% of the general population, its complications account for 5% of all hospital bed days, and 4% of all ambulatory care or hospital admissions, which is more than angina or asthma.[236]

2.8.1 Type 1 diabetes and children's academic achievement

Children with type 1 diabetes may potentially be at risk of poor cognitive function and educational outcomes, due to poor glycaemic control,[237, 238] illness-related hospitalization,[12] absences from school,[13 239] and psychological challenges.[14] As glucose is a major fuel source for the brain, both hyperglycemic[240-242] and hypoglycemic[243] episodes have the potential to impact brain functioning.[238] Diabetic ketoacidosis (DKA) at type 1 diabetes presentation has both short and long term consequences for glycaemic control (through exacerbation of β-cell loss),[240] and memory function.[241, 242] For example, a cohort study reported morphological and functional brain changes among children with DKA at initial type 1 diabetes diagnosis, demonstrating that while most of the changes resolved, brain changes during DKA were associated with alterations in attention and memory at six months follow-up.[238] The psychosocial challenges[14] that children with type 1 diabetes face could also impact their quality of life and school outcomes. A study from Victoria, Australia that interviewed children 12 years

after initial recruitment into a longitudinal study reported that mental health referral rates were 19% higher among children with type 1 diabetes compared with healthy controls.[14]

Despite the multiple potential mechanisms linking type 1 diabetes with children's educational outcomes, there is uncertainty around the effect of childhood type 1 diabetes on children's educational outcome. Small cross sectional studies have demonstrated poor cognitive functions among children with type 1 diabetes.[244, 245] Large population-based studies that use data from three decades ago demonstrated negative effects,[35, 36] and two recent studies showed no effect of type 1 diabetes on educational outcomes.[13, 37] Recently, a large Swedish population-based study showed a negative effect of type 1 diabetes on children's educational outcomes.[246] This study showed that children with type 1 diabetes are still disadvantaged in educational outcome in some developed countries, even in an era of improved type 1 diabetes management. Similarly, a nationwide Swedish population-based study that used data from two decades ago (births 1973-1986, school data from 1988-2003) reported poor school outcomes in mathematics, English, Swedish and sports for children with type 1 diabetes at the end of compulsory schooling (<16 years of age) compared to children without type 1 diabetes.[35] A recent Western Australian study that linked state-wide diabetes register data with school assessment data (2008-2011) demonstrated that type 1 diabetes was not associated with a decrement in school performance.[13] These divergent findings could be due to variability in adjustment for confounders, and due to advances in management of type 1 diabetes. For example, most previous studies did not adjust for father's age, pre-pregnancy hypertension and diabetes, ethnicity and other socioeconomic indicators, and were at risk of bias due to unmeasured confounding.[26, 30, 31] In addition, mostly small cross sectional studies have estimated the effect of type 1 diabetes on the cognitive

functions of children using cognitive tests. In my fourth study, I used academic test scores as an outcome from national school assessment program, which measures the real-life educational experiences of children, and important for moving ahead in school, and for future employment.[247-249] In the last decade there has been substantial improvement in type 1 diabetes management, I wanted to explore the extend to which these improvement have benefited children in South Australia in reaching optimum learning capabilities at schools. I used the potential outcome approach to compute the average treatment effect of type 1 diabetes on children's educational outcome, using AIPW. The results obtained from a potential outcome approach (the outcome which would have been observed if the exposed person had not been exposed) such has AIPW may be interpreted as the outcome as though the entire population were exposed or unexposed to type 1 diabetes. AIPW procedure can have a similar interpretation to a randomized control trial (RCT) in the absence of unmeasured confounding[250] (more detail on AIPW is given in the next method chapter, and detail about this study is provided in Chapter 7).

CHAPTER 3 METHODS

In this chapter, I describe; 1) the study population and design, 2) the multiple administrative datasets used for my doctoral research; 3) the data linkage process; 4) the epidemiological concepts including methods to deal with confounding; and 5) the analytical approaches used in each of the four research questions.

My doctoral research includes studies of the incidence, risk factors and outcomes of type 1 diabetes for children in South Australia, born from 1999-2013. South Australia is a southern State in Australia with a population of 1.677 million, and accounts for 7% of the total Australian population.[251]South Australia's population is representative of Australia in terms of socioeconomic characterises and heath indicators such as life expectancy,[252] and maternal and child health characteristics.[253, 254]

Most of the data used in my doctoral thesis was sourced from the South Australian heath system's hospitalization and perinatal statistics. In Australia, mostly the first contact an individual has with the health system is often with primary health care (delivered by general practitioners, dentists, nurses, allied health professionals etc), which broadly encompasses care unrelated to a hospital visit. Sometimes individuals may also be admitted to hospital following presentation at an emergency department. When there is need, individuals are referred by primary health care providers to the secondary care services provided in both private and public hospitals.[255] Public hospitals are funded and managed by the state and federal government, and

private hospitals are licensed and regulated by governments, but owned and operated by the private sector. In Australia, of births that occur in hospital, most births occur in public hospitals (74% in 2017), and 26% of children are born in private hospitals.[253] The Australian Government has funded a universal public health insurance scheme called Medicare since 1984.[255] Medicare covers hospital, medical and pharmaceutical expenses, and is funded through general taxation revenue and a specific 2% Medicare levy. Ambulance services, dental treatment, physiotherapy and other allied health services, glasses and contact lenses are not subsidized through Medicare. Medicare is only available to Australian and New Zealand citizens, Australian permanent residents, and people from countries with health insurance agreements.[255] Other people outside these categories must take out private health insurance, or pay full fees for health services.

3.1 Study population

To investigate the four research questions in my dissertation, I used multiple administrative, whole-of-population linked data sources as part of the South Australian Early Childhood Data Project (SA ECDP), which are routinely collected by various government departments in South Australia. Whole-of-population here means that every South Australian child, born from 1999-2013 who has a birth, perinatal or hospitalisation record is included in my studies. A range of information on births, perinatal characteristics, hospitalizations, and school assessments were linked by an independent data linkage agency (explained in section 3.3). Overall there were n =

286,058 South Australian children born from 1999-2013. Of these, n =285,871 with a South Australian birth or perinatal record, and 87 had a hospitalisation record (but no birth or perinatal record). Of the total children, 557 children or 0.2% were identified as having type 1 diabetes from all South Australian public hospital admission data from the calendar years 2001-2014. Children entered the SA ECDP at birth, in each successive birth years from 1999-2013, therefore each birth cohort consist of different age groups of children. For example birth cohort of 1999-2013 consist of children aged ≤15 years. And birth cohort 1999-2001 only consist of ≤ 2 years old children. Children were followed from birth until the diagnosis of type 1 diabetes, or censored at the end of follow up in June 2014. The study population size, the birth years included and length of follow-up differed in each study depending on the administrative datasets required to answer each research question (Table 3.1, details are in section 3.6).

3.2 Data sources

In this thesis, I used data from the South Australian Early Childhood Data Project (SA ECDP Figure 3-1), [256-258] which has been created by linking multiple government administrative datasets. The SA ECDP, since its inception in 2008 has been led by Professor John Lynch, and been sustained by continuous project grants and government partnerships secured by the SA ECDP lead researchers. New data is added to the existing SA ECDP database,[256-258] in an ongoing process of grant applications, ethics approvals, and data linkage. The de-identified data from the SA ECDP is used for research and government reporting.[256-258] The datasets included

in the SA ECDP have information on pregnancy and birth, maternal and child heath, child development, child protection, education, and hospitalizations (Figure 3-1). The datasets from the SA ECDP used in my doctoral project included birth registration data, perinatal statistics, hospitalization data, school enrolment and school assessment data. A summary of the data sources used in each study is given in Table 3-1 with a detailed description in the following sections.

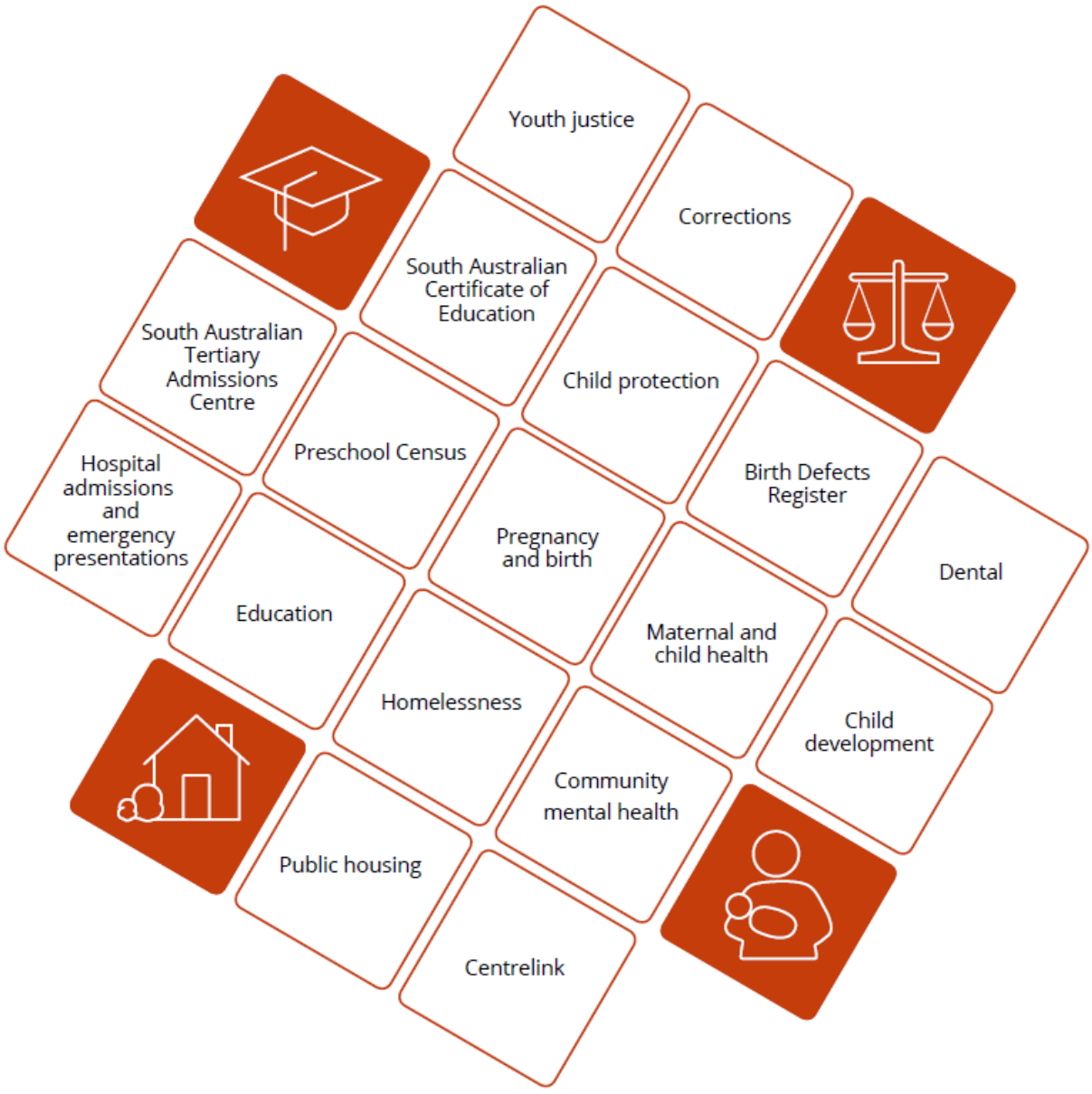


Figure 3-1: Data Sources held in the South Australia Early Childhood Data Project (SA ECDP)

Figure source : Nuske et al. 2016, Pilkington et al. 2017, and Pilkington et al. 2019 [256-258]

3.2.1 Birth Registration data

The first source of data is birth registrations on South Australian children born from 1999-2013, collected by the South Australian government's office of Births, Deaths and Marriages. In South Australia, all babies born >20 weeks gestation and >400 grams are legally required to be registered within 60 days after birth.[256, 259] This important source of information is used by the government for counting citizens, and identification for issuing driver's licences and passports, and for government services such as schools, Medicare and Centrelink (social security support) enrolment. The data collected through the birth registry includes mother's and father's age, ethnicity, child demographics, and basic clinical information such as birthweight and plurality. In this doctoral thesis, I used this dataset as a supplement and cross check for the more comprehensive perinatal statistics data.

3.2.2 Perinatal Statistics Collection

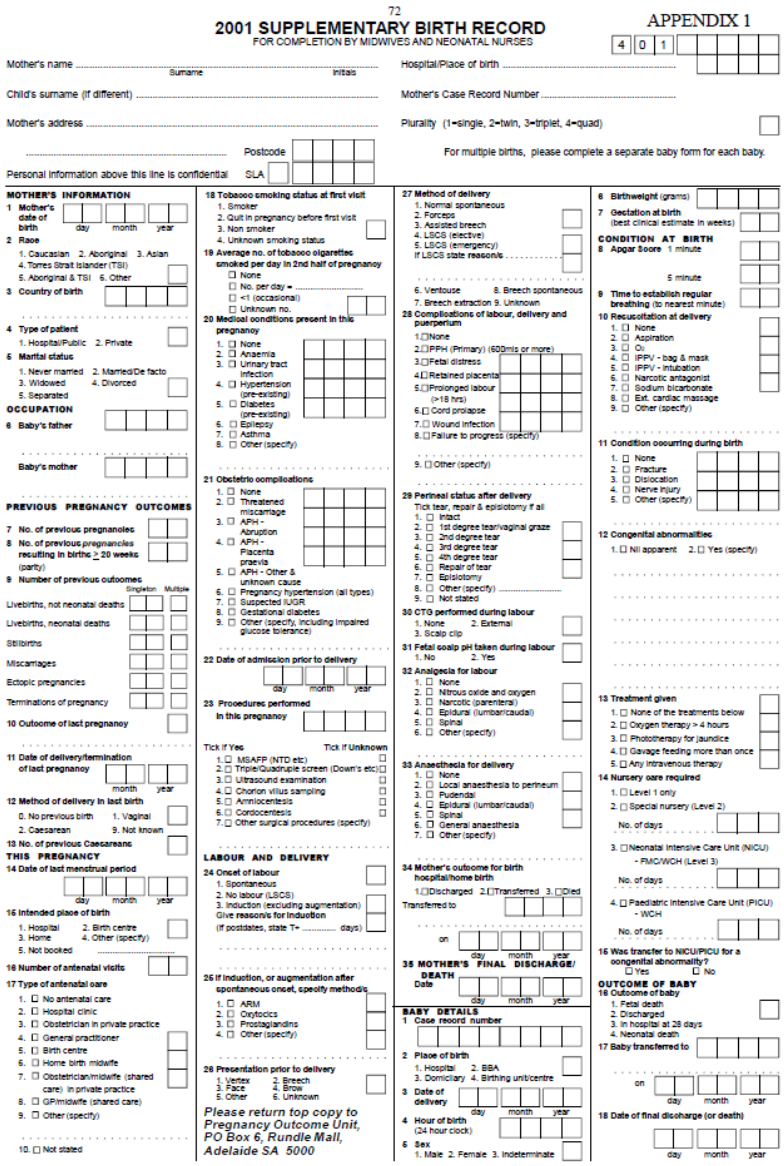
According to South Australian law enacted in the Health Care Regulations 2008[260], it is mandatory to report perinatal information for every birth (>20 weeks gestation and >400 g) occurring in South Australia to the Minister for Health and Wellbeing within ~30 days after birth.[259] This also includes births to women who reside interstate, if the birth occurs in South

Australia. The South Australian Ministry of Health and Wellbeing has been collecting this information since 1981, and has been using the data for producing annual pregnancy reports to track maternal and child health indicators.[259] All Australian jurisdictions report perinatal data to the federal government, which is used to monitor maternal and child health indicators at a national level.[253] Perinatal data is collected by midwives or neonatal nurses usually at the time of birth either at home or at hospital using a validated form called the Supplementary Birth Record. In Australia in 2017, 97% of births occurred in hospitals, 2.4% of births occurred in birth centres, 0.3% were home births, and 0.7% occurred before arriving at hospital.[253] Irrespective of location, the Supplementary Birth Record is mandated to be filled out for all births.

The information collected by nurses and midwives using the Supplementary Birth Record was validated in 1991 (records of 1986) and 2001 (records of 1994) against medical records collected by South Australian hospitals.[261] The validations were performed by the South Australian Pregnancy Outcome Unit, Department of Human Services, to check the accuracy of the perinatal data collection process. Perinatal information is important for identifying population groups at risk of poor pregnancy outcomes, for preventive interventions, and for monitoring maternal and child health indicators. In the second validation using records from 1994, 121 variables in the Supplementary Birth Record form (including all perinatal variables used in my doctoral thesis) were examined against the hospital records, and found to be valid in capturing the perinatal information.[261] The maternal smoking and caesarean section information from the perinatal statistics were used as exposures for studies 2 and 3 (Chapter

5 and 6). Data on caesarean birth and onset of labour was part of the 1994 Supplementary Birth Record form, which was validated. Collection of information on maternal smoking (as used in study 3 in this doctoral dissertation) in the first and second half of pregnancy has been included in the Supplementary Birth Record since 1998 (Figure 3-2).[262] Therefore, maternal smoking information was not part of the 1994 Supplementary Birth Record form that was validated. However, other studies have demonstrated that maternal smoking reported in pregnancy is reflective of actual tobacco exposure in pregnancy.[171]

The perinatal statistics data used in this project consists of all children born in South Australia from 1999-2013. Perinatal statistics includes basic demographic, pregnancy and birth information about the mother; e.g. age, employment, ethnicity, smoking status, maternal BMI (body mass index), postcode, parity, and any medical condition or complications of pregnancy, labour or delivery.[259] It also includes neonatal information, such as infant gestation, birthweight, sex and birth outcomes. The perinatal statistics data provided information about the exposures in study 1 (socioeconomic variables), 2 (caesarean section), and 3 (maternal smoking in pregnancy); and relevant confounders for each research question.



72

2001 SUPPLEMENTARY BIRTH RECORD

FOR COMPLETION BY MIDWIVES AND NEONATAL NURSES

APPENDIX 1

4 0 1

Mother's name ..... Surname ..... Initials

Hospital/Place of birth .....

Child's surname (if different) .....

Mother's Case Record Number .....

Mother's address .....

Plurality (1=single, 2=twin, 3=triplet, 4=quad)

..... Postcode

For multiple births, please complete a separate baby form for each baby.

Personal information above this line is confidential SLA

MOTHER'S INFORMATION

1 Mother's date of birth day month year

2 Race
1. Caucasian 2. Aboriginal 3. Asian 4. Torres Strait Islander (TSI) 5. Aboriginal & TSI 6. Other

3 Country of birth

4 Type of patient
1. Hospital/Public 2. Private

5 Marital status
1. Never married 2. Married/De facto 3. Widowed 4. Divorced 5. Separated

OCCUPATION

6 Baby's father

Baby's mother

PREVIOUS PREGNANCY OUTCOMES

7 No. of previous pregnancies

8 No. of previous pregnancies resulting in births ≥ 20 weeks (parity)

9 Number of previous outcomes

	Singleton	Multiple
Livebirths, not neonatal deaths		
Livebirths, neonatal deaths		
Stillbirths		
Miscarriages		
Ectopic pregnancies		
Terminations of pregnancy		

10 Outcome of last pregnancy

11 Date of delivery/termination of last pregnancy month year

12 Method of delivery in last birth
0. No previous birth 1. Vaginal 2. Caesarean 9. Not known

13 No. of previous Caesareans

THIS PREGNANCY

14 Date of last menstrual period day month year

15 Intended place of birth
1. Hospital 2. Birth centre 3. Home 4. Other (specify) 5. Not booked .....

16 Number of antenatal visits

17 Type of antenatal care
1. ☐ No antenatal care
2. ☐ Hospital clinic
3. ☐ Obstetrician in private practice
4. ☐ General practitioner
5. ☐ Birth centre
6. ☐ Home birth midwife
7. ☐ Obstetrician/midwife (shared care) in private practice
8. ☐ GP/midwife (shared care)
9. ☐ Other (specify) .....
10. ☐ Not stated

18 Tobacco smoking status at first visit
1. Smoker
2. Quit in pregnancy before first visit
3. Non smoker
4. Unknown smoking status

19 Average no. of tobacco cigarettes smoked per day in 2nd half of pregnancy
☐ None
☐ No. per day = ...
☐ <1 (occasional)
☐ Unknown no.

20 Medical conditions present in this pregnancy
1. ☐ None
2. ☐ Anaemia
3. ☐ Urinary tract infection
4. ☐ Hypertension (pre-existing)
5. ☐ Diabetes (pre-existing)
6. ☐ Epilepsy
7. ☐ Asthma
8. ☐ Other (specify) .....

21 Obstetric complications
1. ☐ None
2. ☐ Threatened miscarriage
3. ☐ APH - Abruption
4. ☐ APH - Placenta praevia
5. ☐ APH - Other & unknown cause
6. ☐ Pregnancy hypertension (all types)
7. ☐ Suspected IUGR
8. ☐ Gestational diabetes
9. ☐ Other (specify, including impaired glucose tolerance) .....

22 Date of admission prior to delivery day month year

23 Procedures performed in this pregnancy

	Tick if Yes	Tick if Unknown
1. ☐ MSAFP (NTD etc)		☐
2. ☐ Triple/Quadruple screen (Down's etc)		☐
3. ☐ Ultrasound examination		☐
4. ☐ Chorion villus sampling		☐
5. ☐ Amniocentesis		☐
6. ☐ Cordocentesis		☐
7. ☐ Other surgical procedures (specify)		

LABOUR AND DELIVERY

24 Onset of labour
1. Spontaneous
2. No labour (LSCS)
3. Induction (excluding augmentation)
Give reason/s for induction
(If postdates, state T+ ..... days)

25 If induction, or augmentation after spontaneous onset, specify method/s
1. ☐ ARM
2. ☐ Oxytocics
3. ☐ Prostaglandins
4. ☐ Other (specify) .....

28 Presentation prior to delivery
1. Vertex 2. Breech 3. Face 4. Brow 5. Other 6. Unknown

Please return top copy to Pregnancy Outcome Unit, PO Box 6, Rundle Mall, Adelaide SA 5000

27 Method of delivery
1. Normal spontaneous
2. Forceps
3. Assisted breech
4. LSCS (elective)
5. LSCS (emergency)
If LSCS state reason/s .....
6. Ventouse 8. Breech spontaneous
7. Breech extraction 9. Unknown

28 Complications of labour, delivery and puerperium
1. ☐ None
2. ☐ PPH (Primary) (600mls or more)
3. ☐ Fetal distress
4. ☐ Retained placenta
5. ☐ Prolonged labour (>18 hrs)
6. ☐ Cord prolapse
7. ☐ Wound infection
8. ☐ Failure to progress (specify) .....
9. ☐ Other (specify) .....

29 Perineal status after delivery
Tick tear, repair & episiotomy if all
1. ☐ Intact
2. ☐ 1st degree tear/vaginal graze
3. ☐ 2nd degree tear
4. ☐ 3rd degree tear
5. ☐ 4th degree tear
6. ☐ Repair of tear
7. ☐ Episiotomy
8. ☐ Other (specify) .....
9. ☐ Not stated

30 CTG performed during labour
1. None 2. External 3. Scalp clip

31 Fetal scalp pH taken during labour
1. No 2. Yes

32 Analgesia for labour
1. ☐ None
2. ☐ Nitrous oxide and oxygen
3. ☐ Narcotic (parenteral)
4. ☐ Epidural (lumbar/caudal)
5. ☐ Spinal
6. ☐ Other (specify) .....

33 Anaesthesia for delivery
1. ☐ None
2. ☐ Local anaesthesia to perineum
3. ☐ Pudendal
4. ☐ Epidural (lumbar/caudal)
5. ☐ Spinal
6. ☐ General anaesthesia
7. ☐ Other (specify) .....

34 Mother's outcome for birth hospital/home birth
1. ☐ Discharged 2. ☐ Transferred 3. ☐ Died
Transferred to
on day month year

35 MOTHER'S FINAL DISCHARGE/ DEATH
Date day month year

BABY DETAILS

1 Case record number

2 Place of birth
1. Hospital 2. BBA 3. Domiciliary 4. Birthing unit/centre

3 Date of delivery day month year

4 Hour of birth (24 hour clock)

5 Sex
1. Male 2. Female 3. Indeterminate

6 Birthweight (grams)

7 Gestation at birth (best clinical estimate in weeks)

CONDITION AT BIRTH

8 Apgar Score 1 minute

5 minute

9 Time to establish regular breathing (to nearest minute)

10 Resuscitation at delivery
1. ☐ None
2. ☐ Aspiration
3. ☐ O2
4. ☐ IPPV - bag & mask
5. ☐ IPPV - intubation
6. ☐ Narcotic antagonist
7. ☐ Sodium bicarbonate
8. ☐ Ext. cardiac massage
9. ☐ Other (specify) .....

11 Condition occurring during birth
1. ☐ None
2. ☐ Fracture
3. ☐ Dislocation
4. ☐ Nerve injury
5. ☐ Other (specify) .....

12 Congenital abnormalities
1. ☐ Nil apparent 2. ☐ Yes (specify) .....

13 Treatment given
1. ☐ None of the treatments below
2. ☐ Oxygen therapy > 4 hours
3. ☐ Phototherapy for jaundice
4. ☐ Gavage feeding more than once
5. ☐ Any intravenous therapy

14 Nursery care required
1. ☐ Level 1 only
2. ☐ Special nursery (Level 2) No. of days .....
3. ☐ Neonatal Intensive Care Unit (NICU) - FMC/WCH (Level 3) No. of days .....
4. ☐ Paediatric Intensive Care Unit (PICU) - WCH No. of days .....

15 Was transfer to NICU/PICU for a congenital abnormality?
☐ Yes ☐ No

OUTCOME OF BABY

16 Outcome of baby
1. Fetal death
2. Discharged
3. In hospital at 28 days
4. Neonatal death

17 Baby transferred to
on day month year

18 Date of final discharge (or death) day month year

Figure 3-2: Supplementary Birth Records form 2001

Source: South Australian pregnancy report 2001[262]

3.2.3 Hospital Admissions data

Children with type 1 diabetes were identified from the Integrated South Australian Activity Collection (ISAAC), which contains admissions to all public hospitals in South Australia.[263] I used admission data from July 2001 to June 2014 in this thesis as hospital data prior to July 2001 was not available for linkage.[256] The hospital admission data consists information on date of birth, sex, date of admission and discharge, age at admission, ethnicity, postcode, and up to 26 different diagnoses codes (depending on number of illnesses diagnosed in each hospital admission episode). The diagnose-1 is the primary diagnosis, and the remaining 25 diagnoses are additional or secondary diagnoses on each admission episode.

I used the International Classification of Disease, Tenth revision, Australian Modification (ICD-10-AM) codes E10, E101-E109 to identify children with type 1 diabetes from all public hospitals data in South Australia. The ICD-10-AM codes are assigned to each patient by the trained hospital staff, following discharge after each hospital admission episode.[264] In South Australia there is only one hospital with a specialized paediatric endocrinology unit, where children are admitted for stabilization after type 1 diabetes diagnoses, and that is a public hospital so was included in this study along with all other public hospitals data.[265]

Children appearing in the public hospital system may have multiple admission episodes from 2001-2014 and each hospitalisation may have up to 26 different ICD-10-AM diagnosis codes, depending on the health conditions of the child. For example, if a child is hospitalized

after experiencing an injury, the injury is recorded as the primary diagnosis, and other health conditions (asthma, type 1 diabetes etc.) are recorded as additional diagnoses. I used the primary and the 25 additional ICD-10-AM diagnosis codes to identify children with type 1 diabetes. For children with multiple hospitalisations for type 1 diabetes, I extracted the first hospital admission with type 1 diabetes to identify the number of individuals with type 1 diabetes. The date of admission with first type 1 diabetes hospitalization was used as age of the first type 1 diabetes diagnosis.

From 2001-2014, 557 children aged ≤15 years were identified as having type 1 diabetes from the hospitalization data. This equates to an incidence rate of 25.3 per 100,000 person-years, which is consistent with the incidence of 24.4 per 100,000 population (from 2000-2016) estimated using Australian national diabetes register data.[175] The consistency of type 1 diabetes incidence in my thesis with the Australian national diabetes register supports the high case ascertainment in this study.[175] The National Diabetes Register is a database of Australians who use insulin for treating diabetes. The National Diabetes Register consists of multiple data sources; including National Diabetes Services Scheme (NDSS); the NDSS Sales data, which records the start of first insulin use based on purchase of insulin and injections; and the Australasian Paediatric Endocrine Groups (APEG) state-based registers. NDSS is how people access subsidised insulin related products, and the reason why the national diabetes registers is believed to have every child diagnosed with type 1 diabetes in Australia.[266]

In my thesis, children with type 1 diabetes were identified from all public hospital admissions. There are no private paediatric hospitals in South Australia, where these kids could be hsopitalised. The National Diabetes Register data is not linked to the SA ECDP as was not ethically approved and had no funding. To confirm whether numbers were right, I compared the annual incidence in the SA ECDP with published NDR data and it was found to be consistent. Others have also shown that similar public hospital data are suitable for identifying cases of type 1 diabetes with 99.8% case ascertainment, and this is also likely to be the case for the SA ECDP.[267] The inclusion of both primary and additional/secondary ICD-10-AM diagnoses codes to identify children with type 1 diabetes helped ensure high case ascertainment.

3.2.4 Education data

3.2.4.1 School assessment data

The South Australian Department for Education provided all the public schools data on children's educational outcomes used for this doctoral research.[268] The school data is collected under a national program called "The National Assessment Program—Literacy and Numeracy – NAPLAN". The NAPLAN assessments used from 2008-2015 are a paper and pencil test administered by teachers in children's classrooms. Australian school children sit

the NAPLAN assessments on the same day in grade/years 3, 5, 7 and 9. NAPLAN assesses children's understanding of the Australian curriculum for literacy and numeracy skills.[268] NAPLAN scores are important to determine children's progress through school.[247-249] NAPLAN results are reported using five scales; one for each domain of writing, reading, spelling, grammar and numeracy (Figure 3-3). Raw NAPLAN scores are converted to scaled scores, ranging from 0-1000. I used year 5, age ~10 NAPLAN data in study 4 (explained in section 3.6.4). In 2017, the average year 5 NAPLAN scale scores for South Australian children ranged from 455 to 494, and the Australian national average ranged from 472 to 505 (Figure 3-3); South Australian children scored lower across all NAPLAN domains compared with the Australian national average.[268]

NAPLAN assessment data from 2008-2015 were used in this project because NAPLAN assessments commenced in 2008, and the latest available data were from 2015. Available information in the school assessment dataset includes the test scores undertaken by students, achievement against proficiency bands, and student demographics.

3.2.4.2 South Australian School Enrolment Census

The School Enrolment Census is collected each year in term 1 and term 3 of the school year (there are four 10-week terms in a school year) by the Government of South Australia, Department for Education. The School Enrolment Census contains important information on parental education, and country of birth that were used as confounders in this research. There are multiple census enrolment records for each child, but only the first school enrolment

census record for each child was used as country of birth and parental education were not likely to change between school enrolment and the NAPLAN assessment. If parents' education data was missing, then parental education data was extracted using their sibling's school enrolment information. Parent's education information from this dataset was used as a confounder in study 4 (Chapter 7), to estimate the effect of type 1 diabetes on educational outcomes. Parents' education data is therefore only available for children who have entered school (from approximately age 5) and had a public-school enrolment record.

Reading

	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Aust
Mean scale score / (S.D.)	508.0 (78.3)	514.6 (71.6)	502.8 (75.0)	498.9 (78.5)	494.3 (75.0)	499.1 (81.7)	520.0 (75.6)	429.9 (113.2)	505.7 (77.0)

Writing

	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Aust
Mean scale score / (S.D.)	477.6 (62.3)	485.6 (56.3)	461.5 (65.1)	468.6 (66.5)	455.5 (64.6)	465.2 (68.5)	479.4 (60.4)	395.4 (116.8)	472.5 (64.4)

Spelling

	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Aust
Mean scale score / (S.D.)	508.5 (73.5)	503.5 (65.9)	496.3 (68.6)	498.2 (73.2)	490.6 (68.6)	483.4 (75.5)	498.5 (67.6)	431.2 (103.8)	500.8 (71.3)

Grammar and punctuation

	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Aust
Mean scale score / (S.D.)	505.7 (83.2)	504.7 (73.7)	495.8 (79.3)	492.5 (83.7)	487.2 (77.7)	488.3 (81.2)	507.0 (74.1)	415.2 (122.3)	499.3 (80.9)

Numeracy

	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Aust
Mean scale score / (S.D.)	498.3 (68.2)	501.7 (62.4)	490.4 (62.6)	488.6 (66.1)	477.0 (61.3)	481.3 (62.9)	497.5 (60.9)	432.1 (78.5)	493.8 (65.5)

Figure 3-3: Average NAPLAN scales scores of year 5 children in the five domains of literacy and numeracy in Australia, 2017

Source: NAPLAN 2017 report[268]

3.3 Data linkage

Data linkage brings together multiple sources of previously unconnected information on the same individuals. The data were de-identified to protect privacy, that's how the SA ECDP was granted ethics approval for data linkage. The multiple administrative datasets included in the SA ECDP[256] were linked by an independent data linkage agency - SA NT DataLink.[256, 269] After all the relevant ethics and custodian approvals are obtained, data custodians send the requested dataset to the data linkage agency with key identifying information including name, gender, date of birth, and address. This demographic information is used by the data linkage agency to probabilistically match children across different datasets and, after matching children, create an anonymising project linkage key. The data custodian then attaches the anonymising project specific key to their original detailed dataset and then remove all identifying information.[256] The de-identified datasets can then be provided to the researchers for analysis. During the data linkage process, the separation of the identifying data (name, gender, date of birth, and address), from the clinical or services data protects individual's privacy, while making population level data accessible for researchers.[270] Unlike many Scandinavian countries, in Australia there is no unique individual citizen identifier number that could be used for data-linkage. In Australia, information on the same individuals from multiple data sources are probabilistically matched and linked using basic demographic information, and it is possible that a small degree of incorrect linkage may happen. No studies to measure the false linkage rate have been undertaken in South Australia. Studies from New

South Wales and Western Australia have reported about 0.1-0.5% false linkage rate in Australian data linkage systems.[269, 271]

In the following sections (3.4, 3.5 and 3.6) I will describe the epidemiological concepts such as confounding, and the methods used to deal with confounding, and missing data; and the analytical methods used to answer each research question.

3.4 Confounding

Confounding is a source of bias in the effect estimate of the exposure on outcome, and corresponds to lack of comparability between the exposed and unexposed groups.[272-274] Due to confounding, the effect of the exposure of interest on the outcome is distorted because the effect of extraneous factors (called confounders) is mixed with any actual exposure effect.[272] In modern formal definitions, a confounder is a common cause of both the exposure and outcome.[272, 273] In observational studies treatment assignment is not randomized, therefore the observed effect estimates could be due to differences and imbalances in variables other than the exposure of interest being investigated.[272] Therefore, adjusting for confounders helps achieve conditional exchangeability, and gives an effect estimate which is due to the exposure of interest[272, 273, 275 276] However, adjustment for confounding is unlikely to ever perfectly occur in practice in observational studies.

In the following section I will discuss the different analytical approaches used in this doctoral thesis to deal with the issue of confounding including; the use of DAGs to identify confounders (in study 2, 3 and 4); and use of AIPW- a potential outcome approach as a different method to dealing with confounding (in study 4). I adjusted for a wide range of relevant confounders in each study in an attempt to reduce bias by confounding (discussed in chapters 4, 5, 6, and 7). I will also discuss the methods used in this thesis to assess the risk of bias due to unmeasured confounding including negative control outcome analysis and E-value.

3.4.1 Directed acyclic graphs (DAGs)

DAGs are an approach to constructing models to assess what should be adjusted for to achieve conditional exchangeability. Before carrying out any analyses in this doctoral thesis, a DAG was drawn to identify possible confounding pathways. (A simplified examples of a DAGs is given in Figure 3-4 below, with actual DAGs used in studies 2 and 3 are given in chapter 5, and 6 respectively).[277] DAGs are a simple, flexible tools for demonstrating the statistical associations between variables given a set of assumptions about the causal structure.[278-280] Depending on the research question and a priori subject matter knowledge, the causal relationship between exposures, outcomes and confounders, are shown in DAGS by single headed arrows. Arrows in DAGs depict the direction of a causal effect from one

variable to the other, either harmful or protective.[281] In Figure 3-4 below, the confounder is a cause of the exposure and the outcome, and this is why confounders are frequently referred to as 'common causes'. Variables that are in the causal pathway from the exposure to the outcome are called mediators; for example, birth weight is a mediator between maternal smoking and childhood type 1 diabetes. Adjusting for mediators can attenuate the total effect of the exposure on the outcome and considered over-adjustment bias in epidemiology[282], therefore I did not adjust for mediators in my studies.

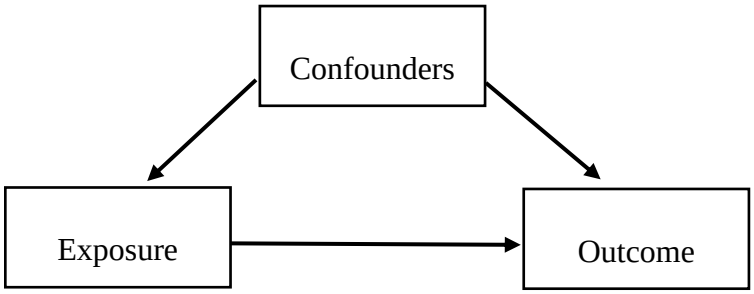


Figure 3-4: Directed acyclic graph (DAG)

3.4.2 Augmented inverse probability weighting

Conceptually the AIPW method fits under the potential outcomes theory, where the results may be interpreted as the education outcome as though the entire population were exposed to type 1 diabetes, versus their education outcome as though they did not have type 1 diabetes. The doubly robust AIPW procedure can have a similar interpretation to a randomised control trial (RCT) in the absence of unmeasured confounding.[250] In large RCTs the treated and

untreated are exchangeable because both the treated and untreated groups are similar across all measured and unmeasured characteristics due to the randomization. Therefore the difference between the treated and untreated in the outcome is considered to be due to the exposure.[276] Hence a causal effect of the treatment on the outcome could be estimated in a perfectly randomized RCT, given no loss to follow up and full compliance.[276] However, in observational studies, since treatment is not randomly assigned, the exposed and unexposed may differ on many characteristics and hence they may not be exchangeable, leading to confounding bias in the effect estimates. In conventional regression, adjustment for confounders is an attempt to obtain conditional exchangeability. Only in potential outcome approach and AIPW a hypothetical population is created called the pseudo population, where every child appears both as exposed (type 1 diabetes) and unexposed (no type 1 diabetes). It can't happen in real life, that is why it is called the *potential outcome* – the outcome that child would have had, had the child been assigned to the alternative exposure (or treatment if it was a RCT). The pseudo population is created by weighting each child in the study by the inverse of the conditional probability of receiving the exposure given covariates.[250, 283, 284] In other words, inverse probability weighting removes confounding by comparing the same child under two different exposure conditions.[250, 283, 284]

AIPW is a potential outcome approach to deal with confounding beyond traditional regression adjustment. Augmented Inverse Probability Weighting (AIPW) is also known as a doubly robust method[250] because it allows the specification of a treatment model for computing the probability to be exposed, given covariates, and the outcome model; and gives

a robust estimate if one of those models is correctly specified. In study 4 of this thesis, to computed the average treatment effect (ATE) of type 1 diabetes on educational outcomes of children in year 5, two models were specified; the treatment model (logit) for computing probability of type 1 diabetes given observed covariates; and the outcome model (linear). A priori identified confounders (socioeconomic and perinatal characteristics mentioned in Table 7.1 of chapter 7) were included in both the outcome and treatment model, based on DAGs.

3.4.3 Negative control outcome analysis

Negative control outcome analysis is a tool used in epidemiology to detect the risk of bias in the effect estimate due to unmeasured confounding and other threats to the validity of causal inference.[285] Negative controls in epidemiological studies are similar to those in experimental studies conducted in laboratories, where researchers leave out an essential ingredient or inactivate the active ingredient, to test an effect that would be implausible by the hypothesized mechanism.[286] In the third study of this thesis, I was particularly concerned about the potential for residual confounding. A negative control outcome analysis[285] offered the potential to detect the presence of residual or unmeasured confounding.

Most previous studies had reported a protective effect of maternal smoking for childhood type 1 diabetes but many of these did not adjust for some important confounders. I also acknowledge that I did not have information on father's type 1 diabetes, which is potentially an important confounder of the maternal smoking and childhood type 1 diabetes association. Therefore, I estimated the risk of bias due to unmeasured confounding in the effect of maternal smoking on childhood type 1 diabetes using a negative control analysis.

One of the assumptions underlying choosing a negative control outcome is to select an outcome which is not directly caused by the exposure.[285] Ideally, the association between the exposure and the–negative control outcome association should have similar set of measured and unmeasured confounders as the exposure and outcome of interest (Figure 3-5). This condition is called perfect U-comparability.[285] In routinely collected observational data finding a perfectly U-comparable negative control outcome is difficult, and the negative control outcome is more likely to only achieve approximate U-comparability with the primary outcome of interest.[285] Looking across all available variables in the linked datasets, I decided to use a variable "children not having a school card" as the negative control outcome. This information was obtained from school enrolment census. A school card is given to socioeconomically disadvantaged Australian children for financial assistance with school fees and does not have any direct plausible relationship with maternal smoking in pregnancy.

As shown in Figure 3-5, socioeconomic position is a potential confounder (C) of the maternal smoking (X) and childhood type 1 diabetes (Y) association. This is because socioeconomic position causes both maternal smoking and childhood type 1 diabetes. Socioeconomically disadvantaged mothers are more likely to smoke in pregnancy.[253] And socioeconomically advantaged children are more likely to have childhood type 1 diabetes. Socioeconomic position also acts as a confounder in the negative control outcome. Similarly, socioeconomically advantaged children are more likely to "not have a school card".

However, there is no plausible association between maternal smoking and "child not having a school card", and as can be seen in Figure 3-5 there is no direct arrow leading from maternal smoking to "child not having a school card". Therefore, if after adjusting for the measured confounders, any observed effect of maternal smoking on the negative control outcome must be due to back door paths or unmeasured confounding (U) (Figure 3-5). In practice there would be many arrows relevant to each confounder, but for simplicity all of the measured confounders have been bundled up together in Figure 3-5. In Figure 3-5, the red and orange arrows from measured and unmeasured confounders going toward maternal smoking and type 1 diabetes are depicting backdoor or indirect path (confounding path). A backdoor path is a non-causal path from the exposure to the outcome, and indicate common causes (confounders) of the exposure and outcome.[280]

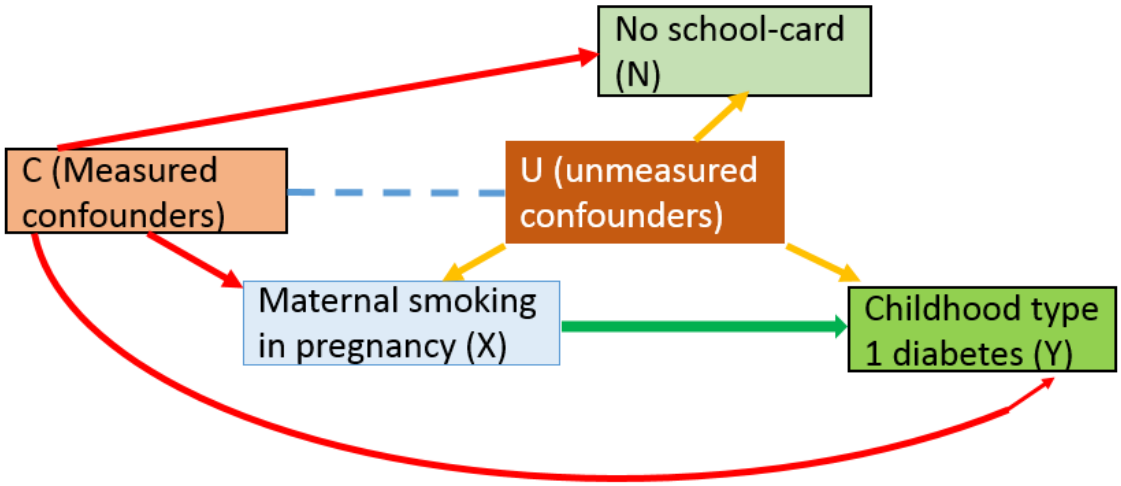


Figure 3-5: Directed acyclic graph (DAG) showing negative-control outcome analysis

Measured confounders (C): Parents' age, parents' occupation, mother's birth region, maternal ethnicity, remoteness, IRSAD (Index of Relative Socio-economic Advantage and Disadvantage), mother type of patient (health care), private or public hospital for child birth, parity, pre-pregnancy hypertension, and diabetes

3.4.4 E-value

The potential for bias from factors that were not measured (unmeasured confounding) is a common issue in observational studies. However, methods have been developed to perform sensitivity analyses to quantify this bias. One of the simplest approaches proposed by Ding and VanderWeele, is calculating the E-Value.[287, 288] The E-value quantifies the minimum strength of association (on a risk ratio scale) that an unmeasured confounder needs to have with the exposure and the outcome to explain away the observed effect.[287, 288] Haneuse, VanderWeele and David stated that "*E-value analysis asks the question: how strong would the unmeasured confounding have to be to negate the observed results?"*[287, 288] *The* E-value also assesses the extent of unmeasured confounding that would be needed to shift the confidence intervals to include the null.[288] E-value is another method for understanding confounders in causal analyses; and it assesses how much evidence there is for causality; the outcome is caused by the exposure of interest, and not attributed to other unmeasured confounders.

In study 3 (section 3.6.4) I calculated E-value for the hazard ratio of maternal smoking in pregnancy on childhood type 1 diabetes following Ding and VanderWeele, in Stata[287-289] Although the E-value is on a risk ratio scale, I used hazard ratio scale because my effect estimates were hazard ratios.

3.5 Multiple imputation for missing data

Missing data can result in biased estimates of the effect and can lead to loss of power and precision.[290, 291] A number of techniques have previously been used to deal with missing data including; extreme case analysis (imputing values with best or worst possible values); overall mean imputation (replacing missing data with the mean values of the observed data); or last value forward method (where the last measured value is used for replacing the missing data).[290 292-294] Other methods have been based on inverse probability weighting to make the complete case representative of all the cases; and approaches based on maximum likelihood have also been used to deal with missing data.[290 292, 293] However, the single value imputation techniques have limitations and may generate biased estimates as they do not take into account any uncertainty around the imputed missing value.[292, 294] Among the statistical approaches to analyse missing data, multiple imputation has been shown to be an efficient and least biased approach compared with complete-case and extreme-case analysis[294-296]

First introduced by Rubin,[297, 298] multiple imputation is a flexible way of dealing with missing data, and it creates multiple copies (each copy has a different estimated value for missing data) of the dataset to allow for the uncertainty about the missing value.[290, 293] Multiple imputation has been used to improve the validity of the estimates, to increase precision and power, and to avoid bias due to the loss of information.[290, 293, 299] However, it is important to carefully think about the reasons of missing data before deploying any method to deal with missing data. There are three assumptions about the mechanism for missing data; 1) missing

completely at random (MCAR); 2) missing at random (MAR); and 3) missing not at random (MNAR).[290, 293, 300] The data are said to be MCAR when the missing data is neither predicted by the observed data nor dependent on the missing data.[300, 301] In case of MCAR, there are no systematic differences between the observed and missing data, hence complete case analysis does not give a biased estimate.[300, 301] The MAR assumption implies that missing data can be predicted by the observed data.[299] There might be systematic differences between the observed and missing values, but these differences can be explained by other observed variables, hence multiple imputation can be used under the MAR assumption.[290, 301] The MNAR assumption implies that missingness depends on the missing data, and could not be predicted using the observed data.[290] MNAR can not really be tested, although the MAR assumption holds if missingness is predicted by measured variables.

Multiple imputation could be performed either by multiple imputation using multivariate normal distribution (MVN) or by multiple imputation by chained equation (MICE).[292] MICE is a flexible approach to impute categorical variables.[299] In MICE different regression models (logit, multinomial, linear) can be specified for each variable depending on the type of variables (binary, categorical, ordinal, continuous) included in the imputation model.[292 299] Whereas MVN assumes a joint multivariate normal distribution, and needs variables to be normally distributed or transformed to approximate normality.[292] In this doctoral thesis, MICE was performed given the different types of variables included in the datasets.[292, 299] Multiple imputation was conducted in all four studies (Chapters 4-7) to deal with missing data. Detail about the proportion of missing data is given in each Chapter 4, 5, 6, and 7.

Analyses involving multiple imputation were conducted in three phases: an imputation phase, analysis phase, and pooling phase. In the imputation phase, twenty datasets were created, with fifty iterations each or fifty cycles of regressions before switching to create the next imputed dataset.[293] At the end of 50 cycles of regression, one imputed dataset is created and saved, and this process is repeated for generating all twenty datasets. Enders (2017) writes about the reason for creating multiple iterations as "*to avoid imputations based on a single set of regression parameters, an iterative algorithm uses Bayesian estimation to update the regression model parameters, and it uses new estimates to generate each set of imputations.*[293] Thus each of the twenty datasets contains a different set of imputed estimates for the missing values. The twenty datasets and the fifty cycles of regressions for each imputed dataset were generated to account for the uncertainty about the missing values.[290, 293] Then in the analysis phase, the twenty datasets were analysed separately using models appropriate for each question (question 1- 4). In the last pooling phase, the estimates and standard errors from the twenty datasets were combined using Rubin's rules.[298, 299] Rubin's rules take in to account between and within imputation variance.[293, 297, 298] For standard regression estimates statistical packages have written programs that combine the estimates across multiple imputed datasets. However, there were no packages available to pool incidence rates and average treatment effects from the AIPW results in paper 1 and paper 4. Therefore, I developed Stata syntax using Rubin's formula, to pool the incidence rates from the 20 imputed datasets in study 1 (Chapter 4), and to combine average treatment effects

from the twenty datasets in study 4 (Chapter 7). All the analyses were conducted on complete-case and imputed data.

3.6 Analytical approaches used in each research question

In the following section I will describe the analytical methods deployed to answer each research question. As mentioned above, children identified with type 1 diabetes from hospital data were merged with perinatal and birth registration data, and other required datasets depending on each research question for analyses (Table 3-1).

Table 3-1: *Summary of the datasets, year of data and analytical methods used in each study*

Chapter	Study	Exposure	Outcome	N	Data Sources	Years	Statistical methods
4	Socioeconomic characteristics of children with type 1 diabetes	Socioeconomic condition	Childhood type 1 diabetes	N = 231,685 T1D = 333	Perinatal Statistics Birth Registrations Hospitalizations	2002-2013 2002-2013 2002-2014	Multiple Imputation Incidence rates
5	Caesarean section and risk of type 1 diabetes	Caesarean birth (prelabour and intrapartum)	Childhood type 1 diabetes	N = 286,058 T1D = 557	Perinatal Statistics Birth Registrations Hospitalizations	1999-2013 1999-2013 2001-2014	Multiple imputation Cox regression
6	Maternal smoking in pregnancy and risk of type 1 diabetes	Maternal smoking in pregnancy	Childhood type 1 diabetes	N = 286,058 T1D = 557	Perinatal Statistics Birth Registrations Hospitalizations School Enrolment Census	1999-2013 1999-2013 2001-2014 2005-2015	Multiple imputation Cox regression Meta-analysis Negative control outcome E-value
7	Effect of type 1 diabetes on children's educational outcome	Childhood type 1 diabetes	Children's educational outcomes in year 5	N = 61,445 T1D =162	Perinatal Statistics Birth Registrations Hospitalizations School Enrolment Census School assessment	1999-2005 1999-2005 2001-2014 2008-2015 2008-2015	Multiple imputation AIPW (Augmented Inverse Probability Weighting)

3.6.1 Research question 1: What are the socioeconomic characteristics of children with type 1 diabetes?

In the first descriptive study (Chapter 4), I calculated incidence rates of type 1 diabetes by individual and area-level socioeconomic indicators, among children in South Australia, born from 2002-2013. Although hospitalization data in the SA ECDP started July 2001, births registration and perinatal data started from 1999. In order to keep the same period of births and hospitalisation data and avoid misclassification of children with type 1 diabetes born before 2002, I used data for the 2002-2013 birth years and the 2002-2014 hospitalization calendar years for study 1.

As mentioned in chapter 2 , socioeconomic position is one of the major determinants of health and a fundamental cause of health inequalities that exist between people with and without socioeconomic resources.[302] Socioeconomic position can influence people's lifestyle, health behaviours, healthcare and environmental exposure, which may have implications for type 1 diabetes. Therefore, the purpose of this study was to describe socioeconomic patterning of childhood type 1 diabetes incidence in South Australia. South Australia is the most disadvantaged mainland state of Australia, and to date there has been no population-level study of type 1 diabetes among South Australian children. Mostly previous studies of type 1 diabetes in childhood have used area-level or country-level measures of socioeconomic position and as described in section 2.7.1, the associations with type 1 diabetes incidence

have been inconsistent.[95, 98, 99, 105, 151, 303, 304] However, very few studies reported individual-level socioeconomic indicators and type 1 diabetes incidence.[32, 155] Therefore, in the first study I compared and described individual and area-level socioeconomic pattering of type 1 diabetes incidence.

For this study, children with type 1 diabetes were identified from hospital admissions data from July 2002 – June 2014 using ICD-10-AM diagnoses codes as described above. The hospital admissions data was merged with the perinatal statistics and births registration data of children. Birth registration data was used to supplement any missing information collected in the perinatal data. The follow-up time for each child started from birth and ended at diagnosis for children with type 1 diabetes or was censored at the end of follow-up (June 2014) for children without type 1 diabetes. For example, children born in 2002 were followed for eleven years and children born in 2013 were followed for one year. The mean follow-up time was 6 years. All children (n = 231,685) born from 2002-2013, contributed to 1,443,756 person-time at risk, and 333 children were diagnosed with type 1 diabetes during the follow-up from 2002-2014.

The age-specific incidence of type 1 diabetes (per 100,000 population) was calculated by using total number of children in each age-group as the denominator. Due to the differences in the follow up time for each child, incidence rate (per 100,000 person-years) by individual and area-level socioeconomic characteristics was calculated using total person-time at risk as the denominator. Type 1 diabetes incidence rates were calculated by individual-level

socioeconomic indicators such as parent's employment, whether mother was a private or public patient (health insurance) at the time of childbirth, or whether the child was delivered in a private or public hospital. In addition, I also calculated type 1 diabetes incidence rate by area-level socioeconomic position, which was measured by the Index of Relative Socioeconomic Advantage and Disadvantage (IRSAD). The individual-level socioeconomic indicators considered here (parents' employment, private or public healthcare, private or public hospital birth) are well known and widely used internationally. However, since the IRSAD measure is unique to Australia, I will describe it further in the following section.

3.6.1.1 IRSAD: An area-level measure of socioeconomic position

The Index of Relative Socio-economic Advantage and Disadvantage (IRSAD) was used as an area-level measure of socioeconomic position in this study.[305] IRSAD is developed by the Australian Bureau of Statistics using five-yearly census data, to assess people's access to resources and their ability to participate in society.[305] IRSAD ranks areas from most disadvantaged to most advantaged. As commonly practiced in Australia, I used the IRSAD quintiles; quintile 1 represented most disadvantaged, and quintile 5 depicted most advantaged areas. IRSAD is constructed based on a weighted combination of selected variables from the Census such as income, education, occupation, housing and others, using principal component analysis (detail of variables in Table 3-2).[305] Each child born in South Australia included in the perinatal dataset was assigned an IRSAD score based on their mother's postcode at the time of birth.

Table 3-2: Census variables used by the Australian Bureau of Statistics for creating the Index of Relative Socio-economic Advantage and Disadvantage (IRSAD)

Low income	% People with stated annual household equivalised income between $1 and $20,799 (approx. 1st and 2nd deciles)
No internet	% Occupied private dwellings with no internet connection
Education <year 11	% People aged 15 years and over whose highest level of education is Year 11 or lower. Includes Certificate I and II
Child living with jobless parents	% Families with children under 15 years of age who live with jobless parents
Occupation (labour)	% Employed people classified as 'labourers"
One parent families	% One parent families with dependent offspring only
Unemployed	% People (in the labour force) unemployed
Disability (age <70 years)	% People aged under 70 who have a long-term health condition or disability and need assistance with core activities
Low rent	% Occupied private dwellings paying rent less than $166 per week (excluding $0 per week)
Separated /divorced	% People aged 15 and over who are separated or divorced
Occupation	Employed people classified as Machinery Operators and Drivers
Low skill workers	% Employed people classified as Low Skill Community and Personal Service Workers
No car	% Occupied private dwellings with no cars
Overcrowded households	% Occupied private dwellings requiring one or more extra bedrooms (based on Canadian National Occupancy Standard)
No education	% People aged 15 years and over who have no educational attainment
3 or more cars	% Occupied private dwellings with three or more cars
At university	% People aged 15 years and over at university or other tertiary institution

Spare bedroom	% Occupied private dwellings with one or more bedrooms spare
High rent	% Occupied private dwellings paying rent greater than $370 per week
Occupation (managers)	% employed people classified as Managers
4 or more bedrooms	% Occupied private dwellings with four or more bedrooms
Occupation (Professionals)	% Employed people classified as Professionals
Diploma	% People aged 15 years and over whose highest level of education attainment is a diploma qualification
High mortgage	% Occupied private dwellings paying mortgage greater than $2,800 per month
High income	% People with stated annual household equivalised income greater than $52,000 (approx. 9th and 10th deciles)

Source: Australian Bureau of Statistics. Socio-Economic Indexes for Areas (SEIFA) technical report 2011[305]

3.6.2 Cox proportional hazard regression

In study 2 and 3, Cox proportional hazard regression was used. Cox proportional hazard regression is used to estimate the effect of an exposure on the outcome; when the outcome is time to event or survival data. In the second and third study of this thesis, I used Cox proportional hazard regression to estimate the effect of caesarean birth and maternal smoking on childhood type 1 diabetes. I used Cox regression to account for different time to diagnosis, and different length of follow-up for each birth cohorts. One of the assumptions of Cox proportional hazard regression is that the hazard ratio is constant over time, or the effect of given covariate does not change overtime.[306] Therefore, before conducting Cox proportional hazard regression analyses in study 2 and 3, I tested the proportional hazard assumption by Schoenfled residual test,[306, 307] which showed non-violation of the proportional hazard assumption.

3.6.3 Research question 2: What is the effect of caesarean birth on childhood type 1 diabetes? Does the risk of type 1 diabetes differ by prelabour and intrapartum caesarean?

A range of perinatal factors have been linked with type 1 diabetes. Caesareans that are not medically indicated are potentially preventable, therefore if caesareans increase the risk of type 1 diabetes they may represent a modifiable risk factor.[308] However evidence about the impact of caesarean birth on childhood type 1 diabetes is mixed, [20-22, 25, 164] and less is known about whether type 1 diabetes risk differs between prelabour and intrapartum caesarean births.[21] Therefore in the second study (Chapter 5) I estimated the effect of caesarean birth on childhood type 1 diabetes, extending the question to whether the risk of type 1 diabetes differed by caesarean type; intrapartum or pre-labour caesarean. As mentioned in chapter 2, birth method induced disparity in gut microbiota is a hypothesized link between caesarean birth and childhood type 1 diabetes.[134] For this research question, it was hypothesized that type 1 diabetes risk would be higher among children born by prelabour caesarean than intrapartum caesarean births. That is because children born by prelabour caesarean do not get exposed to maternal vaginal microbiota, and intrapartum caesareans births presumably have some exposure.

Children with type 1 diabetes identified from hospital data using ICD-10-AM codes (2001-2014) were merged with perinatal statistics and birth registration data (1999-2013) for the analysis of this question (Table 3-1). Data on exposure (caesarean birth) and confounders were sourced from perinatal statistics. Information on birth method (normal spontaneous and instrumental vaginal deliveries, elective and emergency caesarean) and onset of labour (spontaneous, no labour and induction) were combined to classify prelabour or intrapartum caesarean.[259] Prelabour caesarean included all caesareans performed before the onset of

spontaneous or induced labour, and intrapartum caesarean included all emergency and elective caesareans performed after the onset of labour (spontaneous or induced labour).

For this study, children were followed from birth until type 1 diabetes diagnosis, or until censored at the end of follow up (June 2014). The follow up time ranged from one year (for 2013 births) to 15 years (for 1999 births), with a mean follow up of eight years. Due to these differences in follow-up time, Cox proportional hazard regression was used to estimate the incidence of type 1 diabetes for children who had caesarean birth compared with normal vaginal delivery (primary analysis). In secondary analysis, the crude and adjusted effect of intrapartum and prelabour caesarean birth on childhood type 1 diabetes was estimated, compared with normal vaginal delivery, using Cox proportional hazards regression. Both primary and secondary analyses were adjusted for a wide range of a priori identified confounders, in order to reduce confounding bias. The confounders were identified based on subject matter knowledge and previous research, using a DAG (Figure 5-2). Detail about the confounders for this study is given in Chapter 5.

In addition to the primary and secondary analyses, five different sensitivity analyses were performed, to investigate the robustness of the effect estimate, which are discussed in Chapter 5. Multiple imputation by chained equation was used to impute the missing data (discussed above in section 3.5).[290] Analyses were performed on both complete-case and imputed data.

3.6.4 Research question 3: What is the effect of maternal smoking in pregnancy on the risk of childhood type 1 diabetes? What is the risk of bias due to residual confounding in the effect estimate?

Similar to caesarean birth, maternal smoking in pregnancy is a modifiable risk factor for childhood type 1 diabetes, however the evidence is mixed. Maternal smoking in pregnancy has been linked with both higher and lower risk of childhood type 1 diabetes.[26, 29-31, 34] In the third study an extensive and rigorous epidemiological approach was deployed to understand the association between maternal smoking in pregnancy and childhood type 1 diabetes. Firstly, in the primary and secondary analyses, I estimated the effect of maternal smoking on childhood type 1 diabetes using whole-of-population South Australian linked data. Secondly, because of the small number of children exposed to maternal smoking in previous studies as well as in my dataset, in order to get a more precise estimate, I performed meta-analyses of the findings from published studies and my study. Thirdly, I assessed the risk of bias due to residual or unmeasured confounding in the estimate of maternal smoking on childhood type 1 diabetes in my findings of the South Australian linked data study using negative control and E value approaches.

3.6.4.1 Primary and secondary analyses

Primary and secondary analyses involved use of multiple administrative datasets to estimate the effect of maternal smoking in pregnancy on childhood type 1 diabetes. For this study,

similar to study 2, children identified with type 1 diabetes from hospital data were merged with the perinatal statistics and birth registration data (1999-2013). Data on maternal smoking in pregnancy and confounders were sourced from perinatal statistics. Similar to question 2, children were followed from birth until type 1 diabetes diagnosis, or until the end of follow up in June 2014, with the mean follow up of eight years (follow up ranged 1-13 years).

In the primary analysis, crude and adjusted risk of childhood type 1 diabetes for children exposed to maternal smoking in pregnancy was estimated by Cox proportional hazard regression, compared with unexposed children. A priori identified confounders were included in the adjusted model, to reduce confounding bias (Figure 6-2). Detail about these confounders are discussed in Chapter 6. Information on maternal smoking collected in the first and second half of pregnancy were combined to create the maternal smoking during pregnancy exposure. For the primary analysis, maternal smoking was categorised as; 1) non-smokers, 2) those who smoked only in the first or second half of pregnancy, 3) consistent smokers. In order to get a cleaner effect of consistent maternal smoking throughout pregnancy, the children who were only exposed to maternal smoking for short time (only in the first or the second half of pregnancy) were grouped separately. Those women who reported smoking in both the first and second half pregnancy of pregnancy were categorized as “consistent smokers”. I wanted to look at the impact of maternal smoking only in the first half or second half of pregnancy separately, but the numbers were too small to estimate this

effect. Therefore, due to small numbers, smoking only in the first or second half of pregnancy category was created by combining the following four groups;

1. women who quit smoking in the first half of pregnancy and never smoked again in pregnancy (n = 10,605, type 1 diabetes = 22),
2. women who quit smoking at/before first antenatal visit but resumed smoking later in pregnancy (n = 400, type 1 diabetes = 0),
3. women who reported smoking only in the first half of pregnancy (total n = 3,869, type 1 diabetes = 8),
4. Women who reported smoking only in the second half of pregnancy (n = 316, type 1 diabetes = 1).

A total of 316 women reported smoking only in the second half of pregnancy from 1999-2013. There were no reports of smoking only in the second half of pregnancy in 2008, 2009, 2010, 2011, and 2013. As can be seen from the above-mentioned numbers in each group, the majority of women quit smoking in the first half of pregnancy and never smoked again in pregnancy. Therefore, exposure to maternal smoking was mostly limited to the first half of pregnancy, among those children whose mothers were categorized as “smokers only in the first and second half of pregnancy”.

For the secondary analysis, the crude and adjusted effect of any smoking in pregnancy was estimated, compared with non-smokers, using Cox proportional hazard regression. Due to small numbers of children exposed to maternal smoking, those who smoked only in the first

and second half of pregnancy, and consistent smokers were combined to create "any smoking in pregnancy" category.

In addition to the primary and secondary analyses, five sensitivity analyses were performed, details are discussed in chapter 6. Multiple imputation was performed to impute the missing data (discussed in section 3.5)[290], and analyses were conducted on both complete-case and imputed data.

3.6.4.2 Meta-analyses

It was evident from the published literature that even the large studies on maternal smoking and childhood type 1 diabetes have imprecise estimates, and this is also true for my study. Previously published population-based studies also had small numbers of children exposed to maternal smoking in pregnancy that impacted their precision.[30] There was no published meta-analysis on maternal smoking and childhood type 1 diabetes when this study was conducted. Therefore, in order to get a more precise estimate of the effect of maternal smoking in pregnancy on childhood type 1 diabetes, findings from my study and previous case-control and population-based cohort studies were meta-analysed.[26, 27, 29-31, 34] I searched Web of Science, PubMed, and EMBASE databases systematically using the terms type 1 diabetes, insulin depended diabetes mellitus, childhood diabetes, maternal smoking in pregnancy, prenatal, environmental and perinatal factors. While studying the literature for my doctoral thesis I observed that the above mentioned terms were commonly used in the

papers to report the incidence and perinatal risk factors of type 1 diabetes. Therefore, these terms were chosen to search for studies on maternal smoking and childhood type 1 diabetes.

Literature showed that most cohort studies are conducted in high risk population to explore the risk factors of type 1 diabetes and mostly reported multiple risk factors and preclinical type 1 diabetes as their outcome. None of those studies had specifically focused on maternal smoking as their main exposure of interest, and maternal smoking was only reported as one of the baseline characteristics. Therefore, I chose only those studies for meta-analysis that had reported clinical type 1 diabetes as their outcome and had adjusted for some confounders in their estimate of the effect of maternal smoking on childhood type 1 diabetes. Case-control studies reported odds ratios, and population-based cohort studies reported hazard ratios. Therefore, I performed meta-analyses separately for case-control,[26, 31-33, 309, 310] and population-based cohort studies,[27, 29, 30] using random effects models, assuming non-homogeneity among studies.

3.6.4.3 Assessing risk of bias due to unmeasured confounding

As mentioned above, unmeasured confounding is a major issue threatening the validity of the effect estimate in observational studies. I assessed the risk of bias due to residual confounding in the effect estimate of maternal smoking on childhood type 1 diabetes (in my original findings obtained from SA ECDP data) using negative control outcome analysis (discussed above in section 3.4.3).[285] I also estimated the risk of bias due to unmeasured confounding in the effect estimate of maternal smoking on childhood type 1 diabetes (my

original findings) using E-value (discussed above in section 3.4.4) following Ding and VanderWeele, in Stata[287, 288] 289

3.6.5 Research question 4: What are the implications of type 1 diabetes for children's educational outcomes?

I estimated the average treatment effect (ATE) of type 1 diabetes on children's educational outcomes in year 5, using augmented inverse probability weighting method (AIPW, discussed above in section 3.4.2).[250] For this study, children identified with type 1 diabetes from hospital data (2001-2014) were merged with school enrolment, school assessment data (2008-2015), perinatal statistics and births registration data (1999-2005). Data on educational outcomes of children born 1999-2005 were sourced from the nationally administered school assessment data, called NAPLAN[268] collected from 2008-2015. Although there were 557 children identified with type 1 diabetes from 2001-2014 from hospitalisation data, only 162 children identified with type 1 diabetes had taken year 5 NAPLAN assessment after their diagnosis and were thus included in this study. A large number of children with type 1 diabetes had taken year 5 NAPLAN assessment, compared with NAPLAN assessments in other year levels (year 3, 7 and 9) in the SA ECDP dataset. Therefore, year 5 NAPLAN assessment was the outcome investigated in this study. In this project I used NAPLAN scale

scores of children as an outcome, keeping in mind that fluctuations in children's blood glucose might be reflected in small variations in continuous NAPLAN scale scores.

CHAPTER 4 INCIDENCE OF TYPE 1 DIABETES BY SOCIODEMOGRAPHIC CHARACTERISTICS AMONG SOUTH AUSTRALIAN CHILDREN: WHOLE-OF-POPULATION STUDY

4.1 Preface

Chapter 4 contains the first epidemiological descriptive study contributing to this thesis. (This paper was under review in the Journal of Paediatrics and Child Health, when the thesis was submitted for examination. It was accepted for publication while the thesis was under examination).

This paper addresses the first question of my doctoral project.

> ***What are the socioeconomic characteristics of children with type 1 diabetes in South Australia born from 2002-2013?***

My doctoral studies began by examining the incidence and socioeconomic characteristics of South Australian children who developed type 1 diabetes. Understanding the basic epidemiological patterns of type 1 diabetes was the logical first step, before exploring in-depth other risk factors and implications of type 1 diabetes. Patterns of incidence of type 1 diabetes in children has never been studied in South Australia before.

Most previous studies (including other Australian studies) have used area-level socioeconomic position to describe the socioeconomic patterning of type 1 diabetes incidence. Only two non-Australian studies have reported individual level socioeconomic patterning of type 1 diabetes in incidence. Area-level measures many not reflect individual-level variation in socioeconomic condition, and are prone to ecological fallacy. Therefore, I took advantage of the linked SA ECDP data with information on individual level socioeconomic position, to estimate the incidence of type 1 diabetes by individual and area-level measures of socioeconomic position.

Highlights

- In this study, individual-level socioeconomic factors supported the hygiene hypothesis, which links more advantaged socioeconomic circumstances to greater risk of type 1 diabetes.

- No association was found between area level socioeconomic position measured using the Index of Relative Socioeconomic Advantage and Disadvantage (IRSAD). The IRSAD applies combinations of different socioeconomic measures from households in the area to individuals.

4.2 PUBLICATION 1: Incidence of type 1 diabetes by sociodemographic characteristics among South Australian children: whole-of-population study

4.2.1 Statement of Authorship

Title of Paper	**Incidence of type 1 diabetes by sociodemographic characteristics among South Australian children: whole-of-population study**
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	

Principal Author

Name of Principal Author (Candidate)	Mumtaz Begum
Contribution to the Paper	MB prepared the draft, conducted the analyses, interpreted the results and helped in study design.
Overall percentage (%)	
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a

	third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	17/01/2020

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated

Name of Co-Author	Catherine Ruth Chittleborough		
Contribution to the Paper	CRC provided critical input in the study design, data analysis and interpretation of findings, and reviewed and revised the draft		
Signature		Date	17/1/2020

Name of Co-Author	Rhiannon Megan Pilkington		
Contribution to the Paper	RMP conceptualized the project, assisted in data acquisition and analysis and interpreting findings, and reviewed and revised the draft.		
Signature		Date	17/1/2020

Name of Co-Author	Murthy Mittinty
Contribution to the Paper	MM assisted with data analysis, interpretation of findings and revision of draft.

Signature		Date	15/01/2020

Name of Co-Author	John Lynch		
Contribution to the Paper	JL was responsible for acquisition of data, provided intellectual input for study design, data analysis and interpretation, and critically reviewed the manuscript.		
Signature		Date	17/1/20

Name of Co-Author	Megan Penno		
Contribution to the Paper	MP provided intellectual input for study design, interpretation of findings, and critically reviewed the manuscript.		
Signature		Date	14/01/2020

Name of Co-Author	Lisa Gaye Smithers		
Contribution to the Paper	LGS provided guidance in the study design and data analysis, assisted in data acquisition and interpreting the results, and reviewed and revised the manuscript.		
Signature		Date	17 JAN 2020

What is already known?

- Past studies have reported mixed evidence about the association between various area-level socioeconomic indicators and the incidence of T1D.
- Previous studies reported higher type 1 diabetes incidence in affluent areas. However, much less is known about individual-level patterning of socioeconomic characteristics of children with T1D.

What this study adds?

- Area and individual-level measures of socioeconomic circumstances were not consistently associated with T1D incidence rates in South Australia.
- Higher T1D incidence rates were observed in the most advantaged groups of individual-level socioeconomic indicators (both parents employed, mother Caucasian, mother had private healthcare).
- There was no clear area-level socioeconomic patterning of T1D incidence.

4.2.2 Abstract

Objective

To describe and compare the incidence of Type 1 diabetes (T1D) in South Australia by individual and area-level socioeconomic characteristics among children aged ≤11 years.

Design

Whole-of-population, data linkage study (n=231,685).

Setting

South Australia, children born from 2002-2013, hospitalization followed from 2002-2014.

Data Source

De-identified, linked administrative hospitalization, birth and perinatal data from the South Australian Early Childhood Data Project.

Outcome Measure

Incidence was calculated by identifying T1D cases from T1D-related hospitalizations using ICD-10-AM diagnosis codes (E10, E101-E109).

Results

Overall, 333 children aged ≤11-years (173 boys) were identified as having T1D. The T1D incidence rate was 23.0 per 100,000 person-years (95% CI: 20.7-25.7), with no sex difference.

T1D incidence was higher among children whose mothers were Caucasian, private patients, and whose parents were employed. For example, T1D incidence was 26.0 per 100,000 (95% CI: 22.8-29.5) among children with both parents employed, compared to 20.0 per 100,000 (95% CI: 12.3- 30.6) among children with both parents unemployed.

There was no clear gradient in the association between area-level socioeconomic position and T1D, with highest incidence for the fourth quintile (26.5 per 100,000 [95% CI: 20.9-33.1]). The most advantaged area (19.4 per 100,000 [95% CI: 13.8-26.5]) had lower incidence than the most disadvantaged area (23.5 per 100,000 [95% CI: 18.9-28.9]).

Conclusion

T1D incidence rates differed depending on the measures of socioeconomic characteristics. Individual-level indicators showed higher incidence among more advantaged children, however, there was no clear area-level socioeconomic patterning of T1D.

Key words: Endocrinology, Adolescent, General Paediatrics

4.2.3 Introduction

Type 1 diabetes (T1D) is a chronic childhood disease and incidence is increasing in many countries.[54] In 2015, the International Diabetes Federation reported more than half a million children aged 0-14 years globally have T1D and 86,000 new cases were diagnosed annually.[311] The World Health Organization's DIAMOND project highlighted wide variation in age standardized incidence of T1D; from 0.1 per 100,000/year in China to 40.9 per 100,000/year in Finland, with around 3% annual increase in incidence from 1990-1999.[6] Australia ranks ninth highest among countries with published T1D incidence (22.5 per 100,000 population) and tenth highest for prevalence among OECD countries with 6,091 children aged 0-14 years with T1D.[54] Although T1D is characterized by autoimmune destruction of insulin producing pancreatic β-cells, the exact cause remains unknown. Studies attribute increasing incidence to environmental factors in early life interacting with genetic predisposition.[106] Socioeconomic conditions influence the environment of an individual, and hence may alter susceptibility to T1D.

There is mixed evidence regarding the association between national, regional and neighbourhood level socioeconomic indicators and T1D incidence. Studies using different socioeconomic measures from Poland (country-level gross domestic product, access to water supply and sewage system, and life expectancy)[105], Northern Ireland (area's population density and household crowding)[95], and UK (area's population density and Townsend deprivation score)[151] have reported high T1D incidence in affluent areas. A US population-

based study demonstrated that neighbourhood-level affluence indicators (household income, vehicle ownership, high education) are linked with higher T1D risk.[96] The hygiene hypothesis has been highlighted as one factor that drives higher T1D rates in advantaged areas[106], but the evidence for higher T1D incidence in affluent areas is contentious. Evidence from Germany[304] showed high T1D incidence in the most deprived areas. However, some studies demonstrated no socioeconomic patterning of T1D. For example, a study in Western Australia (WA) based on diabetes-register data (aged <15 years) found no association between area-level socioeconomic disadvantage and T1D incidence.[152]

All these studies have used different indicators to measure national, regional and neighbourhood socioeconomic conditions, and demonstrated inconsistent evidence. Area-level studies may not represent individual-level risk factors. Therefore, individual-level studies are needed to see whether individual socioeconomic conditions have any role in influencing T1D risk. Only two studies in Italy[155] and Washington[32] focused on individual-level socioeconomic characteristics and T1D incidence and reported lower T1D risk in disadvantaged children. No previous Australian study has focused on individual-level socioeconomic measures and T1D incidence. Therefore, this cohort study describes the incidence of T1D in South Australia (SA) by individual and area-level socioeconomic characteristics among children, born from 2002-2013.

4.2.4 Methods

4.2.4.1 Data source and participants

This population-based study used linked, de-identified government administrative data from the South Australian Early Childhood Data Project (n~280,000).[256] Children born from 2002 to 2013 were followed for hospitalization from 2002 to 2014 (aged ≤11 years-old).

Routinely collected birth registration, perinatal, and hospitalization data were probabilistically linked by an independent linkage agency (SA-NT DataLink; www.santdatalink.org.au, accessed 3rd December 2019) using demographic information including name, birth date, sex and address. Researchers receive de-identified data from custodians following the data linkage. Australian data linkage systems (SA-NT DataLink; www.santdatalink.org.au, accessed 3rd December 2019) typically estimate a false linkage rate of 0.1-0.5%.

4.2.4.2 Type 1 diabetes

Hospitalization data recording all admissions to public hospitals in SA was used to identify T1D incident cases. Children with their first T1D-related hospitalization were identified, based on the International Classification of Disease, Australian-modification (ICD-10-AM), 10th edition codes (E101 to E109), using both primary and additional diagnoses. ICD-10-AM codes are a hospital reporting tool assigned to patient records by trained staff following

discharge. Children with neonatal diabetes and who had diabetes secondary to other causes were classified as not having T1D.

4.2.4.3 Socioeconomic characteristics

Information on socioeconomic characteristics was sourced from the South Australian Perinatal Statistics Collection, for all births in SA from 2002-2013, which we validated and supplemented with Births Registration data.[256] The perinatal statistics collection at the time of birth (home and hospital) by midwives/neonatal nurses is mandatory for every birth in SA. Births registration data is collected as a part of the SA Births, Deaths and Marriages registry, and all births are legally required to be registered within 60 days of birth. Birth registration data includes parental and child demographic information and basic clinical birth data.

Individual-level socioeconomic variables included parents' employment status (both unemployed, one parent employed, both employed), type of hospital where child was born (public or private), and whether mother was a public patient or a private patient (with private health insurance) at time of delivery. Demographic and behavioural characteristics include maternal ethnicity (Caucasian, Aboriginal or Torres Strait Islander, Asian and other), and maternal smoking at first antenatal visit (yes/no), respectively.

Area-level socioeconomic variables were derived using the mother's postcode of residence at the child's birth. Living in a remote or accessible area (major cities, inner regional, outer regional or remote) was coded using the Accessibility and Remoteness Index of Australia. The Index of Relative Socioeconomic Advantage and Disadvantage (IRSAD) was used as a

neighbourhood-level summary measure of socioeconomic conditions. IRSAD was created by the Australian Bureau of Statistics (ABS) from Census information collected in 2001, 2006 and 2011 (year of index applied depended on year of birth). The IRSAD score ranks geographic areas from most disadvantaged to most advantaged. The score has been categorized into quintiles for this study. Area-level variables used in the IRSAD include annual household income, education, employment status, occupation, home ownership, car ownership, internet connection, English language, child disability and one-parent households.

4.2.4.4 Statistical analysis

The age-specific T1D incidence was calculated as the number of cases per 100,000 individuals, with the total number of children in each age-group as the denominator. The incidence rate per 100,000 person-years by socioeconomic characteristics was calculated using total person-time as the denominator.

The proportion of children with missing data on each socioeconomic variable ranged from 0.02-1.7% (Figure 4-1), with the exception of parental employment (12.0% missing), number of antenatal visits (7.4%) and father's age (4.6%). Multiple imputation was conducted to maintain the validity of the association between socioeconomic characteristics and T1D, and to account for potential bias if the association differed for children with or without complete data.[290] Multiple imputation by chained equation was performed to impute missing values, in Stata SE version 15.0. Missing data were imputed for all variables included in Table 4-1.

The outcome variable (T1D) was not imputed. Mother's birth in Australia and type of antenatal care were included as auxiliary variables. Twenty datasets were generated, with 50 iterations per imputed dataset.

One cannot apply Rubin's rules directly to our imputed data as the denominators (person-time) change in every imputation, Due to the changing denominator (person-time) in each imputed dataset, we did not simply take the mean of incidence rates. Instead we computed the average incidence rate as follows;

$$Average\ Incidence\ Rate\ (\bar{\lambda}) = \frac{Average\ T1D\ cases\ (\bar{n}\ )}{Average\ person\ time} \times 100{,}000$$

And the 95% lower and upper bounds are computed as;[312]

$$95\%\ Lower\ CI = \bar{\lambda} * \left(1 - \frac{1.96}{2 * \sqrt{(\bar{n})}}\right)^2,$$

$$95\%\ Upper\ CI = \bar{\lambda} * \left(\frac{(\bar{n}+1)}{\bar{n}}\right) * \left(1 + \frac{1.96}{\left(2 * \sqrt{(\bar{n}+1)}\right)}\right)^2.$$

Results from complete case and response sample are provided as supplementary material (Supplementary table 4-2 and 4-3). Inferences from complete case, response sample and imputed data (main analysis presented below) are consistent. The estimates in our study are presented and interpreted following American Statistical Association's Statement on p-values.[313]

4.2.5 Results

Among children born from 2002-2013, whose hospitalizations were followed from 2002-2014, 333 (173 boys) were diagnosed with T1D. The overall T1D incidence rate among ≤11-year-olds was 23.0 per 100,000 person-years (95% CI: 20.7-25.7). There was no sex difference in the overall T1D incidence rate. However, age and sex specific incidence (Figure 4-2) shows that peak age of diagnosis occurred earlier in boys (age 5) than girls (age 6).

Table 4-1 depicts the socioeconomic characteristics of children with T1D. The T1D incidence rate was higher among children whose parents were both employed (26.0 per 100,000 [95% CI: 22.8-29.5]) compared to both unemployed (20.0 per 100,000 [95% CI: 12.3-30.6]).

Children whose mother was a private patient had a higher T1D incidence (26.0 per 100,000 [95% CI: 21.6-31.1]) than those whose mother was a public patient (21.6 per 100,000 [95% CI: 18.8-24.8]). The T1D incidence rate was lower for children with a mother who identified as Aboriginal or Torres Strait Islander (7.5 per 100,000 [95% CI: 1.51-21.8]) compared to Caucasians (24.5 per 100,000 [95% CI: 21.9-27.4]).

T1D incidence was lower in the most advantaged IRSAD quintile area (19.4 per 100,000 [95% CI: 13.8-26.5]) compared to the most disadvantaged area (23.5 per 100,000 [95% CI: 18.9-28.9]), but there was no clear gradient in the association between IRSAD and T1D incidence. T1D incidence was highest in the fourth IRSAD quintile (26.5 per 100, 000 [95% CI: 20.9-33.1]). There was no clear parental age patterning of T1D. Children whose mother smoked during pregnancy had lower T1D incidence rate compared to non-smokers.

4.2.6 Discussion

In this population-based study, we found inconsistent evidence of the association between individual and area-level socioeconomic and demographic characteristics and T1D incidence. Individual-level measures of socioeconomic disadvantaged were consistently associated with lower T1D incidence.

In this study, when socioeconomic position was measured by individual characteristics, T1D incidence was higher among advantaged children, whose parents were both employed and whose mothers were private patients. These results are consistent with the US population-based study that reported low T1D risk in disadvantaged children.[32] Higher T1D incidence in advantaged children could possibly be due to differences in lifestyle and health behaviour such as living in less crowded homes, caesarean-section, and breast-feeding/weaning practices not measured in this study. Caesarean-sections have been linked to T1D potentially via the microbiome and perinatal stress, although evidence is mixed.[314] Household crowding has been associated with lower T1D incidence[95], perhaps due to more microbial contact resulting in immune stimulations. Other studies link increasing T1D incidence to reduced herd-immunity to enterovirus infections, because early enterovirus and other viral infections have been associated with high T1D risk.[106] This is supported by findings of a reduction in maternal enterovirus antibody levels over 20 years in Finland and Sweden while T1D incidence has increased in these countries during this period, possibly reflecting a lower immunity for fetal/infant enterovirus infections.[109] Additionally, families with both parents employed may eat more processed and ready-to-eat food due to time scarcity. Children of unemployed, or both employed parents have reportedly higher risk of overweight.[315] This

can be linked with the accelerator or beta-cell-stress hypothesis, where environmental factors leading to fast growth/overweight can exhaust pancreatic beta-cells, which eventually fail due to a secondary autoimmune reaction.[106]

The area-level measure of socioeconomic conditions (IRSAD) used in this study showed lower T1D incidence in the most advantaged areas compared to the most disadvantaged areas of SA, however, there was no dose-response pattern across the five IRSAD quintiles (most disadvantaged to most advantaged). Similar to our findings, low T1D incidence was observed in the most advantaged areas in Germany.[304] Contrary to our area-level findings many previous studies [73, 95, 105] reported high T1D incidence in affluent areas. However, they have used different measures of area-level socioeconomic position, ranging from a few separate variables to combinations of different variables to create an index, such as the Townsend deprivation index in the UK[151] and the Index of Relative Socioeconomic Disadvantage (IRSD) in WA.[73] Both the Townsend deprivation index (unweighted sum of employment, car ownership, home ownership, overcrowding) and IRSD (weighted sum of socioeconomic variables representing disadvantage) are measures of disadvantage, and are slightly different measures than the IRSAD (which takes in to account area-level indicators of advantage and disadvantage). There have also been two studies from England using the Townsend deprivation index that have reported inconsistent findings even with the same index.[99, 151] Similarly, two Western Australian studies (2006[73], 2007[152]) reported inconsistent results between the associations of IRSD with T1D, depending on whether IRSD was measured at

birth or diagnosis. Together, these results suggest that small area-level socioeconomic measures may not be a consistent indicator of T1D risk.

The area-level variation in T1D incidence observed in our study could be due to differences in individual characteristics. Individual characteristics not measured in this study that might have an effect on T1D incidence include genetic make-up, the pre-and-postnatal nutritional environment and other factors affecting parental health behaviour, such as income. It is possible these play a more crucial role than the small area environment. In the context of the hygiene hypothesis, country-level affluence and determinants of overall health such as hygiene, sewerage systems, and clean water supply may help explain the wide range of global T1D incidence, and increasing T1D incidence corresponding to country level socioeconomic improvements.[54, 105] Area-level socioeconomic differences within a high income country like Australia do not necessarily reflect a large difference in hygiene, sewerage or clean water accessibility. Therefore, the hygiene hypothesis might not be relevant in countries like Australia where baseline levels of hygiene are high. This is supported by a Swedish study that used multilevel analysis and demonstrated that administrative areas (counties, municipalities) have minor relevance for individual risk of T1D in Sweden[316], because differences disappeared after adjusting for individual-level factors. Moreover, national reports show individual socioeconomic variation within each IRSAD deciles, suggesting area-level measure do not represent individual-level differences.[154]

We found that children born to Aboriginal or Torres Strait Islander mothers have lower T1D incidence compared to Caucasians and others, consistent with national reports.[62] A higher proportion of Aboriginal or Torres Strait Islander people experience socioeconomic disadvantage, and are more likely to live in overcrowded households compared to non-Aboriginal Australians.[317] Both household crowding and socioeconomic disadvantage have been associated with lower T1D incidence.[32, 95, 96] Ethnic disparities in T1D incidence have also been reported in the US (higher incidence in white-Americans than other ethnicities).[32]

T1D incidence was lower among children whose mothers smoked during pregnancy. Least advantaged mothers are more likely to smoke during pregnancy, and disadvantage has been linked with lower T1D incidence.[96]

This large whole-of-population study brings together multiple data sources to enable investigation of area and individual-level characteristics among children with T1D. As the study only includes data from public hospitals, it is possible that children who are diagnosed and treated for T1D in a private setting may be misclassified, however there are a number of reasons why case ascertainment is high. The incidence in this study (23 per 100,000 person-years among children aged ≤11-year-olds) is consistent with a report of national diabetes register data (24 per 100,000 population for 0-14-year-olds).[62] In a similar study in WA, 99.6% of T1D cases in a national diabetes register were ascertained by capture-recapture method in hospital data.[73] In SA, children with incident T1D diagnosis will normally be admitted for stabilization and education on diabetes management. In SA, there is one public

hospital with a paediatric endocrinology and diabetes service for children, which is included in these data. Furthermore, children are included in the case definition even when they are admitted for reasons other than T1D (e.g. injury), as T1D is included in their additional diagnosis codes.

4.2.7 Conclusion

In this large population study, area and individual-level measures of socioeconomic condition were not consistently associated with T1D in SA. There was no clear area-level socioeconomic patterning of T1D. However, individual-level socioeconomic indicators showed higher T1D incidence in more advantaged children.

4.2.8 Tables and Figures

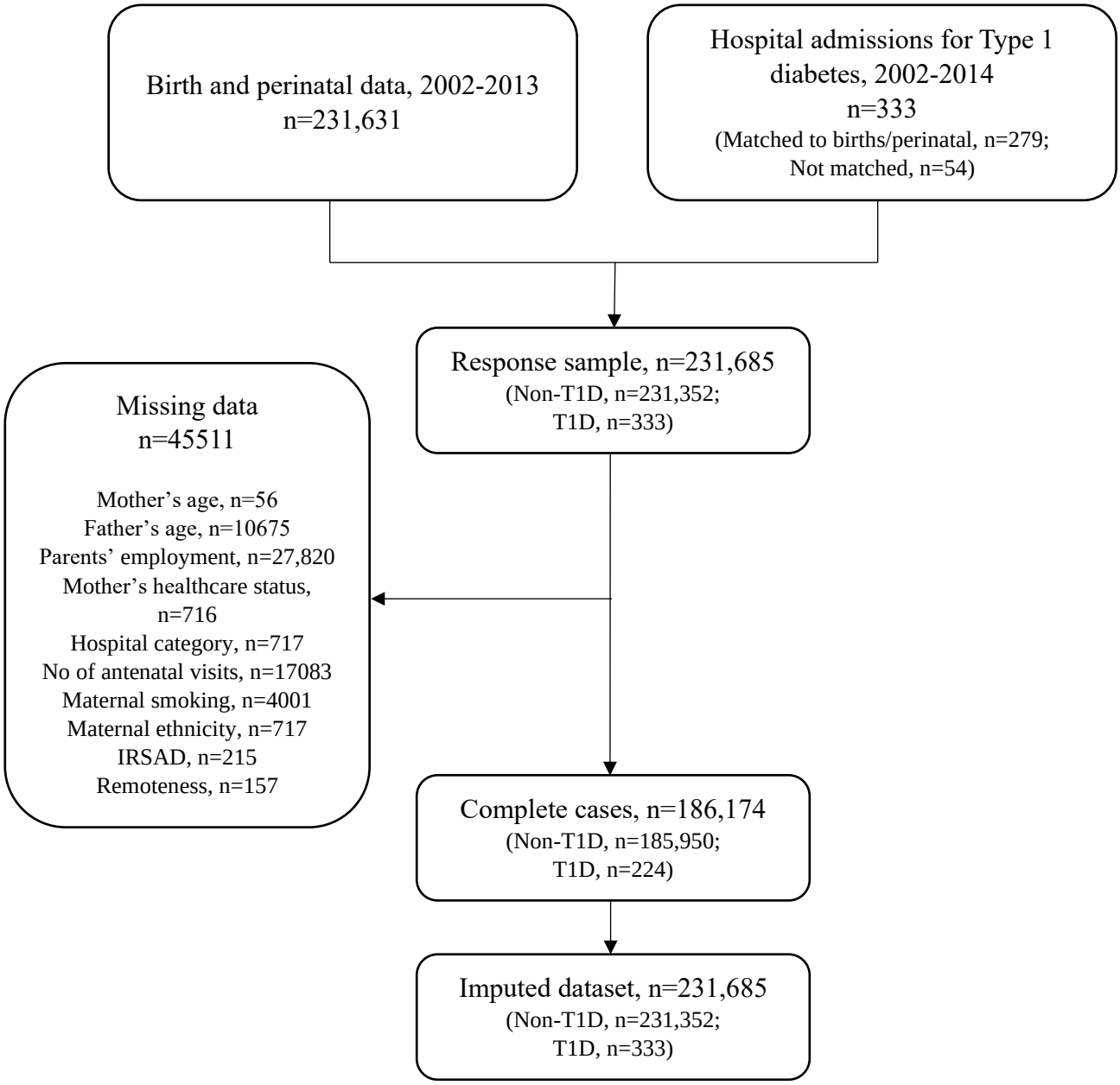


Figure 4-1: Flow chart showing number of children included in data analysis. T1D, Type 1 diabetes; IRSAD, Index of Relative Socioeconomic Advantage and Disadvantage

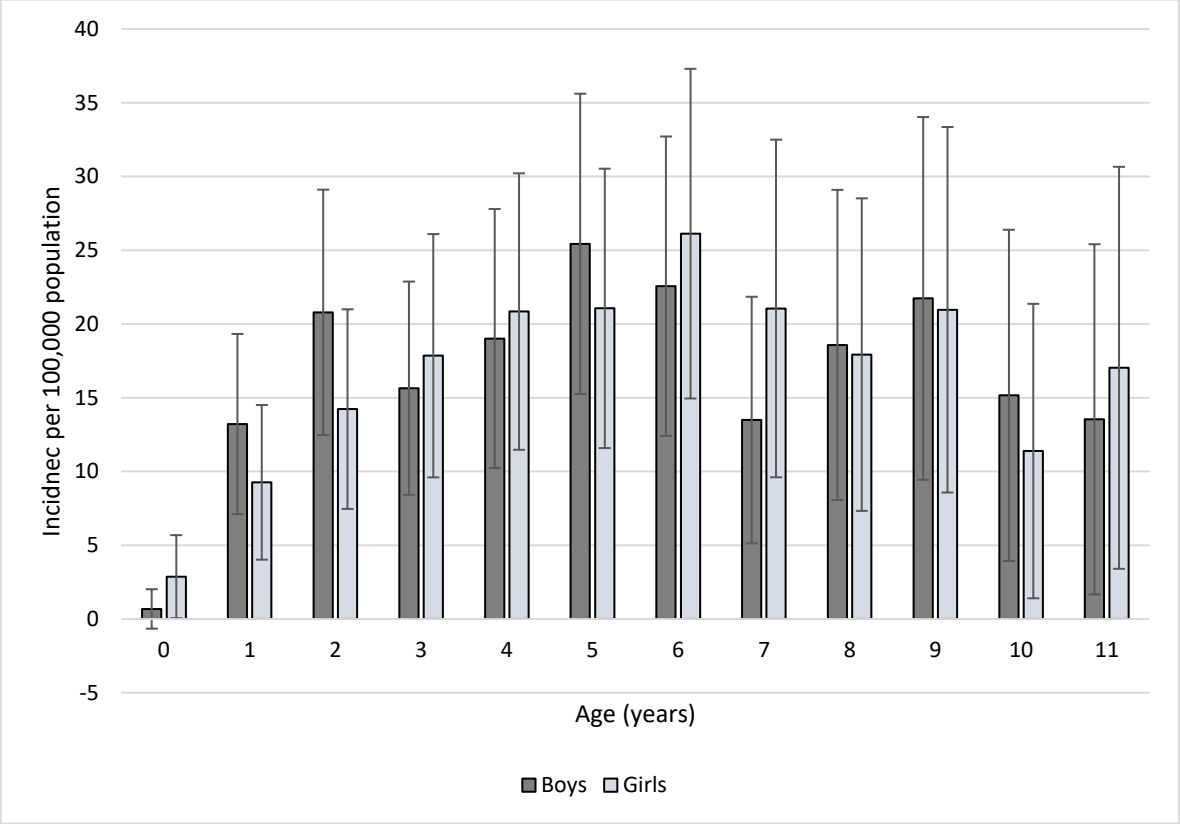


Figure 4-2: Age-specific incidence of T1D per 100,000 population in South Australia, born from 2002-2013, with hospitalizations from 2002-2014

Table 4-1: Characteristics of children with type 1 diabetes in South Australia born 2002-2013 with hospitalizations until 2014 (N=231,685: 231,352 Non T1D; 333 T1D)

	Person time	T1D, n	IR	(95% CI)
Mother's age (years)				
<25	288982.4	66	22.9	(17.7-29.1)
25-29	408222.5	88	21.5	(17.2-26.5)
30-34	465326.9	112	24.1	(19.9-29.1)
≥35	281224.4	67	23.8	(18.4-30.2)
Father's age (years)				
<25	162754.8	39	24.1	(17.1-32.9)
25-29	322821.6	69	21.2	(16.5-26.9)
30-34	468428.1	103	21.9	(17.9-26.6)
≥35	489751.6	122	25.0	(20.8-29.9)
Parents' employment				
Both unemployed	1 04902.7	21	20.0	(12.3-30.6)
One parent employed	425959.3	75	17.6	(13.8-22.0)
Both employed	912894.2	237	26.0	(22.8-29.5)
Mother's healthcare status				
Private	470721.7	122	26.0	(21.6-31.1)
Public	973034.5	211	21.6	(18.8-24.8)
Hospital category				
Private	384961.7	94	24.4	(19.7-29.9)
Public	1058794.0	239	22.6	(19.8-25.6)
Antenatal visits				
<7 visits	118222.0	20	17.3	(10.6-26.6)
7-14 visits	1157203.0	270	23.3	(20.6-26.3)
>14 visits	168331.3	43	25.2	(18.2-34.1)
Maternal smoking				
No smoking	1180970.0	282	23.8	(21.1-26.8)
Smoking	262785.8	51	19.6	(14.6-25.7)
Mother's Ethnicity				
Caucasian	1272363.0	312	24.5	(21.9-27.4)
Aboriginal or Torres Strait Islander	**	**	7.5	(1.5-21.8)
Asian and other	**	**	13.8	(8.1-21.9)
IRSAD quintile				
Most disadvantaged (1)	381984.0	90	23.5	(18.9-28.9)
2nd quintile	320913.1	62	19.4	(14.9-24.9)
3rd quintile	249842.7	65	26.1	(20.1-33.2)
4th quintile	288934.4	77	26.5	(20.9-33.1)
Most advantaged (5)	202082.1	39	19.4	(13.8-26.5)
Remoteness				
Major cities	1044753.0	239	22.9	(20.1-26.0)
Inner regional	138163.5	36	25.8	(18.0-35.8)
Outer regional/remote	260839.7	58	22.4	(17.0-29.0)

T1D, Type 1 diabetes; IR, Incidence rate per 100,000 person-years; IRSAD, Index of Relative Socioeconomic Advantage and Disadvantage; **Mother's ethnicity - due to small cell sizes, number and person time has not been shown.

Table 4-2: (Supplementary Table) Response Sample: Characteristics of children with type 1 diabetes

	NON-T1D		T1D cases			
	N	**%**	**N**	**%**	**IR**	**(95% CI)**
Mother's age (years)	**231352**		**277**			
<25	45013	19. 5	56	20.2	**19.4**	(14.9-25.2)
25-29	66011	28.5	72	25.9	**17.6**	(14.0-22.2)
30-34	73945	32.0	94	33.9	**20.2**	(16.5-24.7)
≥35	46383	20.1	55	19.9	**19.6**	(15.0-25.5)
Father's age (years)	**220742**		**268**			
<25	22800	10.3	31	11.6	**21.1**	(14.9-30.1)
25-29	48784	22.1	53	19.8	**17.6**	(13.4-23.0)
30-34	71422	32.4	83	31.0	**18.4**	(14.8-22.8)
≥35	77736	35.2	101	37.7	**21.2**	(17.5-25.8)
Parents' employment	**203622**		**243**			
Both unemployed	14709	7.2	13	5.4	**14.2**	(8.3-24.5)
One parent employed	59065	29.0	51	21.0	**13.7**	(10.4-18.0)
Both employed	129848	63.8	179	73.7	**22.3**	(19.3-25.8)
Mother's healthcare status	**230692**		**277**			
Private	72690	31.5	99	35.7	**21.1**	(17.3-25.7)
Public	158002	68.5	178	64.3	**18.4**	(15.8-21.3)
Hospital category	**230691**		**277**			
Private	59558	25.8	80	28.9	**20.8**	(16.7-25.9)
Public	171133	74.2	197	71.1	**18.7**	(16.2-21.5)
Antenatal visits	**214345**		**257**			
<7 visits	17639	8.2	15	5.8	**14.3**	(8.6-23.7)
7-14 visits	172479	80.5	209	81.3	**19.7**	(17.2-22.5)
>14 visits	24227	11.3	33	12.8	**21.3**	(15.1-29.9)
Maternal smoking	**227410**		**274**			
No smoking	190108	83.6	230	83.9	**19.8**	(17.4-22.6)
Smoking	37302	16.4	44	16.1	**17.0**	(12.7-22.9)
IRSAD quintile	**231193**		**277**			
Most disadvantaged (1)	65389	28.3	78	28.2	**20.5**	(16.4-25.5)
2nd quintile	52714	22.8	51	18.4	**15.9**	(12.1-20.9)
3rd quintile	38783	16.8	54	19.5	**21.6**	(16.6-28.3)
4th quintile	44805	19.4	64	23.1	**22.2**	(17.4-28.3
Most advantaged (5)	29502	12.8	30	10.8	**14.9**	(10.4-21.2)
Remoteness	**231251**		**277**			
Major cities	168673	72.9	199	71.8	**19.1**	(16.6-21.9)
Inner regional	22050	9.5	29	10.5	**21.0**	(14.6-30.2)
Outer regional/remote	40528	17.5	49	17.7	**18.8**	(14.2-24.9)

Table 4-3: (Supplementary Table) Complete cases (N=186174, Non-T1D n=185950, T1D cases n=224) Characteristics of children with type 1 diabetes

	NON-T1D		T1D cases			
	N	**%**	**N**	**%**	**IR**	**(95% CI)**
Mother's age (years)						
<25	32901	17.7	35	15.6	**16.8**	(12.1-23.5)
25-29	54227	29.2	62	27.7	**18.7**	(14.6-24.0)
30-34	61360	33.0	80	35.7	**20.9**	(16.8-26.1)
≥35	37462	20.2	47	21.0	**20.9**	(15.7-27.8)
Father's age (years)						
<25	18458	9.9	24	10.7	**20.5**	(13.8-30.6)
25-29	41472	22.3	44	19.6	**17.3**	(12.9-23.3)
30-34	60944	32.8	70	31.3	**18.3**	(14.5-23.2)
≥35	65076	35.0	86	38.4	**21.9**	(17.7-27.0)
Parents' employment						
Both unemployed	12285	6.6	11	4.9	**14.6**	(8.1-26.4)
One parent employed	53522	28.8	47	21.0	**14.0**	(10.5-18.7)
Both employed	120143	64.6	166	74.1	**22.6**	(19.4-26.3)
Mother's healthcare status						
Private	60874	32.7	85	38.0	**21.8**	(17.6-26.9)
Public	125076	67.3	139	62.1	**18.4**	(15.6-21.7)
Hospital category						
Private	52851	28.4	73	32.6	**21.4**	(17.0-26.9)
Public	133099	71.6	151	67.4	**18.8**	(16.0-22.0)
Antenatal visits						
<7 visits	11828	6.4	7	3.1	**10.3**	(4.9-21.5)
7-14 visits	152681	82.1	186	83.0	**19.8**	(17.1-22.9)
>14 visits	21441	11.5	31	13.8	**22.5**	(15.8-32.0)
Maternal smoking						
No smoking	159543	85.8	188	83.9	**19.5**	(16.9-22.5)
Smoking	26407	14.2	36	16.1	**19.7**	(14.2-27.3)
IRSAD						
Most disadvantaged (1)	50663	27.3	51	22.8	**17.6**	(13.3-23.1)
2nd quintile	42389	22.8	43	19.2	**16.9**	(12.5-22.8)
3rd quintile	31607	17.0	50	22.3	**24.8**	(18.8-32.8)
4th quintile	37211	20.0	56	25.0	**23.7**	(18.2-30.7)
Most advantaged (5)	24080	13.0	24	10.7	**14.7**	(9.9-22.0)
Remoteness						
Major Cities	137012	73.7	158	70.5	**18.9**	(16.1-22.1)
Inner Regional	17765	9.6	26	11.6	**23.4**	(15.9-34.4)
Outer Regional/ Remote	31173	16.8	40	17.9	**20.3**	(14.9-27.6)

CHAPTER 5 CAESAREAN SECTION AND RISK OF TYPE 1 DIABETES: WHOLE-OF-POPULATION STUDY

5.1 Preface

This Chapter contains the second paper contributing to this thesis. This paper has been published in Diabetic Medicine.

This paper addresses the second questions of this doctoral thesis.

What is the effect of caesarean birth on childhood type 1 diabetes? Does the risk of type 1 diabetes differ by prelabour and intrapartum caesarean?

As mentioned in Chapter 2 that socioeconomic condition is major determinant of health, many risk factors of type 1 diabetes are socioeconomically patterned. For example, caesarean birth is more common among socioeconomically advantaged women. After studying the socioeconomic characteristics of children with type 1 diabetes in the first descriptive study, in study 2 (Chapter 5), I studied the association between caesarean birth and type 1 diabetes, using the SA ECDP data. First reason to explore this question was the rising incidence of type 1 diabetes and the parallel increasing rates of caesarean birth at national level in Australia. Secondly, keeping the microbiome theory in mind, I wanted to explore the effect of caesarean birth occurring before and after the onset of labour because that may effect neonatal exposure to maternal vaginal microbiota. Very few studies reported the effect of

caesarean types (emergency and elective caesarean) on childhood type 1 diabetes. However, elective and emergency caesarean do not differentiate whether caesarean occurred before or after the onset of labour and after the rupture of membranes. Therefore, In Chapter 5, I explored the question whether type 1 diabetes risk differed for prelabour or intrapartum caesarean births.

Highlights

The effect of caesarean birth on childhood type 1 diabetes is negligible. Similarly, for the effect of intrapartum caesarean on type 1 diabetes. Together this suggests that the differences in the neonatal microbiota as a result of caesarean birth is unlikely to impact on type 1 diabetes risk pathogenesis.

5.2 PUBLICATION 2: Caesarean section and risk of type 1 diabetes: whole-of-population study

5.2.1 Statement of Authorship

Title of Paper	**Caesarean section and risk of type 1 diabetes: whole-of-population study**
Publication Status	☑ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	

Principal Author

Name of Principal Author (Candidate)	Mumtaz Begum
Contribution to the Paper	MB prepared the draft, helped in study design, conducted the analyses and interpreted the results.
Overall percentage (%)	
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am

	the primary author of this paper.		
Signature		Date	17/01/2020

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Rhiannon Pilkington		
Contribution to the Paper	RP conceptualised the project, assisted in data acquisition, analysis and interpreting the findings, reviewed and revised the draft.		
Signature		Date	17-1-2020

Name of Co-Author	Catherine Chittleborough		
Contribution to the Paper	CC reviewed and revised the draft, and provided critical input in the study design, data analysis and interpretation of the findings and helped in data acquisition.		
Signature		Date	17/1/2020

Name of Co-Author	John Lynch

Contribution to the Paper	JL acquired the data, provided intellectual input in study design, critically reviewed and revised the manuscript.		
Signature		Date	17/1/20

Name of Co-Author	Megan Penno		
Contribution to the Paper	MP provided intellectual input into analysis and interpreting the findings, and reviewed and revised the draft.		
Signature		Date	14/01/2020

Name of Co-Author	Lisa Smithers		
Contribution to the Paper	LS conceptualised the project, provided guidance in the study design and data analysis, assisted in data acquisition, interpreting the results, and reviewed and revised the manuscript.		
Signature		Date	17 JAN 2020

What is New?

- Evidence about caesarean section and childhood type 1 diabetes risk is mixed. Only one study reported whether type 1 diabetes risk differs by prelabour or intrapartum caesarean.

- A potential link between type 1 diabetes and caesarean section is lack of exposure to the vaginal microbiota. Prelabour caesarean births are not exposed to the vaginal microbiota, whereas intrapartum caesareans presumably have some exposure.

- We found a negligible risk of type 1 diabetes for children who had intrapartum or prelabour caesarean, and the 95% CI were wide and included the null suggesting that neonatal vaginal microbiota might not be involved in type 1 diabetes.

5.2.2 Abstract

Background

A hypothesized mechanism for increased type 1 diabetes risk among caesarean births is lack of exposure to the vaginal microbiota. Children born by prelabour caesarean are not exposed to the vaginal microbiota, while caesarean births during labour (intrapartum) may be exposed. The aim of this study was to estimate type 1 diabetes risk among children born by caesarean compared with normal vaginal delivery.

Methods

This whole-of-population study linked routinely-collected, de-identified administrative data from the South Australian Early Childhood Data Project for all births from 1999-2013. Type 1 diabetes cases were identified using inpatient hospitalisations from 2001-2014 (ICD-10-AM codes E10-E109). Type 1 diabetes risk for caesarean was assessed by Cox regression using two models: 1) caesarean versus vaginal, 2) prelabour or intrapartum caesarean versus vaginal. Analyses were adjusted for confounding and multiple imputation was used to address missing data.

Results

A total of 286,058 children born 1999-2013 contributed to 2,200,252 person-years, of which 557 had type 1 diabetes. Of all births, 90,546 (31.7%) were caesarean, and of these 53.1%

were prelabour and 46.9% intrapartum caesarean. Compared with vaginal delivery, the adjusted hazard ratio for type 1 diabetes was 1.05 (95% confidence interval (CI) 0.86-1.28) for caesarean, 1.02 (95% CI 0.79-1.32) for prelabour caesarean, and 1.08 (95% CI 0.82-1.41) for intrapartum caesarean.

Conclusion

There may be a small increased type 1 diabetes risk following caesarean, but confidence intervals included the null. The lower estimate for prelabour compared with intrapartum caesarean, and the potential for unmeasured confounding suggest that neonatal vaginal microbiota might not be involved in type 1 diabetes.

Keywords: Caesarean section, intrapartum and prelabour caesarean, type 1 diabetes, record linkage, whole-of-population

5.2.3 Introduction

Type 1 diabetes is an autoimmune disorder, characterized by destruction of insulin producing pancreatic beta-cells. The increasing incidence of childhood type 1 diabetes has been linked with environmental risk factors interacting with genetic predisposition. The rise in caesarean births in parallel with increasing type 1 diabetes incidence is one reason caesarean births have been suggested as a risk factor for type 1 diabetes. For example, the average global annual increase in caesarean births was 4.4% from 1990 to 2014[18], and from 1990-1999 the Diamond project[6] reported a 2.8% annual global increase in type 1 diabetes incidence. However, not all countries with higher caesarean rates have high type 1 diabetes incidence. Brazil[161] and China[162] have caesarean rates of 77% and 46% respectively, but relatively low type 1 diabetes incidence (12.8 and 1.01 per 100,000 person-years, respectively).[8, 150] Australia has high childhood type 1 diabetes incidence (24.4 per 100,000 population, aged <15 years)[175], as well as caesarean rates (33% in 2015)[318] more than double the World Health Organization recommendation of 10-15%.[319]

It has been hypothesized the neonatal gut microbiota is a link between caesarean births and type 1 diabetes, as early microbial contact may influence the development of the immune system.[133] During a normal vaginal delivery, the neonate is exposed to the vaginal and gastrointestinal microbiota of the mother.[320] Children born by caesarean, particularly prelabor caesarean, are not exposed to vaginal microbiota and their gut microbial composition is more reflective of maternal skin.[320] It has also been demonstrated that gut

bacterial colonisation patterns differ among neonates born by emergency caesarean, elective caesarean and natural birth.[156]

There are inconsistent findings about the association between caesarean birth and type 1 diabetes risk. Multiple studies of varying designs (case-control, meta-analysis) have reported 20-30% increased type 1 diabetes risk for caesarean birth.[25] However, larger population-based cohort studies with better control of confounding, found null or small associations between caesarean birth and type 1 diabetes.[20-22] A Norwegian cohort study [22] and a Swedish case-control study [23] reported a 6% and 2% increased type 1 diabetes risk for caesarean birth, respectively. A Swedish sibling-design study reported a small relative risk for type 1 diabetes of 1.06 for elective and emergency caesarean compared with vaginal birth.[20] A Danish cohort study distinguished whether caesarean occurred before or during labour and demonstrated no type 1 diabetes risk for intrapartum (during labour) and a small risk for prelabor caesarean. [21] Therefore, for the present study, we separated prelabor and intrapartum caesarean births under the assumption children born by prelabor caesarean are not exposed to the maternal vaginal microbiota, whereas births by intrapartum caesarean have some exposure if the membranes have ruptured. The objective of this study was to estimate the association between caesarean birth and type 1 diabetes risk, and to see whether the risk differed by caesarean type stratified as prelabor or intrapartum caesarean, using whole-of-population data.

5.2.4 Materials and methods

5.2.4.1 Study design and population

Routinely collected, de-identified government administrative linked data from the South Australian Early Childhood Data Project (1999-2013)[256] was used in this whole-of-population study (Figure 5-1). The characteristics of the South Australian population are reflective of the Australian population.[318]

Datasets were linked by SA NT Datalink, an independent agency, using a probabilistic linkage algorithm to match children across datasets using demographic information such as name, sex, date of birth, and address. Data custodians provide de-identified data to the researchers following data linkage. Australian data linkage systems typically estimate around 0.1% false linkage rates.[271]

5.2.4.2 Type 1 diabetes

Type 1 diabetes information was sourced from hospitalisation data (2001-2014) from all South Australian public hospitals. Children with type 1 diabetes were identified from their first type 1 diabetes related hospitalisation, using International Classification of Disease, 10th edition, Australian Modification (ICD-10-AM) codes (E10, ranging E101 to E109) for primary and secondary diagnoses.[264] Trained staff use these ICD-10-AM codes as reporting tools and assign them to patient records following discharge.

5.2.4.3 Method of delivery

Information on delivery method (normal spontaneous and instrumental vaginal deliveries, elective and emergency caesarean) and onset of labour (spontaneous, no labour and induction) was obtained from South Australian Perinatal Statistics Collection[259] for all children in South Australia, born from 1999-2013. This is a mandatory collection of perinatal information of all births at hospital or home. This information collected by neonatal nurses/midwives, when validated against an audit of medical records, was highly accurate in capturing perinatal information.[261] Perinatal data was validated and supplemented with Births Registration data. Prelabour caesarean included all caesareans in the absence of spontaneous or induced labour, and intrapartum caesarean included all elective and emergency caesareans performed after spontaneous or induced labour.

5.2.4.4 Confounding

Based on previous studies and literature, potential confounding was identified *a priori* using a directed acyclic graph (Figure 5-2, Supplementary). Confounding factors were sourced from the South Australian Perinatal Statistics Collection[259], and supplemented by Births Registration data. Parental characteristics included parents' age, maternal pre-existing (type 1 and type 2 diabetes) or gestational diabetes (yes/no), hypertension (yes/no), smoking at first antenatal visit (smoker, quit before first visit, non-smoker), maternal birth region (Oceania, Europe, and Africa, Asia, America), mother's ethnicity (White European ancestry, Aboriginal and Torres Strait Islander, and Asian and others), public or private healthcare,

public or private hospital of birth, and number of antenatal visits (continuous variable). Parents' highest occupation (four categories; I) Managers, administrators and professionals, II) Para-professionals, tradespersons, clerks, salespersons, and personal service, III) Plant, machine operators, drivers and labourers, IV) Students, pensioner, home duties and unemployed) was also included. Other child-related factors included are birth order (1st, 2nd, 3rd, ≥4th child), birthweight for gestational age z-score (calculated using Australian birthweight standards[321]), area-level socioeconomic disadvantage (Index of Relative Socioeconomic Advantage and Disadvantage, IRSAD [305]), and remoteness (measured using the Australian Remoteness Index for Areas, ARIA).[322] Maternal body mass index (BMI) at first antenatal visit (BMI<25, 25-<30, ≥30 kg/m^2) was only included in the sensitivity analyses, due to high proportion of missing data as it was not routinely collected until 2007.

5.2.4.5 Statistical analysis

Crude and adjusted hazard ratios (HR) for the risk of type 1 diabetes for children born by caesarean section was assessed by Cox regression using two models: 1) caesarean versus vaginal delivery, 2) prelabour or intrapartum caesarean versus vaginal delivery. Cox proportional hazard regression was used to account for differences in observation time. For instance, children born in 1999 were observed for 15 years and 2013 births were observed for one year, and the mean observation time was eight years. Children were followed from birth until type 1 diabetes diagnosis, or till the end of follow-up (June-2014). The

proportional hazard assumption was checked by the Schoenfeld residual test, and it demonstrated non-violation of the proportional hazard assumption.

Five sensitivity analyses were performed to check if the association between caesarean and type 1 diabetes was similar to the main findings. The first sensitivity analysis was adjusted for maternal BMI in addition to other confounders. Maternal BMI was collected from 2007 onwards and was not included in the main analyses. The second sensitivity analysis was restricted to singleton births, to make our study comparable with previous studies, which were conducted on singleton births.[21] The third sensitivity analysis was restricted to children born from July 2001 to December, 2013 (similar starting point for hospital and births data), as hospital data was only available from July 2001 and any diagnosis of type 1 diabetes from 1999 to mid-2001 may have been misclassified. The fourth sensitivity analysis was restricted to children born in South Australia, for whom we had more complete data. Lastly all the above reasons were combined, and analysis was restricted to singleton children who were born in South Australia from July 2001 to December 2013, with maternal BMI included in the model as a confounder.

The exposure variable (delivery method) and confounding factors mentioned in Table 5-1 had missing data ranging from 0.03% to 1.9%, except father's age (4.6%), number of antenatal visits (8.7%) and maternal BMI (66.6%). Maternal BMI had a high proportion of missing data as collection of this information started in 2007 and hence is only included in sensitivity analyses.

Multiple imputation was performed to account for potential bias if the association was different for children with or without complete data.[323] Multiple imputation by chained equation was performed, and 20 datasets were created with 50 iterations. All variables included in the adjusted and sensitivity analyses were included in the imputation model. Estimates were combined from the 20 imputed datasets following Rubin's rules.

The results from imputed analyses are presented. Imputed results were consistent with the complete case results that are provided in supplementary material (Table 5-4 and Table 5-5). All analyses were performed using Stata SE version 15.0 (Stata Corp, College Station, Texas, USA).

The estimates and confidence intervals in our study are interpreted based on the American Statistical Association's Statement on p-values.[324]

5.2.4.6 Ethics Approval

Ethics approval was granted by the Human Research Ethics Committees of the South Australian Department of Health (HREC/13/SAH/106), and the Aboriginal Health Council of South Australia (04-13-538).

5.2.5 Results

A total of 286,058 children born 1999-2013, contributed to 2,200,252 person-years, during which 557 children were diagnosed with type 1 diabetes. The incidence rate of type 1 diabetes for children born from 1999-2013 (aged ≤15 years) was 25.3 per 100,000 person-years. Among 286,058 children born from 1999-2013, 31.7% had a caesarean birth and 68.3% had a vaginal birth. Among 90,546 children who had a caesarean birth, 53.1% were prelabour and 46.9% were intrapartum caesarean (Figure 5-1). Type 1 diabetes was diagnosed following hospitalisation from 2001 to 2014 in 557 cases, of which 381 (68.5%) were a normal vaginal delivery, 89 (16.0%) were prelabour caesarean and 87 (15.6%) were intrapartum caesarean.

Table 5-1 shows the sociodemographic and perinatal characteristics of children born by caesarean and vaginal delivery. There were differences in the characteristics of children born by caesarean compared with vaginal delivery. For example, children who had caesarean births were more likely to be from advantaged areas, born in private hospitals, their mother was more likely to have had private healthcare, pre-existing diabetes, hypertension, or high BMI (≥30 kg/m^2).

Table 5-2 illustrates the association of caesarean with type 1 diabetes, compared with vaginal delivery. Children with caesarean birth had a higher estimate for type 1 diabetes than vaginal

delivery (crude HR = 1.09, 95% CI 0.90-1.33), which was attenuated after adjustment for confounding (HR = 1.05, 95% CI 0.86-1.28).

Table 5-3 shows the risk of type 1 diabetes for caesarean types. Adjusted estimates show little evidence of increased type 1 diabetes risk for prelabour (HR = 1.02, 95% CI 0.79-1.32) and slightly higher risk for intrapartum caesarean (HR = 1.08, 95% CI 0.82-1.41) compared with vaginal delivery, but confidence intervals were wide.

The association between caesarean and type 1 diabetes was similar to the main analysis in the five sensitivity analyses (Table 5-6, Supplementary) when restricted to singleton infants, children born in South Australia with available perinatal and birth data, and 2001-2013 births with similar data commencement periods for births and hospitalisation. Inclusion of maternal BMI as a confounder did not change the findings. Finally, restricting the analysis to singleton children who were born in South Australia from 2001-2013, and including maternal BMI in the model, showed a similar pattern as the main findings and had wide confidence intervals.

5.2.6 Discussion

This study compared the risk of type 1 diabetes for children born by caesarean with children who had normal vaginal birth, in a population-based cohort born from 1999-2013 (aged ≤15 years). Adjusted estimates showed a 5% higher type 1 diabetes risk for caesarean births

compared to vaginal delivery, but 95% CIs were wide and included the null. Contrary to what we had hypothesized (higher type 1 diabetes risk for prelabour than intrapartum caesarean), the estimates showed a slightly higher type 1 diabetes risk for intrapartum caesarean (8%) than prelabour caesarean (2%). This reversal of expected risk, together with wide confidence intervals and the potential for unmeasured confounding suggest the small increased type 1 diabetes risk for caesarean birth may be due to unmeasured confounding.

Contrary to our findings of a small increased type 1 diabetes risk, a meta-analysis (including children aged 0-14 years in 18 out of 20 studies) reported around 20% increased type 1 diabetes risk for caesarean births.[25] The meta-analysis of 20 studies included 17 case-control studies with limited information on potential confounding, and notably only eight studies adjusted for maternal type 1 diabetes.[25] Our findings of 5% increased type 1 diabetes risk for caesarean birth are consistent with large population-based studies in Sweden (odds ratio (OR) = 1.02, 95% CI 0.94-1.10)[23], and Norway (rate ratio = 1.06, 95% CI 0.91-1.23)[22] that adjusted for similar confounding factors, and included similar age ranges (0-19 years, 0-15 years). A Swedish population-based study (type 1 diabetes = 10,428, aged <15 years) reported 15% increased type 1 diabetes risk for elective caesareans (relative risk = 1.15, 95% CI 1.06-1.25), which attenuated to 6% (relative risk = 1.06, 95% CI 0.85-1.31) in the sibling-design analysis, with wide confidence intervals that include the null. This attenuation in the sibling study also suggests the potential for residual confounding. However, the 2% (relative risk = 1.02, 95% CI 0.95-1.11) increased type 1 diabetes risk for emergency caesarean changed to 6% (relative risk = 1.06, 95% CI 0.88-1.28) higher risk in the sibling-design, which is difficult to

interpret.[20] Also consistent with our findings, a Danish nationwide population-based study (type 1 diabetes ~ 4,000, aged <15 years) reported a small increased risk of type 1 diabetes for prelabour (HR = 1.1, 95% CI 0.95-1.2), and no risk for intrapartum (HR = 1.0, 95% CI 0.89-1.1) caesarean births.[21]

The speculated mechanism between caesarean and type 1 diabetes risk comes from studies demonstrating disparity in the gut microbiota of children who had caesarean section versus vaginal birth, which was thought to impact immune development.[156] Gut microbiota composition of children with type 1 diabetes also differs from that of healthy children.[142] However, our study findings do not support the theory that exposure to the maternal vaginal microbiota during birth plays a role in type 1 diabetes risk, as the risk for prelabour would be higher if vaginal microbiota were involved. It is possible that previously reported neonatal disparities in gut microbiota composition related to delivery method could be temporary changes and may not induce long term type 1 diabetes risk. In support of this, a study found delivery method was associated with different gut bacterial colonisation at one week across caesarean and normal vaginal deliveries, but these differences became less prominent at one month and almost disappeared by one year of age.[156] Additionally, apart from the islet-autoantibodies acquired from mothers' placenta, islet-autoantibodies rarely appear before age 6-months,[325] which could suggest there is no involvement of the neonatal microbiota inherited at the time of birth in the initiation of type 1 diabetes. Previously observed disparities in the gut microbiota of children with and without type 1 diabetes may not be due to delivery method. This is because during early childhood the gut microbiota goes through

an intense phase of remodelling, and by age 3-years it transitions to an adult-like composition.[133] Therefore, other potential risk factors such as breastfeeding, dietary practices, medications, or infections might play a more important role in altering the microbiota composition of children,[133] who develop type 1 diabetes.

In our whole-of-population study, linkage of perinatal and hospital data enabled us to account for a wide range of confounding factors. Furthermore, routine collection of perinatal data at birth using a standardized tool provided precise and well documented detail of the confounding factors, minimising recall-bias. However, as with all observational studies unmeasured confounding remains possible. Although we included a wide range of parental and child characteristics to reduce confounding, we did not have data on maternal autoimmune conditions, maternal infections, and father's type 1 diabetes. Father's type 1 diabetes is strongly associated with childhood type 1 diabetes, and may potentially impact caesarean delivery through socioeconomic conditions. When Clausen et al.[21] adjusted for father's type 1 diabetes, it did not change their estimate, possibly because the link between father's type 1 diabetes and caesarean is weak. As we have adjusted for a range of socioeconomic variables, we have attempted to block the potential confounding path from father's type 1 diabetes to caesarean birth (Figure 5-2, Supplementary). In addition, our estimates for intrapartum caesarean may be confounded by indication as maternal infection (for example group B streptococcus, GBS) may cause prelabour rupture of membranes and can be an indication for intrapartum caesarean. The consequent exposure of GBS positive

mothers, or mothers with prolonged rupture of membranes, to GBS prophylactic antibiotics may also affect the neonatal microbiota and the child's risk of type 1 diabetes.[326]

Our study only included children attending public hospitals in South Australia, so we may have missed children that have never attended public hospitals. However, we believe that case ascertainment is high, because there is one paediatric public hospital in South Australia with a specialised endocrinology unit, where children are admitted for stabilization after diagnosis with type 1 diabetes. In addition, by using primary and secondary ICD-10-AM type 1 diabetes diagnoses codes, we have identified children with type 1 diabetes admitted to hospital for other reasons (e.g. injury), as type 1 diabetes is included as their secondary diagnoses. High case ascertainment is substantiated after comparing the incidence reported here (25.3 per 100,000 person-years) with national registry data (24.4 per 100,000 population) collected from 2000 to 2016.[175]

In this whole-of-population study, results indicated there was a small increased risk of type 1 diabetes for all caesarean births, but confidence intervals were wide and included a null effect. The small estimates, and lower type 1 diabetes risk for prelabour caesarean than intrapartum caesarean suggest our findings do not support the theory that birth method induced differences in neonatal microbiota composition have a role in type 1 diabetes risk.

5.2.7 Tables and Figures

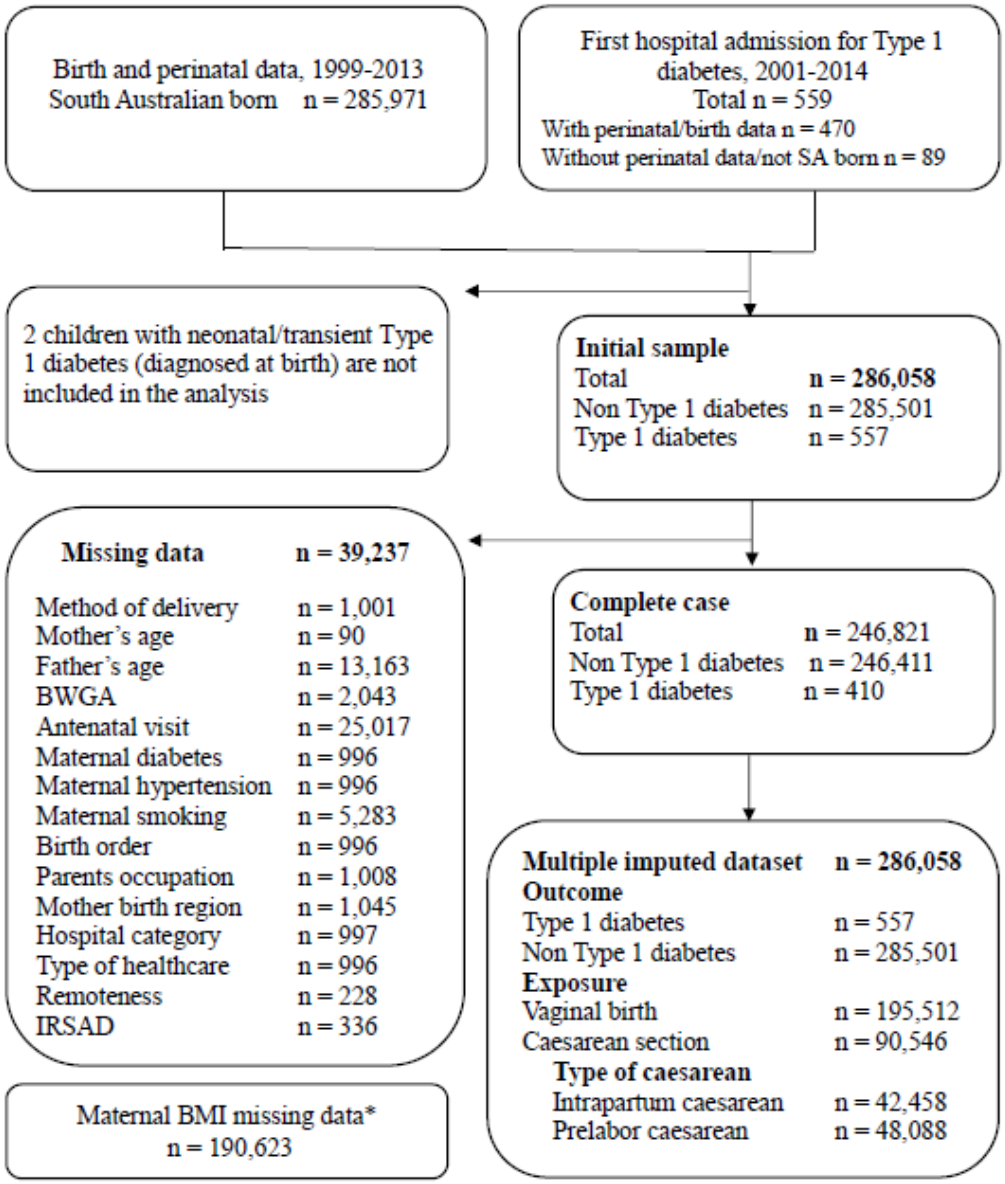


Figure 5-1: Flow chart of the study population. IRSAD (Index of Relative Socioeconomic Advantaged and Disadvantaged), BWGA (birthweight for gestational age), SA (South Australia) *Maternal BMI was measured from 2007 onwards

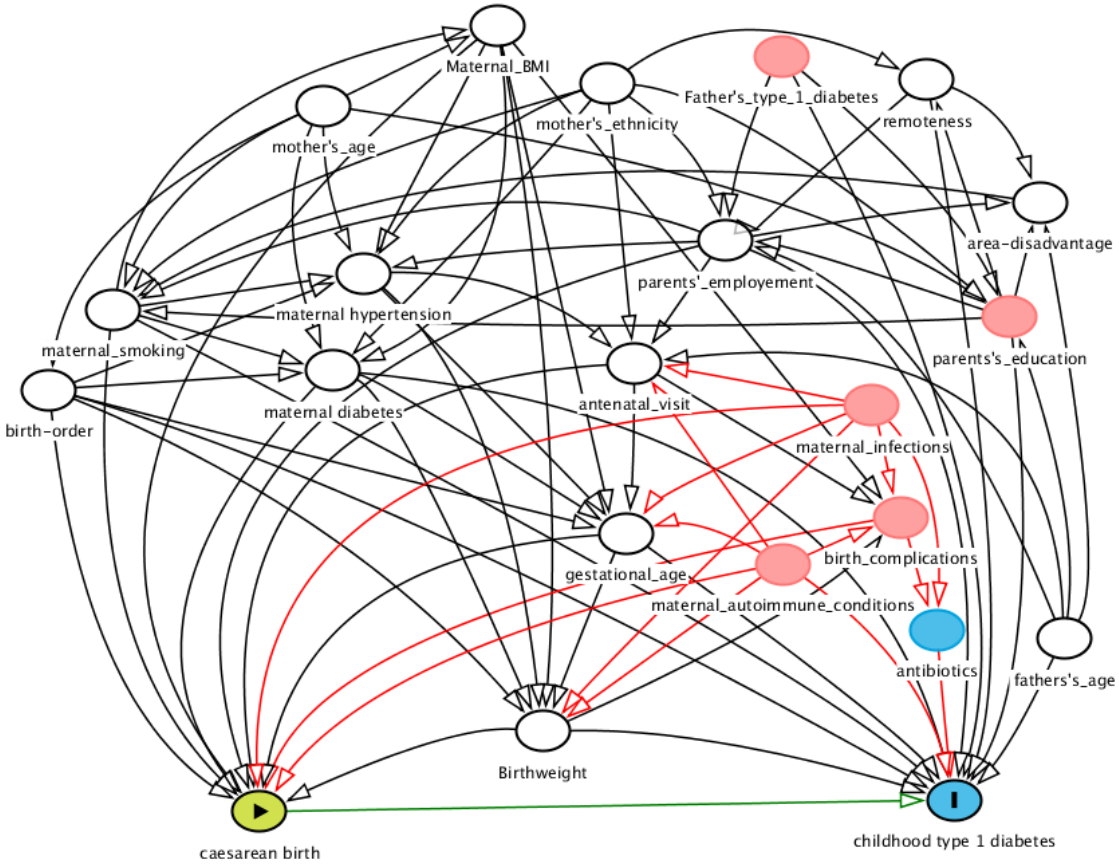


Figure 5-2: (Supplementary Figure) Directed acyclic graph (DAG) showing proposed confounding structure

Exposure = Caesarean birth

Outcome = Childhood type 1 diabetes

Confounders = Adjusted for in our model

Confounders = Not measured, and not adjusted for in our model

Path showing unmeasured confounding = 

Link between maternal infection and childhood type 1 diabetes (not a confounder) 

- As shown in the DAG, adjusting for socioeconomic variables (parent's employment, remoteness, and area-level index of socioeconomic advantage and disadvantage) potentially blocked the path from father's type 1 diabetes to caesarean births.
- Maternal infection, maternal autoimmune conditions, and birth complications remain possible causes of unmeasured confounding.
- Birthweight and gestational age are not mediators and colliders, they are confounders and effect both the exposure (caesarean birth) and the outcome (childhood type 1 diabetes), and therefore, we have adjusted for them in our model.

Table 5-1: Sociodemographic and perinatal characteristics of children/parents by delivery method (N = 286,058)

Characteristics	**Vaginal Delivery**		**Caesarean delivery**	
	N = 195,512	**(%)**	**N = 90,546**	**(%)**
Type 1 diabetes				
No	195,131	(99.8)	90,370	(99.8)
Yes	381	(0.2)	176	(0.2)
Mother's age (years)*	28.96 ± 5.6		30.65 ± 5.5	
Father's age (years)*	32.31 ± 6.5		33.63 ± 6.5	
BWGAZ*	-0.04 ± 0.9		0.07 ± 1.1	
Number of antenatal visits*	10.36 ± 2.9		10.60 ± 2.9	
Maternal diabetes (gestational or pre-existing)				
No	186,786	(96)	83,236	(92)
Yes	8,726	(4.5)	7,310	(8.1)
Maternal hypertension				
No	182,228	(93)	80,085	(88)
Yes	13,284	(6.8)	10,461	(12)
Mother BMI				
<25 kg/m^2 (underweight and normal)	106,740	(55)	38,376	(42)
25 – <30 kg/m^2 (overweight)	50,718	(26)	25,846	(29)
≥30 kg/m^2 (obese/severely obese)	38,054	(20)	26,324	(29)
Maternal smoking				
Smoker	36,438	(19)	13,973	(15)
Quit in pregnancy before first visit	7,753	(4.0)	3,716	(4.1)
Non-smoker	151,321	(77)	72,857	(81)
Birth order				
1st child	79,841	(41)	40,037	(44)
2nd child	67,879	(35)	32,175	(36)
3rd child	30,056	(15)	12,411	(14)
≥4 child	17,736	(9.1)	5,923	(6.6)
Parent's highest occupation				
Managers and professionals	76,112	(39)	39,439	(44)
Para professionals, tradesperson	73,802	(38)	34,066	(38)
Machines operators, drivers, labourers	23,362	(12)	9,386	(10)
Students, pensioners, unemployed	22,236	(11)	7,655	(8.4)

Mother's birth region				
Oceania and Antarctica	163,343	(84)	75,478	(83)
Europe	13,723	(7.1)	6,183	(6.8)
Africa, Asia, Americas	18,446	(9.4)	8,885	(9.8)
Hospital category				
Private	42,819	(22)	29,670	(33)
Public	152,693	(78)	60,876	(67)
Healthcare of mother				
Private	52,838	(27)	35,530	(39)
Public	142,674	(73)	55,016	(61)
Remoteness				
Major cities of South Australia	141,024	(72)	66,543	(74)
Inner regional South Australia	19,323	(9.9)	7,985	(8.8)
Remote and very remote South Australia	35,165	(18)	16,018	(18)
IRSAD quintile				
Most disadvantaged (1)	56,179	(29)	23,328	(26)
2nd quintile	43,731	(22)	19,966	(22)
3rd quintile	32,664	(17)	14,888	(16)
4th quintile	37,132	(19)	18,773	(21)
Most advantaged (5)	25,806	(13)	13,591	(15)
Mother's ethnicity				
White European ancestry	169,538	(87)	79,033	(87)
Aboriginal or Torres Strait Islander	5,817	(3.0)	2,548	(2.8)
Asian, other	20,157	(10)	8,965	(10)

*Data presented are mean ± SD, all others are n (%)
IRSAD (Index of Relative Socioeconomic Advantage and Disadvantage)
BWGAZ (Birthweight for gestational age z-score)
Percentages may not add to 100% due to rounding

Table 5-2: Method of delivery and risk of type 1 diabetes (n = 286,058, type 1 diabetes = 557)

	Unadjusted		**Adjusted**	
	HR	**(95% CI)**	**HR**	**(95% CI)**
Vaginal delivery	**Ref**		**Ref**	
All caesarean births	**1.09**	(0.90-1.33)	**1.05**	(0.86-1.28)

Adjusted for birthweight for gestational age z score, parental age, parental occupation, maternal diabetes and hypertension, maternal region of birth, maternal ethnicity, IRSAD, remoteness, birth at public or private hospital, child's birth order, private or public healthcare, antenatal visit, maternal smoking

Table 5-3: Caesarean type and risk of type 1 diabetes (n = 286,058, type 1 diabetes = 557)

	Unadjusted		**Adjusted**	
	HR	**(95% CI)**	**HR**	**(95% CI)**
Vaginal delivery	**Ref**		**Ref**	
Intrapartum caesarean	**1.12**	(0.87-1.45)	**1.08**	(0.82-1.41)
Prelabour caesarean	**1.06**	(0.83-1.36)	**1.02**	(0.79-1.32)

Adjusted for birthweight for gestational age z score, parental age, parental occupation, maternal diabetes and hypertension, maternal region of birth, maternal ethnicity, IRSAD, remoteness, birth at public or private hospital, child's birth order, private or public healthcare, antenatal visit, maternal smoking

Table 5-4: (Supplementary Table) Complete case analysis: Sociodemographic and perinatal characteristics of children/parents by method of delivery

Perinatal and socioeconomic characteristics	**Vaginal Delivery**		**Caesarean Delivery**	
	N	**(%)**	**N**	**(%)**
Type 1 diabetes	**194,861**		**90,196**	
No	194,539	(99.8)	90,050	(99.8)
Yes	322	(0.2)	146	(0.2)
Mother's age (years)*	**194,861**		**90,198**	
	28.9 ± 5.6		30.6 ± 5.5	
Father's age (years)*	**185,124**		**86,890**	
	32.4 ± 6.5		33.7 ± 6.4	
Number of antenatal visits*	**179,376**		**81,663**	
	10.4 ± 3.0		10.6 ± 2.9	
BWGAZ*	**193,830**		**90,184**	
	-0.04 ± 1.0		0.07 ± 1.1	
Maternal diabetes (gestational and pre-existing)	**194,861**		**90,198**	
No	186,168	(96)	82,919	(92)
Yes	8,693	(4.5)	7,279	(8.1)
Maternal hypertension	**194,861**		**90,198**	
No	181,617	(93)	79,772	(88)
Yes	13,244	(6.8)	10,426	(12)
Maternal BMI	**64,431**		**31,004**	
<25 kg/m^2 (underweight and normal)	34,264	(53)	12,668	(41)
25 – <30 kg/m^2 (overweight)	17,034	(26)	8,741	(28)
≥30 kg/m^2 (obese/severely obese)	13,133	(20)	9,595	(31)
Maternal smoking	**192,261**		**88,513**	
Smoker	35874	(19)	13,686	(16)
Quit in pregnancy before first visit	7,633	(3.9)	3,651	(4.1)
Non-smoker	148,754	(77)	71,176	(80)
Child's birth order	**194,861**		**90,198**	
1st child	79,609	(41)	39,917	(44)
2nd child	67,656	(35)	32,048	(36)
3rd child	29,939	(15)	12,346	(14)
≥4 child	17,657	(9.0)	5,887	(6.5)
Parent's occupation	**194,218**		**89,958**	

Managers and professionals	75,515	(39)	39,139	(44)
Para professionals, tradesperson	73,397	(38)	33,897	(38)
Machines operators, drivers , labours	23,228	(12)	9,332	(10)
Students and pensioners and unemployed	22,078	(11.4)	7,590	(8.4)
Mother's region of birth	**194,826**		**90,184**	
Oceania and Antarctica	162,807	(84)	75,195	(83)
Europe	13,659	(7.0)	6,145	(6.8)
Africa, Asia, Americas	18,360	(9.4)	8844	(9.8)
Hospital category	**194,860**		**90,198**	
Private	42,620	(22)	29,514	(33)
Public	152,240	(78)	60,684	(67)
Healthcare of mother	**194,861**		**90198**	
Private	**52,594**	(27)	35,349	(39)
Public	142,267	(73)	54,849	(61)
Remoteness	**194,779**		**90,149**	
Major cities of South Australia	140,497	(72)	66,255	(75)
Inner regional South Australia	19,218	(9.9)	7,935	(8.8)
Remote & very remote South Australia	35,064	(18)	15,959	(18)
IRSAD-quintile	**194,693**		**90,124**	
Most disadvantaged (1)	56,006	(29)	23,249	(26)
2nd quintile	43,578	(22)	19,889	(22)
3rd quintile	32,511	(17)	14,814	(16)
4th quintile	36,954	(19)	18,673	(21)
Most advantaged (5)	25,644	(13)	13,499	(15)
Mother’s ethnicity	**194,861**		**90,197**	
White European ancestry	168,967	(87)	78,722	(87)
Aboriginal and Torres Strait Islander	5,801	(3.0)	2,540	(2.8)
Asians and others	20,093	(10)	8,935	(9.9)

Data presented are mean ± SD, all others are n (%)

IRSAD (Index of Relative Socioeconomic Advantage and Disadvantage)

BWGAZ (Birthweight for gestational age Z-score)

Table 5-5: (Supplementary Table) Complete case analysis: Caesarean birth, intrapartum and prelabor caesarean and risk of type 1 diabetes

	(Unadjusted)		**(Adjusted)**	
	(N = 285,057, T1D = 468)		(N = 246,821, T1D = 410)	
	HR	**(95% CI)**	**HR**	**(95% CI)**
Vaginal delivery	**Ref**		**Ref**	
All caesarean births	1.07	(0.88-1.31)	1.10	(0.89-1.37)
Vaginal delivery	**Ref**		**Ref**	
Intrapartum caesarean	1.12	(0.87-1.44)	1.16	(0.89-1.54)
Prelabor caesarean	1.04	(0.80-1.34)	1.04	(0.79-1.38)

Adjusted for birthweight for gestational age Z score, parental age, parental occupation, maternal diabetes and hypertension, maternal region of birth, maternal ethnicity, IRSAD, remoteness, public or private hospital, child's birth order, private or public health care, Antenatal visit, maternal smoking

Table 5-6: (Supplementary Table) Sensitivity analyses based on imputed data: Caesarean birth, intrapartum and prelabor caesarean and risk of type 1 diabetes

	BMI included[1] (n = 286,058)		**Singleton[2] (n = 277,215)**		**SA born[3] (n = 285,969)**		**2001-2013 births[4] (n = 240,646)**		**Combined[5] (n = 233,145)**	
	HR	**(95% CI)**	**HR**	**(95% CI)**	**HR**	**(95% CI)**	**HR**	**(95% CI)**	**HR**	**(95% CI)**
Vaginal delivery	**Ref**		**Ref**		**Ref**		**Ref**		**Ref**	
All caesarean births	**1.05**	(0.82-1.36)	**1.08**	(0.84-1.40)	**1.06**	(0.83-1.35)	**1.08**	(0.83-1.41)	**1.17**	(0.89-1.52)
Vaginal delivery	**Ref**		**Ref**		**Ref**		**Ref**		**Ref**	
Intrapartum caesarean	**1.08**	(0.81-1.45)	**1.09**	(0.81-1.47)	**1.10**	(0.83-1.47)	**1.07**	(0.76-1.50)	**1.15**	(0.82-1.61)
Prelabor caesarean	**1.02**	(0.74-1.41)	**1.07**	(0.77-1.49)	**1.02**	(0.75-1.38)	**1.09**	(0.78-1.51)	**1.18**	(0.85-1.65)

[1]Adjusted for maternal BMI, along with others confounders mentioned in the main analysis (Table 2 and 3, Supplementary Table 2)
[2]Restricted to singleton births, and adjusted for all confounders as model 1
[3]Restricted to South Australian born children, and adjusted for all confounders as model 1
[4]Only children born from July-2001 to December-2013 are included, and adjusted for all confounders as model 1
[5]Restricted to singleton children, born in South Australia, form July-2001 to December-2013, and adjusted for all confounders as model 1

CHAPTER 6 EFFECT OF MATERNAL SMOKING DURING PREGNANCY ON CHILDHOOD TYPE 1 DIABETES: WHOLE-OF-POPULATION STUDY

6.1 Preface

This Chapter contains the third study contributing to this thesis. This study was accepted for publication in Diabetologia at the time of thesis submission and published during examination. And it addresses the third question of this doctoral thesis.

What is the effect of maternal smoking in pregnancy on the risk of childhood type 1 diabetes? What is the risk of bias due to residual confounding in the effect estimate?

In Chapter 6, I studied the association between maternal smoking in pregnancy and risk of childhood type 1 diabetes, using a large (n=286,085) whole-of-population routinely-collected linked dataset from South Australia. Similar to caesarean birth, maternal smoking is a potentially modifiable risk factor that has been linked with type 1 diabetes with inconsistent evidence. Smoking during pregnancy has been linked with poorer health outcomes for the developing foetus, including, low birth weight, preterm birth and neurodevelopment issues.

Some reports of epigenetic modifications among children as a results of maternal smoking during pregnancy, and role of nicotine as an immune suppressant and its transfer through placenta makes maternal smoking in pregnancy a potential candidate for childhood type 1 diabetes pathogenesis.

Australia has both high maternal smoking in pregnancy rates and type 1 diabetes incidence. However, maternal smoking in pregnancy is declining and type 1 diabetes incidence is rising. Understanding these somewhat paradoxical trends was one reason why I wanted to explore the relationship between maternal smoking in pregnancy and childhood type 1 diabetes incidence. In addition, previous studies had numerous methodological imitations, which could be explored using contemporary epidemiological methods to better understand the potential for bias.

Highlights

- Maternal smoking in pregnancy is suggested to be associated with a small reduced risk of childhood type 1 diabetes.

- All studies in this area (including mine) face the issue of small numbers of children with type 1 diabetes exposed to maternal smoking in pregnancy, highlighting the need for larger studies, such as multi-country consortia to be able to conduct IPD meta-analyses

6.2 PUBLICATION 3: Effect of maternal smoking during pregnancy on childhood type 1 diabetes: Whole-of-population study

6.2.1 Statement of Authorship

Title of Paper	**Effect of maternal smoking during pregnancy on childhood type 1 diabetes: Whole-of-population study**
Publication Status	☐ Published ☑ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	

Principal Author

Name of Principal Author (Candidate)	Mumtaz Begum
Contribution to the Paper	MB prepared the draft, reviewed and revised the manuscript, analysed the data, and helped in conceptualizing and designing the study.
Overall percentage (%)	

Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	17/01/2020

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Rhiannon Megan Pilkington		
Contribution to the Paper	RMP acquired the data, provided intellectual input in statistical analyses and interpretation of the findings, and critically reviewed and edited the draft.		
Signature		Date	17-01-2020

Name of Co-Author	Catherine Ruth Chittleborough		
Contribution to the Paper	CRC acquired the data, provided intellectual input in statistical analyses and interpretation of the findings, and critically reviewed and edited the draft.		
Signature		Date	17/1/2020

Name of Co-Author	John Lynch		
Contribution to the Paper	JL, acquired the data, provided intellectual input in statistical analyses and interpretation of the findings, and critically reviewed and edited the draft.		
Signature		Date	17/1/20

Name of Co-Author	Megan Penno		
Contribution to the Paper	MP provided intellectual input, helped in interpreting the findings, and critically reviewed and edited the draft		
Signature		Date	14/01/2020

Name of Co-Author	Lisa Gaye Smithers		
Contribution to the Paper	LGS acquired the data, provided intellectual input in statistical analyses and interpretation of the findings, and critically reviewed and edited the draft.		
Signature		Date	17 JAN 2020

What is already known about this subject?

- Evidence is mixed about maternal smoking in pregnancy and risk of childhood type 1 diabetes
- Most case-control and population-based cohort studies reported reduced type 1 diabetes risk for children exposed to maternal smoking in pregnancy

What is the key question?

- What is the effect of maternal smoking during pregnancy on risk of childhood type 1 diabetes and what is the potential for bias in this effect due to unmeasured confounding?

What are the new findings?

- Maternal smoking in pregnancy was associated with 16% lower childhood type 1 diabetes incidence
- Our meta-analytic estimates demonstrated 28-29% reduced type 1 diabetes risk for children exposed to maternal smoking in pregnancy
- The negative control analysis indicated that some of the observed effect of prenatal maternal smoking on childhood type 1 diabetes was due to residual confounding. The E-value indicated that unmeasured confounding associated with prenatal maternal

smoking and childhood type 1 diabetes with a HR of 1.67 could negate the observed effect.

How might this impact on clinical practice in the foreseeable future?

Maternal smoking in pregnancy is not recommended but the mechanism leading to reduced type 1 diabetes should be investigated.

6.2.2 Abstract

Aims/hypothesis

Evidence of an association between maternal smoking during pregnancy (prenatal smoking) and childhood type 1 diabetes is mixed. Previous studies have been small and potentially biased due to unmeasured confounding. The objectives of this study were to estimate the association between prenatal smoking and childhood type 1 diabetes, assess residual confounding with a negative control design and an E-value analysis, and summarize published effect estimates from a meta-analysis.

Method

This whole-of-population study (births 1999-2013, aged ≤15-years) used de-identified linked administrative data from the South Australian Early Childhood Data Project. Type 1 diabetes was diagnosed in 557 children (ICD-10-AM codes, E10, E101-E109) during hospitalization (2001-2014). Families not given financial assistance for school fees was a negative control outcome. Adjusted Cox proportional hazard ratios (HR) were calculated. Analyses were conducted on complete-case (n = 264,542, type 1 diabetes = 442) and imputed (n = 286,058, type 1 diabetes = 557) data. A random-effects meta-analysis was used to summarize effects of prenatal smoking on type 1 diabetes.

Results

Compared with non-smokers, children exposed to maternal smoking only in the first or second-half of pregnancy had 6% higher type 1 diabetes incidence (adjusted-HR 1.06, 95% CI 0.73, 1.55). Type 1 diabetes incidence was 24% lower (adjusted-HR 0.76, 95% CI 0.58, 0.99) among children exposed to consistent prenatal smoking, and 16% lower for exposure to any maternal smoking in pregnancy (adjusted-HR 0.84, 95% CI 0.67, 1.08), compared with unexposed group. Meta-analytic estimates showed 28-29% lower risk of type 1 diabetes among children exposed to prenatal smoking compared with those not exposed. The negative control outcome analysis indicated residual confounding in the prenatal smoking and type 1 diabetes association. E-value analysis indicated that unmeasured confounding associated with prenatal smoking and childhood type 1 diabetes with a HR of 1.67 could negate the observed effect.

Conclusions/interpretation

Our best estimate from the study is that maternal smoking in pregnancy was associated with 16% lower childhood type 1 diabetes incidence, and some of this effect was due to residual confounding.

6.2.3 Introduction

Onset of type 1 diabetes can occur at any age, but many children who develop this condition have detectible autoantibodies targeting beta cell antigens within the first year of life[45], suggesting that early exposures may have a role in the natural history of type 1 diabetes.[145] Increasing trends, country variation in global type 1 diabetes incidence (0.01 to 60 per 100,000 per-year) and the type 1 diabetes discordance in monozygotic twins all suggest a role of non-genetic factors.[3, 6, 11] An environmental exposure implicated in type 1 diabetes pathogenesis is maternal smoking during pregnancy. Mechanisms that link prenatal smoking and childhood type 1 diabetes are not understood but may be associated with altered gene expression or immune function. [87, 172]

Type 1 diabetes incidence has increased in many countries that are ranked in the top ten for type 1 diabetes incidence including Finland, Sweden, USA and UK[3, 54] while prenatal smoking rates have declined or become stable in these countries in the last decade.[19] In Australia, type 1 diabetes incidence among 0-14 year olds has increased from 21.5 per 100,000 population in 2000, to 24.7 per 100,000 population in 2015 [175], whilst prevalence of smoking during late pregnancy reduced from 17.3% in 2006 [176] to 9.9% in 2016. [177] Previous studies of prenatal smoking and childhood type 1 diabetes reported mixed findings, demonstrating increased [26], decreased [29-31], and null type 1 diabetes risk.[34] A Swedish HLA-genotype-matched case-control study demonstrated an increased type 1 diabetes risk for

children exposed to prenatal smoking[26], and a UK record-linkage study did not find any difference in the type 1 diabetes incidence between children exposed and unexposed to prenatal smoking.[34] Conversely, large population-based cohort studies have demonstrated 25-35% lower type 1 diabetes risk for children exposed to prenatal smoking.[29, 30] These mixed findings of an association between prenatal smoking and childhood type 1 diabetes could be due to differences in confounding adjustments. Many population-based studies have adjusted for confounders such as maternal age[26, 29-31], maternal deprivation or socioeconomic position, [30, 31] birth order or parity,[26, 29, 30] maternal education, pre-pregnancy BMI and diabetes, and some have adjusted for mediators.[26, 30, 31] Some studies have not adjusted for father's age, pre-pregnancy diabetes or hypertension, ethnicity or socioeconomic indicators, and one study excluded children whose mothers had pre-existing diabetes.[26, 30, 31] Most studies on smoking during pregnancy and type 1 diabetes risk have small numbers of children exposed to smoking during pregnancy, ranging from 5 to 78 in population-based cohort studies [27, 29, 30] and 29 to 258 in case-control studies.[26, 31, 32, 309, 310, 327] Small sample sizes could be another reason for variable findings in previous studies, therefore, one way to obtain a more precise effect estimate of the prenatal smoking and type 1 diabetes association is to combine these estimates in a meta-analysis.

The objectives of this study were: firstly to estimate the association between prenatal maternal smoking and childhood type 1 diabetes incidence with adjustment for a range of confounding factors defined a priori; secondly, to measure the potential for bias due to unmeasured confounding using a negative control outcome analysis and E-value calculation;

and, to perform a meta-analyses of published population-based cohort and case-control studies.

6.2.4 Methods

6.2.4.1 Study population and design

We used data from the South Australian Early Childhood Data Project, which consists of routinely collected, de-identified, linked government administrative datasets.[256] Datasets were linked by an independent agency (SA NT Datalink).[269] Children were linked across datasets using a probabilistic algorithm that included demographic information such as name, date of birth, sex and address. A 0.5% false linkage rate has been reported in Australia.[269] This study used inpatient hospitalization data from July 2001 to June 2014, and perinatal and birth registration data from 1999-2013 (Figure 6-1).

6.2.4.2 Type 1 diabetes

Individuals with Type 1 diabetes (aged <15 years) were identified from inpatient hospitalization data from all public hospitals in South Australia (SA). Children with index type 1 diabetes hospitalizations were identified using International Classification of Disease, 10th edition, Australian Modification (ICD-10-AM) codes (E10, ranging E101 to E109),

including both principal and additional diagnoses.[264] Diagnoses codes were assigned to each hospitalisation episode by trained hospital staff.

6.2.4.3 Maternal smoking during pregnancy

Data on prenatal smoking was obtained from the South Australian Perinatal Statistics Collection [259] from 1999-2013. It is mandatory for perinatal statistics to be collected for every child born in South Australia. Data are collected by midwives or nurses using a standard data collection form. The perinatal data have been collected by the South Australian government since 1981 to track mother and child health indicators.[177, 259] The data collection form has been validated against an audit of medical records.[261]

Maternal smoking data were collected at the first antenatal visit (≤20 weeks gestation) and in the second-half of pregnancy (≥20 weeks gestation).[259] Most women (74.3%) had their first antenatal visit before 14 weeks gestation.[259] For the primary analysis smoking was categorized into non-smokers, smokers only in first or second half of pregnancy, and consistent smokers. For women who smoked only in the first or second half of pregnancy, no information is available on when they started or stopped smoking, therefore the duration of smoking is unclear. Due to small numbers of offspring with type 1 diabetes among women who smoked, smoking in pregnancy was dichotomized into non-smokers or smokers (including women who smoked at any time during pregnancy) for secondary analyses.

6.2.4.4 Confounding

Confounders of the association between maternal smoking during pregnancy and type 1 diabetes risk were identified a priori based on literature and by using a directed acyclic graph (Figure 6-2, Supplementary). Information on confounders was obtained from the South Australian Perinatal Statistics Collection.[259] Parental characteristics included mother's and father's age (continuous variables), and parental highest occupation in four categories: (1) Managers, administrators and professionals;(2) Para-professionals, tradespersons, clerks, salespersons, and personal service; (3) Plant and machine operators, drivers and labourers; (4) Students, pensioner, home duties and unemployed. Maternal and child characteristics were maternal birth region (Oceania, Europe, or Africa, Asia or America), mother's ethnicity (European decent; Aboriginal or Torres Strait Islander; Asian and other), whether mother was a private or public hospital patient, type of hospital where child was born (private/public), maternal pre-pregnancy hypertension (yes/no), pre-pregnancy diabetes (yes/no), parity (1^{st}, 2^{nd}, 3^{rd}, and $\geq 4^{th}$) and the child's year of birth. Area-level measures of socioeconomic conditions (Index of Relative Socioeconomic Advantage and Disadvantage, IRSAD), and remoteness and accessibility (Australian Remoteness Index for Areas) were based on mother's postcode at the time of birth. Maternal BMI at the first antenatal visit (<25; 25 to <30; ≥30 kg/m^2) was included as a confounder only in sensitivity analysis as maternal BMI data was not collected for births prior to 2007.

6.2.4.5 Statistical Analysis

The crude and adjusted association between prenatal smoking and childhood type 1 diabetes for both primary and secondary analyses was estimated by Cox proportional hazard regression, to account for differences in observation time across successive birth years. These analyses were conducted on both complete-case (Table 6-5, Supplementary) and imputed data (Table 6-2). Schoenfeld residual tests demonstrated non-violation of the proportional hazard assumption. Children were followed from birth until the diagnosis of type 1 diabetes or the end of follow-up (June 2014). The observation time ranged from one year (for 2013 births) to 15 years (for 1999 births), with a mean follow-up of eight years.

A negative control outcome analysis was used to investigate whether any association between maternal smoking in pregnancy and type 1 diabetes could be due to unmeasured confounding.[285] Cox proportional hazard regression analysis estimated the association between maternal smoking during pregnancy and the family not having a school card for financial assistance with school fees. An assumption about negative control outcomes is that the measured and unmeasured confounding pattern for the association between maternal smoking and type 1 diabetes is the same as for the maternal smoking and no school card association (called 'U-comparable').[285] There is no plausible reason for prenatal maternal smoking to directly cause a child to get a school card. If there is any association between maternal smoking in pregnancy and child not having a school card, it must be through confounders (e.g. sociodemographic characteristics), additional backdoor paths or unmeasured confounding

(Figure 3-5 is given in Chapter 3). School card data was sourced from the school enrolment census and was provided by the Department of Education, South Australia. Data were only available for children who had started school (complete-case analysis: n=277,370 [no school card , n=84531]; imputed data: n=184,663[no school card, n= 149,670]).

The E-value was calculated to measure the potential for bias due to unmeasured confounding in the prenatal smoking and type 1 diabetes association. The E-value quantifies the minimum strength of an association that an unmeasured confounder would need to have with the exposure and outcome to negate the observed association between prenatal smoking and type 1 diabetes, given the measured confounders.[288] The E-value for the CI quantifies the strength of an association that unmeasured confounding would need to have with prenatal smoking and the childhood type 1 diabetes, above and beyond the measured confounders, to change the CI to include the null.[288]

Five sensitivity analyses were performed to see if the association between prenatal smoking and childhood type 1 diabetes was similar to the main findings. In the first sensitivity analysis, we examined whether adjusting for pre-pregnancy BMI (in addition to all other confounders) influenced the association between prenatal smoking and type 1 diabetes. Maternal BMI was only collected from 2007 onwards and therefore could not be included in the main analyses. The second sensitivity analysis was restricted to births from July-2001 to December-2013, as hospital data was only available from July-2001 and any diagnosis of type 1 diabetes from 1999 to mid-2001 may have been misclassified. As perinatal data was

only collected for South Australian born children, the third sensitivity analysis was restricted to children born in South Australia, for whom more complete data were available. The fourth sensitivity analysis was restricted to singleton births, to make our study comparable with previous studies that were conducted only on singleton births.[29, 30] Finally, all the above were combined, and the analysis was restricted to singleton children born in South Australia from July-2001 to December-2013. All the sensitivity analyses were adjusted for maternal BMI along with all other confounders.

The amount of missing information on the exposure and confounders in Table 6-1 ranged from 0.03% to 0.35%, except prenatal maternal smoking (3.0%), father's age (4.6%), and maternal BMI (66.6%; not included in primary analysis). Multiple imputation by chained equations was conducted to maintain the association between prenatal smoking and childhood type 1 diabetes, and to account for potential bias if the association differs between children with and without complete data.[290] The outcome variable (type 1 diabetes) was not imputed. The outcome variable, all variables in Table 6-1 and the sensitivity analyses, and the Nelson-Aalen estimator of the cumulative hazard were included in the imputation models.[328] Summary statistics (Table 6-1) and estimates in Table 6-2 and 6-3 were derived by combining the 20 imputed datasets, using Rubin's rules. Analysis from complete-case data are presented in Supplementary Tables (Table 6-5 and Table 6-6).

6.2.4.6 Meta-analysis

Even the largest studies investigating the effect of prenatal maternal smoking on type 1 diabetes only included relatively small numbers of children with type 1 diabetes who were born to mothers who smoked during pregnancy (number ranging from 42 to 78 in population-based cohort studies).[29, 30] Therefore, estimates from individual studies are imprecise. To compute a more precise estimate, a meta-analysis of the current results with previous studies that reported an association between prenatal smoking and childhood type 1 diabetes was conducted in January 2019. PubMed, Web of Science, and EMBASE databases were systematically searched for studies on type 1 diabetes, maternal smoking during diabetes and related terms, without limiting year of publication. Population-based studies written in English that reported maternal smoking during pregnancy as the exposure, and overt or clinical type 1 diabetes in childhood (<19 years) as an outcome were included. Studies that reported beta cell autoimmunity (preclinical type 1 diabetes) as an outcome were excluded. Only population-based studies with some attempt to adjust for confounding were included in the analysis.

Separate random-effects model were performed for the meta-analyses of population-based cohort and case-control studies. The meta-analysis of population-based studies pooled the HRs and the meta-analysis of the case-control studies pooled the odds ratios. All population-based-cohort studies reported HRs and all the case-control studies reported odds ratios. We used a random-effects model as we did not assume homogeneity of effects among studies.

Analyses were conducted in Stata SE version 15.0 (Stata Corp, College Station, Texas, USA).

6.2.4.7 Ethics Approval

Ethics approval was granted by the Human Research Ethics Committees of the South Australian Department of Health (HREC/13/SAH/106), and the Aboriginal Health Council of SA (04-13-538).

6.2.5 Results

A total of 286,058 children (aged ≤15 years) born from 1999-2013, contributed to 2,200,252 person-years of data. During follow up 557 children were diagnosed with type 1 diabetes: an incidence of 25.3 per 100,000 person-years. Among 286,058 children, 62,216 were born to mothers who smoked during pregnancy.

Of the 557 children diagnosed with type 1 diabetes from 2001 to 2014, 118 were exposed to maternal smoking during pregnancy, with 80 exposed to consistent smoking in both the first and second half of pregnancy. The crude type 1 diabetes incidence was 26.3 per 100,000 person-years for children not exposed, and 22.2 per 100,000 person-years for children exposed to maternal smoking.

The numbers in Table 6-1 shows that overall, socioeconomically disadvantaged women, who were from low income occupations, living in most disadvantaged areas, younger at the child's birth, and delivered in a public hospitals had higher prevalence of smoking during pregnancy. The distributions of these characteristics were similar in both the complete-case and imputed analyses (Table 6-1). There were numerically more socioeconomically disadvantaged women in the group that consistently smoked throughout pregnancy (Table 6-4, Supplementary).

6.2.5.1 Primary and Secondary Analyses

In the primary analysis (Table 6-2), following adjustment for confounding, type 1 diabetes incidence was 6% higher for children whose mothers smoked only in first-half or second-half of pregnancy (HR 1.06, 95% CI 0.73,1.55), and 24% lower incidence for children exposed to consistent prenatal maternal smoking (HR 0.76, 95% CI 0.58, 0.99), compared with unexposed children. For smoking in the first or second-half of pregnancy, CIs were wide, ranged from 27% reduced to 55% increased type 1 diabetes incidence. For consistent smoking, CIs ranged from 42% reduced to almost no difference in type 1 diabetes incidence between children exposed and unexposed to maternal smoking.

In secondary adjusted analysis, when the exposure included any smoking (first or second half of pregnancy, and consistent smoking), childhood type 1 diabetes incidence was 16% lower for children exposed to maternal smoking in pregnancy (HR 0.84, 95% CI 0.67, 1.08)

compared with those unexposed. Again, the confidence intervals were wide. Complete-case analyses showed similar associations (Table 6-5, Supplementary).

6.2.5.2 Potential for unmeasured confounding

The negative control outcome analysis (Table 6-3) demonstrated 15% reduced incidence of not having a school card (HR 0.85, 95% CI 0.83, 0.87) for children exposed to consistent prenatal maternal smoking, and 14% reduced incidence of not having a school card (HR 0.86, 95% CI 0.85, 0.88) related to exposure to any maternal smoking in pregnancy after adjustment for confounding. Complete-case analysis showed similar pattern (Table 6-6, Supplementary).

The E-value for the observed point-estimate (HR 0.84, 95% CI 0.67, 1.08) of prenatal smoking and childhood type 1 diabetes was 1.67. The observed 16% reduced incidence of type 1 diabetes for children exposed to prenatal smoking could be explained away by unmeasured confounding that was associated with prenatal smoking and the childhood type 1 diabetes by an HR of 1.67 each, above and beyond the measured confounders. The observed 95% CI already included the null value (95% CI 0.67, 1.08), therefore the E-value for the CI was 1, suggesting that no unmeasured confounding would be needed to move the CI to include the null. For example in our study, type 1 diabetes in the father is a potential unmeasured confounder and a strong predictor of type 1 diabetes in the offspring (paternal type 1 diabetes vs. no paternal type 1 diabetes OR 9.19, 95% CI 3.8, 22.0),[310] which could indirectly impact maternal smoking through father's education and socioeconomic position.[36]

As the E-value indicated the "unmeasured confounder and outcome association", and the "unmeasured confounder and exposure association", each would need to be equal to a HR of 1.67 to negate the observed effect. Although type 1 diabetes in the father is a strong predictor of type 1 diabetes in the offspring, we do not know the strength of the association between type 1 diabetes in the father and maternal smoking in pregnancy. In addition, the low prevalence of type 1 diabetes in the father (<1%) also reduces the potential to confound the maternal smoking and childhood type 1 diabetes association. Therefore, type 1 diabetes in the father may not be a strong enough unmeasured confounder to negate the observed effect of maternal smoking in pregnancy on childhood type 1 diabetes in this study.

6.2.5.3 Sensitivity Analyses

The sensitivity analyses (Table 6-7, Supplementary) were consistent with the main findings.

6.2.5.4 Meta-analysis

In addition to the current study, there were four previous population-based studies describing an association between prenatal smoking and type 1 diabetes available for meta-analysis (Figure 6-3). The meta-analysis showed 28% lower type 1 diabetes incidence for children exposed to prenatal smoking (HR 0.72, 95% CI 0.62, 0.82) compared with unexposed children. Similarly, the meta-analysis of six case-control studies (Figure 6-4, Supplementary) demonstrated 29% reduced type 1 diabetes risk for children whose mothers smoked during pregnancy (OR 0.71, 95% CI 0.55, 0.86) compared with those children whose mothers had not smoked during pregnancy.

6.2.6 Discussion

In this large whole-of-population study, type 1 diabetes incidence was lower for children exposed to maternal smoking in pregnancy compared with children unexposed, after adjusting for a wide range of confounders. Similar to studies in this area, small numbers of type 1 diabetes cases among children exposed to prenatal smoking has impacted on the precision of the effect estimates. The CIs around the adjusted effect estimates in our primary and secondary analyses were wide, but on balance, provided some evidence to suggest a lower incidence of type 1 diabetes (consistent smokers HR 0.76, 95% CI 0.58, 0.99). The crude absolute risk reduction was small, with 4 fewer type 1 diabetes cases per 100,000 person-years among children exposed to maternal smoking in pregnancy versus unexposed children The 6% increased type 1 diabetes incidence for children whose mothers smoked only in the first or second half of pregnancy is difficult to interpret, again because of the very wide confidence intervals (95% CI 0.73,1.55), and the small number of type 1 diabetes cases associated with mothers who smoked (n=38) in the first or second half of pregnancy. In addition, 69% of the 38 women who smoked only in the first or second half of pregnancy and had a child with type 1 diabetes reported quitting before or at their first antenatal visit. This suggests that the exposure to smoking during pregnancy among this group is mostly limited to the first trimester. After triangulating across the main results and the meta-analysis,

and despite the negative control outcome analysis indicating the likelihood of a small amount of unmeasured confounding, the evidence suggests a lower type 1 diabetes incidence for children exposed to maternal smoking during pregnancy as compared with unexposed children. However, because of the unmeasured confounding, the effect of smoking on type 1 diabetes is likely to be smaller than the point estimates suggest. As indicated by the E-value, unmeasured confounding associated with prenatal smoking and childhood type 1 diabetes with an HR of 1.67 could negate the observed effect.

Smoking will never be recommended as an intervention for type 1 diabetes due to the significant harm it causes to both the mother and the foetus. The consequences of prenatal smoking have been well researched and include increased risk of miscarriage, preterm delivery, low birthweight, childhood obesity, respiratory problems, neurodevelopment and behavioural consequences. [329, 330] It is not known which component of tobacco, nicotine or other combustible chemicals, may be an active factor associated with reduced type 1 diabetes risk. Our findings suggest the mechanism of the effect of maternal smoking during pregnancy on type 1 diabetes (e.g. immune suppression by nicotine exposure, or alterations in gene expression) needs further investigation.

The prevalence of smoking in first half of pregnancy (23.8% in 1999, 11.2% in 2013) in our study is similar to reports of smoking in early pregnancy from Australia (11.3% in 2013),[331] and Scandinavia in 2009 (Denmark 12.5%, Norway 16.5%, Finland 15%).[332] Of the women who reported smoking and gave birth from 1999-2013, 25% quit at or before their first

antenatal visit or only smoked in first half of pregnancy, consistent with Australian national reports.[177] Consistent smoking in both first and second half of pregnancy in our study (20.7% in 1999, 9.7% in 2013) is similar to Scandinavian estimates (~10% in Finland and Denmark in 2009).[332]

Our findings of a lower type 1 diabetes incidence following exposure to prenatal smoking were similar to findings of previous case-control and population-based cohort studies, however, the effect sizes in our study (16-24%) were smaller than previously reported estimates (25-35%).[29-31, 33, 310] In our study, adjusting for a range of confounders did not considerably attenuate the effect estimate of maternal smoking on childhood type 1 diabetes. The number of children with type 1 diabetes exposed to prenatal smoking is small across previous studies, ranging from 42 to 72 in other population-based studies from Australia and Norway.[29, 30] This is similar to our study, with 118 out of 557 children with type 1 diabetes exposed to maternal smoking in pregnancy. A meta-analysis was undertaken in our study to address the small samples and indicated that there was a 28% reduced risk of type 1 diabetes using data from population-based studies [27, 29, 30] and 29% reduced risk using data from case-control studies.[26, 31-33, 309, 310] These results involve six countries (Norway, Denmark, Sweden, UK, US, and Australia), and 9,872 children with type 1 diabetes of which 16% were exposed to prenatal smoking. Our meta-analytic estimates are consistent with a recent meta-analysis.[186] However, our results are inconsistent with the HLA-matched case-control study that reported increased type 1 diabetes risk for children exposed to prenatal smoking.[26] Whilst the contribution of the HLA system may explain up to 60% of genetic risk of type 1 diabetes,

it is plausible that other genetic risk variants contributing to susceptibility[75] may be more or less prevalent in the type 1 diabetes cases compared with controls in the HLA-matched study.

Many studies such as the Australian-based Environmental Determinants of Islet Autoimmunity ENDIA study[333] have been established in at-risk populations to elucidate mechanisms that lead to type 1 diabetes, and current thinking is that there are likely to be multiple mechanisms that link prenatal smoking with type 1 diabetes.[165] Epigenetic modifications induced by prenatal smoking, such as DNA methylation, may lead to changes in gene expression. Differences in DNA methylation have been reported among children exposed and unexposed to prenatal smoking,[87] and among children with type 1 diabetes and their type 1 diabetes-discordant twins.[170] In addition, nicotine, a known immune suppressant reported to effect both the innate and adaptive immune responses, can pass through the placenta to foetal circulation.[172] The suppressive impact of smoking on immunity increases the risk of many chronic diseases, but there is some suggestion it could be protective for an autoimmune disease like type 1 diabetes. Smoking during pregnancy has been associated with reduced risks of other diseases with immune elements such as pre-eclampsia, Parkinson's disease, ulcerative colitis, and sarcoidosis.[334-337] The presence of autoantibodies in preeclampsia suggests that it may be a pregnancy-induced autoimmune disease.[338] High risk HLA genotypes for type 1 diabetes are also reported to be involved in ulcerative colitis, [339] suggesting there may be a similar underlying mechanism of smoking in modifying the risk of both of these diseases.

This study included children attending all public hospitals in South Australia, therefore children with type 1 diabetes who never attended any public hospital might have been missed. However, we believe that case ascertainment is high in this study, because, both primary and secondary diagnoses codes were used, and thus even children who are hospitalized for other reasons were included if type 1 diabetes was noted as their secondary diagnosis. In addition, the South Australia paediatric hospital with specialized endocrinology unit and where children are admitted for stabilization after diabetes diagnosis, was included in this study. In a similar setting in Western Australia, 99.8% of type 1 diabetes cases on the state diabetes register were ascertained in public hospital data.[267] Although social desirability bias in collecting maternal smoking is a common issue in observational studies, a Swedish validation study demonstrated that of the women who reported no smoking in pregnancy, 95% were classified as non-smokers based on serum cotinine concentration.[171] In our study, there could be some measurement error in maternal smoking during pregnancy. However, we do not have information about the extent of mismeasurement in the exposure to smoking as the data are not available. Furthermore, the capture of perinatal data is via a form validated against medical records. There may be misclassification of mother's tobacco exposure, because we do not have data on father's smoking, which can affect the foetus through maternal passive smoking. Another limitation is the lack of information on dose or number of cigarette smoked daily, in pregnancy. We used the E-value and negative control outcome to indicate the potential for bias due to unmeasured confounding. The E-value is dependent on the validity of the effect estimate and could be biased if there were selection and measurement biases.

We have tried to improve the internal validity by adjusting for a range of potential confounders, using routinely collected data (reduced risk of recall-bias), and conducting multiple imputation to account for bias due to loss of information. In addition, the negative control outcome analysis is assumed to have exact same set of measured and unmeasured confounders (perfectly U-comparable), as the maternal smoking and childhood type 1 diabetes association. It is rare to have a perfectly U-comparable negative control outcome, and it is most likely to only be approximately U-comparable.[285] However, because of non-perfect U-comparability, the negative control outcomes suggests that the association between smoking and type 1 diabetes is not entirely due to residual confounding.

6.2.7 Conclusion

Our best estimate from this study is there is 16% reduced incidence of type 1 diabetes for children exposed to any maternal smoking in pregnancy as compared to unexposed children, but some of this effect is likely due to residual confounding. With the current analyses we cannot rule out an effect of maternal smoking during pregnancy on childhood type 1 diabetes. However, the absolute reduction in type 1 diabetes cases among children exposed to smoking was small (4 cases per 100,000 person-years). This study along with the results of similar population-based studies, suggests that the mechanism leading to the maternal smoking and type 1 diabetes association needs to be investigated.

6.2.8 Tables and Figures

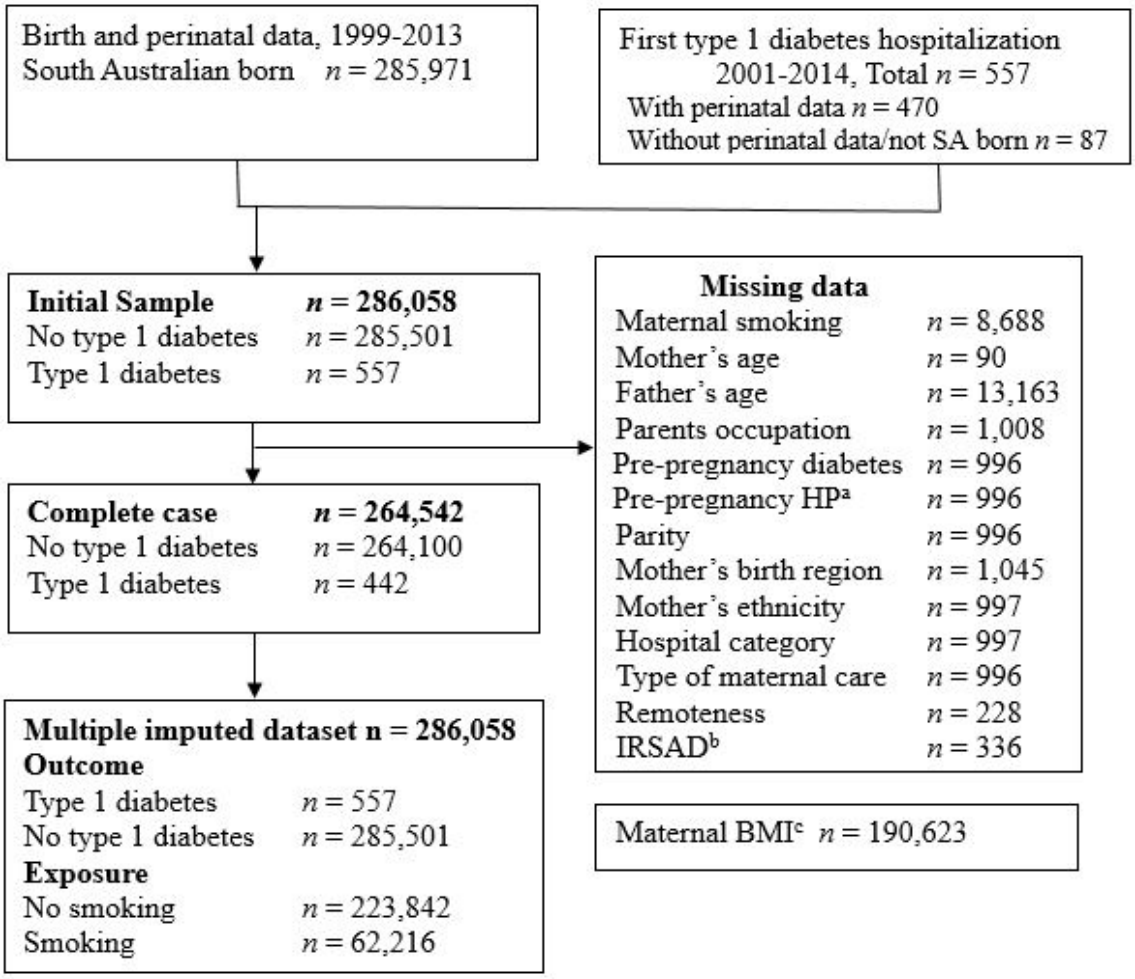


Figure 6-1: Flow chart of the study population

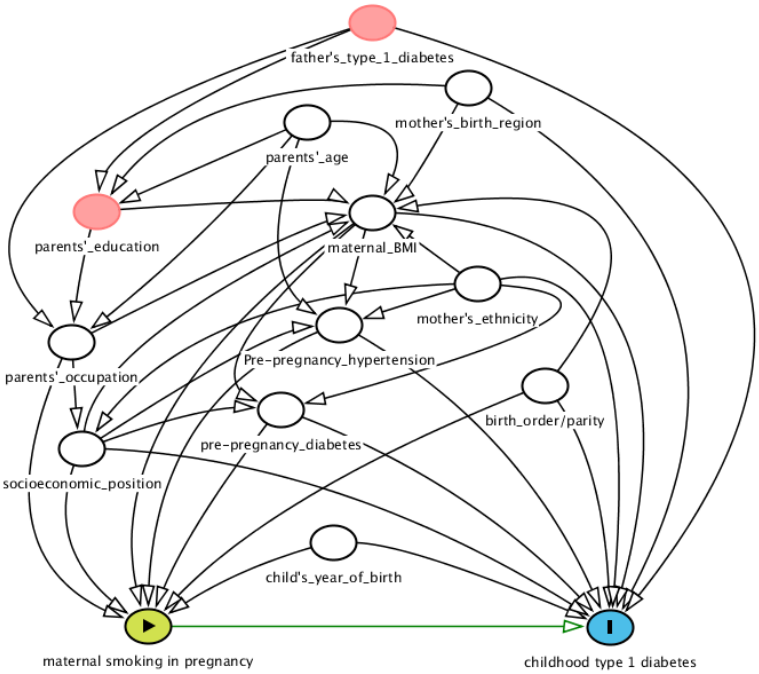


Figure 6-2: (Supplementary Figure) Directed acyclic graph showing proposed confounding structure

Exposure = Maternal smoking in pregnancy

Outcome = Childhood type 1 diabetes

Confounders = Adjusted for in our model

Confounders = Not measured, and not adjusted for in our model

Path showing unmeasured confounding =

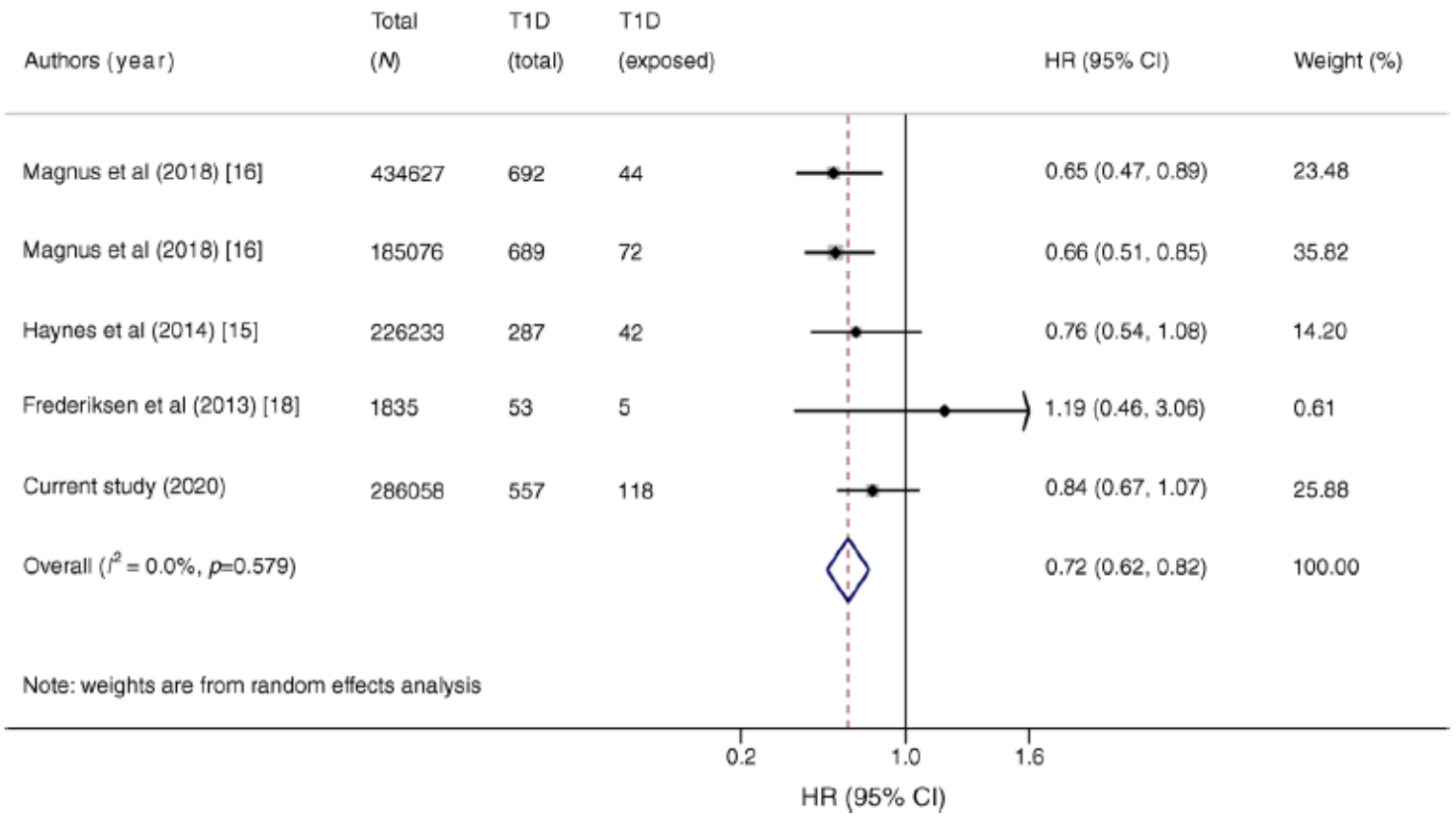


Figure 6-3: Maternal smoking during pregnancy and risk of type 1 diabetes (T1D) in population-based studies

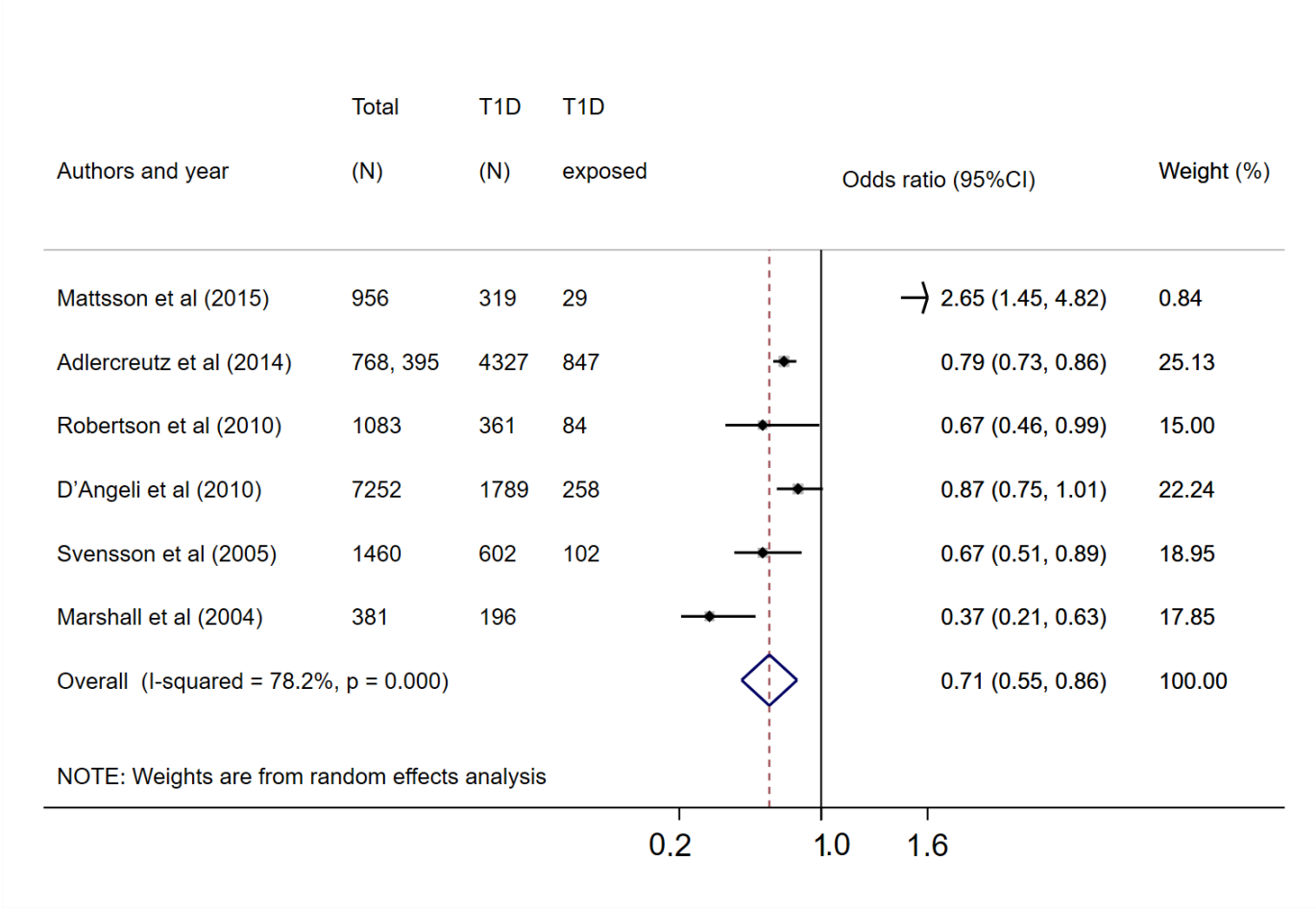


Figure 6-4: (Supplementary Figure) Maternal smoking during pregnancy and risk of type 1 diabetes (T1D) in case-control studies

Table 6-1: Characteristics of children by maternal smoking during pregnancy (N=286,058)

	Imputed analysis N = 286,058[a]		**Complete-case analysis N = 264,542**	
	No smoking n=223,842	**Smoking n=62,216**	**No smoking n=212,414**	**Smoking n=52,128**
Type 1 diabetes				
No	223403 (99.8)	62098 (99.8)	212064 (99.8)	52036 (99.8)
Yes	439 (0.2)	118 (0.2)	350 (0.2)	92 (0.2)
Mother's age (years) (Mean ± SD)	30.1 ± 5.4	27.4 ± 6.1	30.1 ± 5.3	27.6 ± 6.0
Father's age (years) (Mean ± SD)	33.2 ± 6.4	31.0 ± 7.0	33.3 ± 6.3	31.1 ± 6.9
Pre-pregnancy diabetes				
No	222591 (99.4)	61759 (99.3)	211283 (99.5)	51774 (99.3)
Yes	1251(0.6)	457 (0.7)	1131 (0.5)	354 (0.7)
Pre-pregnancy hypertension				
No	221030 (98.7)	61475 (98.8)	209763 (98.8)	51524 (98.8)
Yes	2812 (1.3)	741 (1.2)	2651 (1.2)	604 (1.2)
Hospital category (for child birth)				
Private	67304 (30.1)	5191 (8.3)	65446 (30.8)	4830 (9.3)
Public	156538 (69.9)	57025 (91.7)	146968 (69.2)	47298 (90.7)
Mother type of patient (Healthcare)				
Private	81670 (36.5)	6704 (10.8)	78018 (36.7)	6014 (11.5)
Public	142172 (63.5)	55512 (89.2)	134396 (63.3)	46114 (88.5)
Parity				
1st child	95457 (42.6)	24419 (39.2)	90323 (42.5)	20622 (39.6)
2nd child	81299 (36.3)	18754 (30.1)	77871 (36.7)	16098 (30.9)
3rd child	32074 (14.3)	10392 (16.7)	30440 (14.3)	8696 (16.7)
$\geq 4^{th}$ child	15011 (6.7)	8650 (13.9)	13780 (6.5)	6712 (12.9)
Parents' highest occupation				
Managers, administrators	103424 (46.2)	12088 (19.4)	99549 (46.9)	10994 (21.1)
Para-professional, tradespersons	83595 (37.3)	24295 (39)	80009 (37.7)	21578 (41.4)
Plant machine operators	20554 (9.2)	12203 (19.6)	19459 (9.2)	10598 (20.3)
Students, pensioners	16269 (7.3)	13630 (21.9)	13397 (6.3)	8958 (17.2)
Mother's birth region				
Oceania	180678 (80.7)	58150 (93.5)	171221 (80.6)	48533 (93.1)
Europe	17029 (7.6)	2869 (4.6)	16268 (7.7)	2528 (4.8)
Africa, Asia, Americas	26135 (11.7)	1197 (1.9)	24925 (11.7)	1067 (2.0)
Mother's ethnicity				
Caucasian	192995 (86.2)	55577 (89.3)	183646 (86.5)	47587 (91.3)
Aboriginal or Torres Strait Islander	3283 (1.5)	5083 (8.2)	2543 (1.2)	3191 (6.1)

Asians and others	27564 (12.3)	1555 (2.5)	26225 (12.3)	1350 (2.6)
Remoteness				
Major cities	166731 (74.5)	40835 (65.6)	158572 (74.7)	34628 (66.4)
Inner regional	21161 (9.5)	6147 (9.9)	20054 (9.4)	5165 (9.9)
Outer regional/remote	35950 (16.1)	15233 (24.5)	33788 (15.9)	12335 (23.7)
IRSAD[b]				
Most disadvantaged (1)	54062 (24.2)	25441 (40.9)	50598 (23.8)	20661 (39.6)
2nd quintile	48338 (21.6)	15359 (24.7)	45861 (21.6)	12888 (24.7)
3rd quintile	38442 (17.2)	9111 (14.6)	36669 (17.3)	7827 (15.0)
4th quintile	47752 (21.3)	8152 (13.1)	45731 (21.5)	7154 (13.7)
Most advantaged (5)	35247 (15.7)	4153 (6.7)	33555 (15.8)	3598 (6.9)
Maternal BMI[c]				
<25 kg/m^2 (Underweight/normal)	114211 (51)	29006 (46.6)	38046 (50.4)	6715 (44.5)
25 to <30 kg/m^2 (Overweight)	59981 (26.8)	17167 (27.6)	20500 (27.1)	4044 (26.8)
≥30 kg/m^2 (Obese/severely obese)	49649 (22.2)	16042 (25.8)	17005 (22.5)	4327 (28.7)

Data are presented as mean ± SD or n(%)

[a]For the imputed analysis, the n values of the subgroups of the following do not equate to the total due to rounding: parity (both groups), mother's ethnicity (smoking), remoteness (smoking), IRSAD (no smoking), maternal BMI (both groups)

bFor IRSAD, first quintile is most disadvantaged, fifth quintile is most advantaged

cMissing data: complete-case analysis/no smoking, n=136,863; complete-case analysis/smoking, n=37,042

HT, hypertension; IRSAD, Index of Relative Socioeconomic Advantage and Disadvantage

Table 6-2: Maternal smoking during pregnancy and risk of childhood type 1 diabetes (T1D) (Total n = 286,058, type 1 diabetes n = 557)

		Unadjusted		Adjusted[c]	
	T1D n	**HR[a]**	**(95%CI[b])**	**HR**	**(95%CI)**
Maternal smoking in pregnancy					
Non smoking	439	**Ref**		**Ref**	
Smoked only in first or second half of pregnancy	38	**1.06**	(0.73, 1.54)	**1.06**	(0.73, 1.55)
Consistent smoking	80	**0.72**	(0.56, 0.93)	**0.76**	(0.58, 0.99)
Maternal smoking in pregnancy					
Non smoking	439	**Ref**		**Ref**	
Smoking[d] (any smoking in pregnancy)	118	**0.80**	(0.64, 1.01)	**0.84**	(0.67, 1.08)

[a]HR (hazard ratio)

[b]CI (confidence interval)

[c]Adjusted for parents' age, parents' occupation, mother's birth region, maternal ethnicity, remoteness, IRSAD (Index of Relative Socio-economic Advantage and Disadvantage), mother type of patient, hospital category, parity, pre-pregnancy hypertension, pre-pregnancy diabetes, and child's year of birth

[d]Smoking (combined consistent smoking, and smoking only in first or second half of pregnancy)

Table 6-3: Negative control outcome analysis: Association between maternal smoking during pregnancy and child not having a school card (Total n = 184,663, No school card n = 149,670)

			Unadjusted		Adjusted[c]
	No school card	**HR[a]**	**(95%CI[b])**	**HR**	**(95%CI)**
Maternal smoking in pregnancy		**Ref**		**Ref**	
Non smoking	119,843				
Smoked only in first or second half of pregnancy	8,619	**0.90**	(0.88, 0.93)	**0.89**	(0.87, 0.92)
Consistent smoking	21,208	**0.80**	(0.78, 0. 81)	**0.85**	(0.83, 0.87)
Maternal smoking in pregnancy					
Non smoking	119,843	**Ref**		**Ref**	
Smoking[d] (any smoking in pregnancy)	29,827	**0.82**	(0.81, 0.84)	**0.86**	(0.85, 0.88)

[a]HR (hazard ratio)

[b]CI (confidence interval)

[c]Adjusted for parents' age, parents' occupation, mother's birth region, maternal ethnicity, remoteness, IRSAD (Index of Relative Socio-economic Advantage and Disadvantage), mother type of patient, hospital category, parity, pre-pregnancy hypertension, and pre-pregnancy diabetes

[d]Smoking (combined consistent smoking, and smoking only in first or second half of pregnancy)

Supplementary Tables

Table 6-4: (Supplementary Table) Complete-case analysis: Characteristics of children by maternal smoking during pregnancy (N=264,542)

	No smoking	**Smoking only in 1st or 2nd half of pregnancy**	**Consistent smoking**
	n=212,414	**n=14,315**	**n=37,813**
Type 1 diabetes			
No	212064 (99.8)	14286 (99.8)	37750 (99.8)
Yes	350 (0.2)	29 (0.2)	63 (0.2)
Mother's age (years) (Mean ± SD[a])	30.1 ± 5.3	28.0 ± 5.8	27.5 (6.0)
Father's age (years) (Mean ± SD)	33.3 ± 6.3	31.1 ± 6.5	31.1 (7.0)
Pre-pregnancy diabetes			
No	211283 (99.5)	14245 (99.5)	37529 (99.2)
Yes	1131 (0.5)	70 (0.5)	284 (0.8)
Pre-pregnancy hypertension			
No	209763 (98.8)	14141 (98.8)	37383 (98.9)
Yes	2651 (1.2)	174 (1.2)	430 (1.1)
Hospital category (for child birth)			
Private	65446 (30.8)	2739 (19.1)	2091 (5.5)
Public	146968 (69.2)	11576 (80.9)	35722 (94.5)
Mother type of patient (Healthcare)			
Private	78018 (36.7)	3204 (22.4)	2810 (7.4)
Public	134396 (63.3)	11111 (77.6)	35003 (92.6)
Parity			
1st child	90323 (42.5)	7859 (54.9)	12763 (33.8)
2nd child	77871 (36.7)	4055 (28.3)	12043 (31.8)
3rd child	30440 (14.3)	1621 (11.3)	7075 (18.7)
≥ 4th child	13780 (6.5)	780 (5.4)	5932 (15.7)
Parents' highest occupation			
Managers, administrators	99549 (46.9)	4697 (32.8)	6297 (16.7)
Para-professional, tradespersons	80009 (37.7)	6551 (45.8)	15027 (39.7)
Plant machine operators	19459 (9.2)	1868 (13.0)	8730 (23.1)

Students, pensioners	13397 (6.3)	1199 (8.4)	7759 (20.5)
Mother's birth region			
Oceania	171221 (80.6)	12894 (90.1)	35639 (94.3)
Europe	16268 (7.7)	872 (6.1)	1656 (4.4)
Africa, Asia, Americas	24925 (11.7)	549 (3.8)	518 (1.4)
Mother's ethnicity			
Caucasian	183646 (86.5)	13366 (93.4)	34221 (90.5)
Aboriginal or Torres Strait Islander	2543 (1.2)	375 (2.6)	2816 (7.4)
Asians and others	26225 (12.3)	574 (4.0)	776 (2.1)
Remoteness			
Major cities	158572 (74.7)	10083 (70.4)	24545 (64.9)
Inner regional	20054 (9.4)	1474 (10.3)	3691 (9.8)
Outer regional/remote	33788 (15.9)	2758 (19.3)	9577 (25.3)
IRSAD[b]			
Most disadvantaged (1)	50598 (23.8)	4311 (30.1)	16350 (43.2)
2nd quintile	45861 (21.6)	3283 (22.9)	9605 (25.4)
3rd quintile	36669 (17.3)	2430 (17.0)	5397 (14.3)
4th quintile	45731 (21.5)	2637 (18.4)	4517 (11.9)
Most advantaged (5)	33555 (15.8)	1654 (11.6)	1944 (5.1)
Maternal BMI[c]			
<25 kg/m^2 (Underweight & normal)	38046 (50.4)	2071 (43.4)	4644 (45.0)
25 to <30 kg/m^2 (Overweight)	20500 (27.1)	1375 (28.8)	2669 (25.9)
≥30 kg/m^2 (Obese/severely obese)	17005 (22.5)	1330 (27.8)	2997 (29.1)

[a]SD (standard deviation)
[b]IRSAD (Index of Relative Socio Economic Advantage and Disadvantage)
[c]BMI (Body Mass Index)

Table 6-5: (Supplementary Table) Complete-case analysis: Maternal smoking during pregnancy and risk of childhood type 1 diabetes (T1D)

		Unadjusted (N = 264,542) (T1D n = 442)		Adjusted[c] (N = 264,542) (T1D n = 442)	
	T1D, n	**HR[a]**	**(95%CI[b])**	**HR**	**(95%CI)**
Maternal smoking in pregnancy					
Non smoking	350	**Ref**		**Ref**	
Smoked only in first or second half of pregnancy	29	**1.07**	(0.73, 1.56)	**1.06**	(0.72, 1.55)
Consistent smoking	63	**0.82**	(0.63, 1.08)	**0.82**	(0.61, 1.09)
Maternal smoking in pregnancy					
Non smoking	350	**Ref**		**Ref**	
Smoking[d]	92	**0.89**	(0.71, 1.12)	**0.89**	(0.69, 1.13)

[a]HR (hazard ratio)

[b]CI (confidence interval)

[c]Adjusted for parent's age, parent's occupation, mother's birth region, maternal ethnicity, remoteness, IRSAD (Index of Relative Socio-economic Advantage and Disadvantage), mother type of patient, hospital category, parity, pre-pregnancy hypertension, pre-pregnancy diabetes and child's year of birth

[d]Smoking (combined consistent smoking, and smoking only in first or second half of pregnancy)

Table 6-6: (Supplementary Table) Negative control outcome complete case-analysis Association between maternal smoking during pregnancy and child not having a school card (Total n = 277,370, No school card n = 84,531)

			Unadjusted		Adjusted[c]
	No school card	**HR[a]**	**(95%CI[b])**	**HR**	**(95%CI)**
Maternal smoking in pregnancy		**Ref**		**Ref**	
Non smoking	65,983				
Smoked only in first or second half of pregnancy	5,068	**0.89**	(0.87, 0.92)	**0.88**	(0.86, 0.91)
Consistent smoking	13,480	**0.79**	(0.77, 0. 80)	**0.84**	(0.82, 0.85)
Maternal smoking in pregnancy					
Non smoking	65,983	**Ref**		**Ref**	
Smoking[d] (any smoking in pregnancy)	18,548	**0.81**	(0.80, 0.83)	**0.85**	(0.83, 0.87)

[a]HR (hazard ratio)

[b]CI (confidence interval)

[c]Adjusted for parents' age, parents' occupation, mother's birth region, maternal ethnicity, remoteness, IRSAD (Index of Relative Socio-economic Advantage and Disadvantage), mother type of patient, hospital category, parity, pre-pregnancy hypertension, and pre-pregnancy diabetes

[d]Smoking (combined consistent smoking, and smoking only in first or second half of pregnancy)

Table 6-7: (Supplementary Table) Sensitivity analyses based on imputed data
Maternal smoking during pregnancy and risk of childhood type 1 diabetes

	With BMI[1] n = 286,058		**2001-2013 births[2]** n = 240,646		**SA born[3]** n = 285,969		**Singleton[4]** n = 276,207		**Combined[5]** n = 232,271	
	HR	**(95% CI)**	**HR**	**(95% CI)**	**HR**	**(95% CI)**	**HR**	**(95% CI)**	**HR**	**(95% CI)**
Maternal smoking in pregnancy										
Non smoking	**Ref**		**Ref**		**Ref**		**Ref**		**Ref**	
Smoked only in first or second half of pregnancy	**1.07**	(0.74, 1.55)	**1.20**	(0.77, 1.89)	**1.11**	(0.77, 1.60)	**1.06**	(0.73, 1.55)	**1.21**	(0.77, 1.91)
Consistent smoking	**0.75**	(0.57, 0.99)	**0.80**	(0.57, 1.13)	**0.78**	(0.59, 1.03)	**0.75**	(0.57, 0.99)	**0.80**	(0.56, 1.14)
Maternal smoking in pregnancy										
Non smoking	**Ref**		**Ref**		**Ref**		**Ref**		**Ref**	
Smoking (any smoking)	**0.84**	(0.66, 1.07)	**0.91**	(0.68, 1.23)	**0.86**	(0.69, 1.10)	**0.83**	(0.65, 1.06)	**0.91**	(0.69, 1.23)

[1]With BMI (Model 1) : Adjusted for maternal BMI, along with parent's age, parent's occupation, mothers' birth region, maternal ethnicity, remoteness, IRSAD (Index of Relative Socio-economic Advantage and Disadvantage), mother type of patient, hospital category, parity, pre-pregnancy hypertension, pre-pregnancy diabetes and child's year of birth
[2]Restricted to children born from July-2001 to December-2013, and adjusted for all confounders as model 1
[3]Restricted to South Australian born children, and adjusted for all confounders as model 1
[4]Restricited to singleton births, and adjusted for all confounders as model 1

[5]Restricted to singleton children, born in South Australia form July-2001 to December-2013, and adjusted for all confounders as model-1

CHAPTER 7 EDUCATIONAL OUTCOMES AMONG CHILDREN WITH TYPE 1 DIABETES: WHOLE-OF-POPULATION LINKED-DATA STUDY

7.1 Preface

This Chapter contains the fourth study contributing to this doctoral thesis. (This study was under review in Paediatric Diabetes when the thesis was submitted for examination and was subsequently accepted and published in that journal during thesis examination).

In this paper, I have studied the impact of childhood type 1 diabetes on educational outcomes of South Australian children in grade 5 (~10 years of age). In this paper I used the potential outcomes theory in designing the analysis and estimated the average treatment effect of type 1 diabetes on children's educational outcomes.

This paper addresses the fourth research question planned to be studied in this doctoral project. "***What is the impact of type 1 diabetes on children's educational outcomes?"***

Highlights

- South Australian children with type 1 diabetes are not disadvantaged in terms of educational outcomes, compared to children without type 1 diabetes.

- This paper compares the educational outcomes of children with and without type 1 diabetes at age 10 when we assume there is considerable parental involvement in type 1 diabetes management.

7.2 PUBLICATION 4: Educational outcomes among children with type 1 diabetes: whole-of-population linked-data study

7.2.1 Statement of Authorship

Title of Paper	**Educational outcomes among children with type 1 diabetes: whole-of-population linked-data study**
Publication Status	☐ Published ☐ Accepted for Publication ☑ Submitted for Publication ☐ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	

Principal Author

Name of Principal Author (Candidate)	Mumtaz Begum
Contribution to the Paper	MB prepared the initial draft, reviewed and revised the manuscript, analysed the data, interpreted the findings, and helped in conceptualizing and designing the study.
Overall percentage (%)	
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.

Signature		Date	17/01/2020

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated

Name of Co-Author	Catherine Ruth Chittleborough		
Contribution to the Paper	CRC acquired the data, provided intellectual input in statistical analyses and interpretation of the findings, and critically reviewed and revised the manuscript for important intellectual content.		
Signature		Date	17/1/2020

Name of Co-Author	Rhiannon Megan Pilkington		
Contribution to the Paper	RMP acquired the data, provided intellectual input in statistical analyses and interpretation of the findings, and critically reviewed and revised the manuscript for important intellectual content.		
Signature		Date	17-01-2020

Name of Co-Author	Murthy Mittinty

Contribution to the Paper	MM helped in data analysis, interpreting the findings, and critically reviewed the manuscript.		
Signature		Date	15/01/2020

Name of Co-Author	John Lynch		
Contribution to the Paper	JL acquired the data, provided intellectual input in statistical analyses and interpretation of the findings, and critically reviewed and revised the manuscript for important intellectual content.		
Signature		Date	17/1/20

Name of Co-Author	Megan Penno		
Contribution to the Paper	MP provided intellectual input, helped in interpreting the findings, and critically reviewed the manuscript for important intellectual content.		
Signature		Date	14/01/2020

Name of Co-Author	Lisa Gaye Smithers		
Contribution to the Paper	LGS acquired the data, provided intellectual input in statistical analyses and interpretation of the findings, and critically reviewed and revised the manuscript for important intellectual content.		
Signature		Date	17 JAN 2020

What's known on This Subject?

Evidence about type 1 diabetes and educational outcomes is inconsistent. Cross-sectional and population-based studies showed poorer cognitive and educational outcomes for children with type 1 diabetes. However, two recent studies demonstrated no effect of type 1 diabetes on educational outcomes.

What This Study Adds?

Our whole-of-population study using doubly-robust analytical methods demonstrated little difference in year 5 literacy and numeracy skills of children with and without type 1 diabetes. These literacy and numeracy skills are important for progression through school, higher education and employment.

7.2.2 Abstract

Background

Challenges with type 1 diabetes (T1D) blood glucose management and illness-related school absences potentially influence children's educational outcomes. However, evidence about the impact of T1D on children's education is mixed. The objectives were to estimate the effects of T1D on children's educational outcomes, and compare time since T1D diagnosis (recent diagnosis (≤2 years) and 3-10 years long exposure) on educational outcomes.

Methods

This whole-of-population study used de-identified, administrative linked-data from the South Australian Early Childhood Data Project. T1D was identified from hospital ICD-10-AM diagnosis codes (E10, ranging E101 to E109), from 2001 to 2014. Educational outcomes were measured in grade 5 by the National Assessment Program-Literacy and Numeracy (NAPLAN, 2008-2015) for children born from 1999-2005. Analyses were conducted using augmented inverse probability of treatment weighting. Multiple imputation was used to impute missing data.

Results

Among 61,445 children born in South Australia who had undertaken NAPLAN assessments, 162 had T1D. There were negligible differences in the educational outcomes of children with and without T1D, and between recently diagnosed and those with longer exposure. For example, the mean reading score was 482.8±78.9 for children with T1D and 475.5±74.3 for other children. The average treatment effect of 6.8 (95% CI -6.3–19.9) reflected one-tenth of a standard deviation difference in the mean reading score of children with and without T1D.

Conclusion

Children with T1D performed similarly on literacy and numeracy in grade 5 (age ~10-years) compared to children without type 1 diabetes. This could be due to effective type 1 diabetes management.

Key words: Childhood, Type 1 diabetes, Educational outcomes, Augmented inverse probability weighting, Linked-data

7.2.3 Introduction

Children with Type 1 diabetes (T1D) are dependent on exogenous insulin their entire lives due to the immune-mediated destruction of their insulin-producing pancreatic beta cells. In the last few decades, advances in clinical care and intensive insulin treatment regimens for children with T1D have reduced the risk and severity of long-term complications such as vision loss, renal failure, cardiovascular disease, and T1D specific mortality.[3, 222-224] However, achieving optimum glucose control is challenging in the pediatric population with T1D, and the associated risk of hypo- or hyperglycaemia may have consequences for children's cognitive function and learning outcomes.[237, 340]

Daily glycaemic variations, severe hypo-or hyperglycaemic episodes and other T1D related challenges may impact classroom learning and academic attainments of children in many ways. Firstly, glycaemic variations during school day even without overt hypo-or-hyperglycaemia can affect cerebral glucose transport,[341] and may have implications for children's concentration and memory processes. Secondly, clinically significant severe hypoglycaemia[243, 342, 343] and hyperglycaemia or diabetic ketoacidosis (DKA),[240] can impact brain and cognitive functions during and in the aftermath of the event. Thirdly, early onset of T1D has been associated with changes in brain volumes and altered neural pathways, and may exert chronic effects on children's cognition and presumably academic performance.[344] Children with T1D have shown lower executive functioning including planning, working-memory, processing, attention, problem solving capacity, and visual motor integration.[340, 345]

In addition, compared to children without T1D, children with T1D are reported to have 3% lower school attendance per-year[13], three times more hospital days per person-year,[12] and 19% higher mental health referral rates.[14] Collectively, reduced concentration and focus in classroom during transient glycaemic excursion, decreased cognitive and brain function during overt hypo- or hyperglycaemic episodes, psychological challenges, repeated hospitalization and school absenteeism put children with T1D at risk for poor educational outcomes. Although in the last 10 years there have been advances in T1D therapy, it is unknown to what extent these improvements have helped children in South Australia to reach their full potential in learning and educational achievement.

While there are plausible biological pathways linking T1D with educational outcomes, studies have mixed results that range from poorer outcomes[35, 36, 346] to null association.[13, 37] It is possible these differing results relate to differences in recent improvements in T1D care, such as the use of insulin pumps, better diabetes education programs and training school staff for managing T1D.[347] For example, two Swedish studies used school data from 1988-2003 during a period when use of insulin pumps was lower, and reported negative effects of T1D on educational outcomes.[35, 36] Since that time, insulin pump use has increased, and has been linked with lower risks of severe hypoglycaemia and diabetic ketoacidosis.[348] However, these Swedish studies [35, 36] as well as other study from Western Australia[13] did not adjust for some important confounders including socioeconomic position,[13, 35] ethnicity,[13, 35] maternal diabetes,[13, 35, 36] mother's birth-region,[13, 35] maternal smoking,[13, 35, 36] and child's birthweight.[13, 35, 36] We used recent whole-of-population linked-data, with a wide range of

potential confounders, to compute the average treatment effect (ATE) of T1D on children's educational outcomes using augmented inverse probability weighting (AIPW). AIPW is a doubly robust procedure using which one can obtain results that have similar interpretation as the results obtained from a randomised trial in the absence of unmeasured confounding. AIPW handles observed confounding by creating a pseudo population where every child appears both as exposed (T1D) and unexposed (non- T1D).[250, 283, 284]

7.2.4 Methods

7.2.4.1 Study design and data source

This study used whole-of-population, de-identified, administrative linked data from the South Australian Early Childhood Data Project (SA ECDP).[256] Data linkage was conducted by an independent linkage agency (SA NT Datalink).[269] Children appearing in different administrative datasets were linked using a probabilistic linkage algorithm that matched children using basic demographic information, such as name, sex, address and date of birth. A very small false linkage rate (0.1-0.5%) has been reported in Australian data linkage systems.[269, 271] The data custodians provided de-identified data with a project linkage key to the researchers for analysis.

The impact of T1D on educational outcomes was estimated by linking routinely collected perinatal and births registration data from 1999 to 2005, inpatient hospitalization data from 2001 to 2014, and school assessment data of year 5 children (age ~10 years) from South Australia collected from 2008 to 2015.[259, 268]

7.2.4.2 Type 1 diabetes

Children with T1D were identified from inpatient hospitalization data from all public hospitals in South Australia, using the International Classification of Disease, 10th edition, Australian Modification (ICD-10-AM) codes (E10, ranging E101 to E109), including both primary and secondary diagnoses.[264] These diagnoses codes are assigned by trained hospital staff following discharge from each hospitalisation episode. The age at first T1D code in hospitalization data was taken as age at diagnosis.

7.2.4.3 Educational outcome

De-identified individual-level educational data for all public schools in South Australia were provided by the South Australian Department for Education. These data were collected through the National Assessment Program-Literacy and Numeracy (NAPLAN), a standardized assessment of literacy and numeracy that commenced in 2008, and is administered annually to students in year/grade 3 (age ~8 years), 5 (~10 years), 7 (~12 years)

and 9 (~14 years), in all schools in Australia.[268] NAPLAN assessment was paper-based for all students from 2008-2015. We used year 5 NAPLAN scores as outcomes because year 5 children were the oldest cohort in our dataset with sufficient numbers of children diagnosed with T1D prior to their NAPLAN assessment.

Raw NAPLAN scores are converted to scaled scores ranging from 0-1000. Results are reported using five scales, one for each domain of numeracy, reading, writing, spelling and grammar. The Australian national average scale score for year 5 children is approximately in the high 400s to 500, and the score increases with advancing grade.[268]

7.2.4.4 Confounding

Potential confounders were identified a priori based on content knowledge, previous studies and directed acyclic graphs. These confounders were sourced from the Perinatal Statistics Collection,[259] a mandatory data collection on every child born in South Australia collected at the time of birth, using a standardized tool. Parental characteristics included as confounders were parental age and parents' highest occupation (Managers, administrators and professionals; Para-professionals, tradespersons, clerks, salespersons, and personal service; Plant, machine operators, drivers and labourers; Students, pensioner, home duties and unemployed). Maternal characteristics included pre-pregnancy or gestational diabetes (yes/no), pre-pregnancy or gestational hypertension (yes/no), ethnicity (Caucasian, Aboriginal or Torres Strait Islander, Asian and others), mother's birth region (Oceania,

Europe, and Asia, Africa, America), smoking during pregnancy (yes/no), number of antenatal visits (continuous variable), whether mother was a private or public patient, and whether the child was delivered in a private or public hospital. Child's birthweight for gestational age z-score[321] was also included. Area-level socioeconomic indicator (Index of Relative Socioeconomic Advantage and Disadvantage, IRSAD[305]), and living in an accessible or remote area (measured by the Australian Remoteness Index for Areas, ARIA[322]), were based on mother's residential postcode at the time of child birth. Parents' highest education level (school only; certificate/diploma; bachelor degree or above) was obtained from school enrolment data. If parental education information was missing, then parental education of the child's biological sibling was transposed.

7.2.4.5 Statistical analysis

In the primary analysis, we estimated the average treatment effect (ATE) of T1D on the five educational domains (reading, writing, spelling, grammar and numeracy scale scores), using AIPW.[250, 349] In this case the "treatment" is being exposed to T1D. In AIPW analysis, two models were specified, and all the above-mentioned confounders were included in the outcome model (linear), and also in the treatment model (logistic). The treatment (exposure) model computes the probability of having T1D given the observed confounders, and the reciprocal of this probability is then used as the weight in the outcome regression. AIPW is known as a doubly robust model, and produces unbiased estimates if at least one of the

models (the outcome or treatment model) is correctly specified.[250] The ATE in our study is the marginal difference in NAPLAN scale scores of children with and without T1D.

For the secondary analysis, AIPW for multivalued exposures was used to estimate the ATE of recently diagnosed T1D (≤2 years since diagnosis) and those exposed to T1D for 3-10 years before the time of NAPLAN assessment, compared with no T1D.[350] Recently diagnosed children might have more difficulty in adjusting and managing T1D, which could impact their assessment. There were only 24 children diagnosed with T1D, 6-12 months prior NAPLAN assessment, therefore, we used ≤2 years since T1D as recent diagnosis.

7.2.4.6 Missing data

The amount of missing data on the five educational domains ranged from 5.8-6.3%. The missing data for confounding factors was low (ranging from 0.3-1.7%), except father's age (4.8%), number of antenatal visits (11.6%) and parents' highest education (31.1%) (Figure 7-1). Multiple imputation by chained equation was conducted to impute missing data on confounders and the outcome.[323] Twenty datasets were generated. Multiple imputation was conducted to maintain the association between childhood type 1 diabetes and educational outcomes, and to account for the potential bias if the association differs between children with and without complete data. All the variables mentioned in Table 7-1 were included in the imputation model. In addition, birth plurality, year of birth, sex, school-card (financial assistance with school fees), and year in which NAPLAN was conducted were included as

auxiliary variables to inform the imputation model, but were not part of the main analysis. We applied Rubin's rules[298] to compute the mean ATE, within imputation variance and between imputation variance, and to combine the estimates from the imputed datasets (Supplementary File 7-3).

Analyses based on imputed datasets are presented in the paper. Complete-case analyses are presented in Supplementary Table 7-4 and 7-5. All analyses were conducted in Stata SE (version 15.1).

7.2.4.7 Ethics Approval

Ethics approval for this project was granted by the Human Research Ethics Committees of the South Australian Department of Health (HREC/13/SAH/106), and the Aboriginal Health Council of South Australia (04-13-538).

7.2.5 Results

Among 61,445 children born in South Australia from 1999-2005 with year 5 NAPLAN assessment from 2008-2015, 162 had T1D (Figure 7-1).

Table 7-1 shows some differences in the sociodemographic and perinatal characteristics of children with and without T1D. For example, children with T1D were less likely to be born to mothers who smoked, and were more likely to have highly educated parents. Children with

T1D were more likely to have mothers who had hypertension or diabetes. Complete-case analysis showed similar patterns (Supplementary Table 7-4).

Table 7-2 presents the crude mean, standard deviation (SD), and the ATE (computed by AIPW) of T1D, for the five educational domains. There was negligible difference in the mean reading, writing, spelling, grammar, and numeracy scores of children with and without T1D. For example, the mean reading score±SD was 482.8±78.9 for children with T1D and 475.5±74.3 for children without T1D. The reading scale score ATE, or the difference in potential outcome means of the reading scale scores of children with and without T1D was 6.8 (95% CI -6.3–19.9). This effect represents a difference of around one-tenth of the standard deviation, and at an assessment level it is a difference of less than one mark.[351] Complete-case analysis showed similar findings (Supplementary Table 7-5).

Table 7-3 presents the associations according to the time since T1D diagnosis (≤2 years or 3-10 years) versus no T1D. There were negligible differences in the mean reading, writing, spelling, grammar and numeracy scores according to time since diagnosis. For example the mean reading score±SD was 475.5±81.5 for children without T1D, 488.2±83.0 for recently diagnosed (≤2 years since T1D), and 480.2±75.4 for 3-10 years exposure to T1D. Compared to children without diabetes, the reading scale score ATE was 6.3 (95% CI -29.8–42.3) for the recently diagnosed children (≤2 year since T1D), and 2.6 (95% CI -10.9–16.2) for 3-10 years exposure to T1D. This estimate reflects a difference of one-twelfth of the SD for recently diagnosed children and one twenty-fifth of the SD for 3-10 years exposure to T1D

compared to children without diabetes. Once again, these effects reflect a difference of less than one mark in the raw reading score.[351]

7.2.6 Discussion

In this whole-of-population study a diagnosis of T1D had little impact on educational outcomes of children completing the educational assessment (NAPLAN) in year 5. This was consistent across all NAPLAN domains of reading, writing, spelling, grammar and numeracy. Differences in mean literacy and numeracy scale scores for children with and without T1D were very small, around one tenth of a standard deviation, or equivalent to less than one mark in the raw assessment. In addition, the estimate (ATE) for recently diagnosed children (≤2 years since diagnosis) and those exposed to T1D for 3-10 years, compared to children without T1D reflected a difference equivalent to less than one mark in the raw scores of the five domains of educational outcomes in year 5.

Our findings are in contrast to small cross-sectional studies[244, 245] two large Swedish population-based studies[35, 36] and a recent record linkage study from Scotland[346] that have demonstrated differences in educational or cognitive outcomes between children with and without T1D. The small cross-sectional studies used cognitive tests that measured verbal, performance and overall intelligence quotient such as the Wechsler Intelligence Scale[244, 245] Whereas educational assessments, such as NAPLAN used in our study, measure literacy and numeracy skills required for children to progress through school, and has implications for access to higher education,[247] employment[248] and income in adulthood.[249] The two large

Swedish nationwide population-based studies that demonstrated poor educational outcomes for children with T1D at the end of compulsory schooling (<15 years) and upper secondary school level (<19 years), used data from children born during the 1970s to 1980s (1972-1978; T1D n=2,485,[36] 1973-1986; T1D n=5,159[35]) with school results from 1988 to 2003.[35] Our study used more recent educational outcomes (2008-2015) and hospital data (2001-2014), during a period of improved T1D treatment, which may help explain why we found little association between T1D and educational outcomes. Similar to our findings, no difference in educational outcomes of children with and without T1D was reported in two recent studies from Western Australia (2008-2011) and Denmark (2011-2015) that used nationally administered standardized school tests as their outcome.[13, 37] The Western Australian study[13] did not adjust for parents' socioeconomic position, ethnicity, and maternal characteristics (diabetes, birth-region, smoking), and the Danish[37] study adjusted for similar confounders as in our study.

Improvements in T1D management, diabetes education programs for school staff, and perhaps increasing use of insulin pumps due to government subsidy[233-235] in recent years may have helped children in achieving better glycaemic control.[233, 235, 347, 352] In Australia, 43% of children with T1D aged 0-14 years used a pump to administer insulin in 2013.[53] The risk of severe hypoglycaemia has shown to be lower among insulin pump users compared with injection users.[348, 353] Improvements in glycaemic control could also be attributed to the fact that T1D is a recognized medical condition by the Australian Government, Department of Social Service[354], and it is a legal obligation for schools to make reasonable adjustments to

encourage participation for children with T1D at school.[352] Thus, children with T1D may reach their optimum capabilities and achieve similar educational outcomes as children without T1D, as demonstrated in our study.

One potential reason for no discernible differences in the educational outcomes of children with and without T1D could be the younger age (9-10 years) of children in our study. Studies have shown that younger children are more likely to have better glycaemic control than older children.[355, 356] It is possible that children with severe T1D and poor glycaemic control were absent on the day of the NAPLAN assessments. Therefore, the impact of T1D on educational outcomes may have been underestimated. However, consistency in the effect estimate between complete-case and imputed analyses suggest similarity in the characteristics of children who took the test or were absent on the test day. NAPLAN assesses the learned literacy and numeracy skills, it does not capture day-to-day difficulties children have with learning and other aspects of cognitive and psychosocial development. NAPLAN may not measure higher level reasoning or problem-solving skills or the capacity to analyse complex and lengthy text where success is heavily dependent on intact working memory skills. Therefore, the impact of T1D and its related events on children's overall wellbeing and development cannot be ruled out.

The use of AIPW, an analysis that is robust to model misspecification,[250] is a strength of this study as it provides unbiased results if either the outcome model or the treatment model is correctly specified. The linkage of school data with the perinatal data enabled us to use a range of socioeconomic and perinatal characteristics to account for potential confounding, as

some of them were not adjusted for in previous studies.[13, 35, 36] Our study only included children attending public hospitals, therefore, we may have missed children who never appeared in any public hospital. However, we believe that case-ascertainment is high in our study because firstly, we have used both primary and secondary diagnoses codes, even capturing children hospitalised for other reasons, if T1D is coded as their additional diagnosis. Secondly, the one paediatric public hospital in South Australia with a specialized endocrinology unit, where children get admitted for stabilization after diabetes diagnosis is included in our study. Thirdly, T1D incidence in our study is consistent with Australian national reports.[62] Still, our study has small number of children with T1D (n=162), because not all children with T1D were diagnosed before year 5 NAPLAN assessments or were old enough to sit the NAPLAN. In addition, private schools do not release individual-level NAPLAN data, therefore, were not included in this study. Educational outcomes of children in private and public schools have shown to be similar after adjusting for socioeconomic composition.[357] It is not the school-sector, rather socioeconomic composition that confounds the association between T1D and educational outcomes, therefore having no data on private schools is unlikely to bias our estimate.

7.2.7 Conclusion

This whole-of-population study demonstrated that South Australian children with T1D are not disadvantaged on educational outcomes in year 5 compared to children without T1D.

This supports the idea that the negative effects of T1D on educational outcomes may have equalized possibly through improved T1D management in children.

7.2.8 Tables and Figures

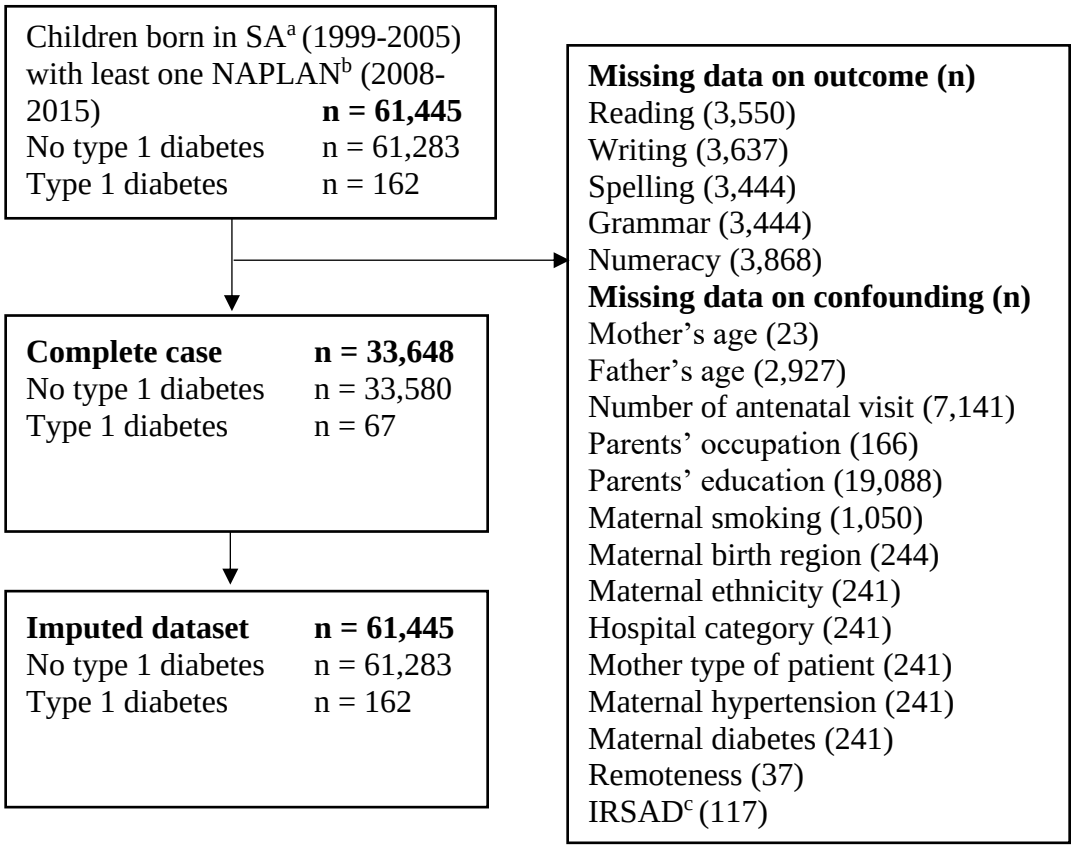


Figure 7-1: Flow chart of the study population

[a]SA (South Australia)
[b]NAPLAN (National Assessment Program-Literacy and Numeracy)
[c]IRSAD (Index of Relative Socioeconomic Advantage and Disadvanatge)

Table 7-1: Characteristics of children with and without type 1 diabetes (N = 61,445)

	No Type 1 diabetes n = 61,283		Type 1 diabetes n =162	
	n	**%**	**n**	**%**
Mother's age (years), mean (SD)[a]	28.6	(5.9)	28.6	(6.0)
Father's age (years), mean (SD)	31.9	(6.7)	31.6	(7.7)
Number of antenatal visits, mean (SD)	10.6	(3.2)	10.6	(2.8)
BWGAZ[b], mean (SD)	0.03	(1.1)	0.12	(1.0)
Maternal diabetes				
No	58,758	95.9	150	92.3
Yes	2,525	4.1	12	7.7
Hypertension				
No	55,475	90.5	138	84.9
Yes	5,808	9.5	24	15.1
Parents' highest occupation				
Managers, administrators	19,160	31.3	54	33.1
Para-Professional, tradespersons	24,858	40.6	75	46.2
Plant machine operators	9,342	15.2	16	9.8
Students, pensioners	7,923	12.9	18	10.9
Parents' highest education				
School only	21,275	34.7	55	33.8
Certificate/diploma	27,081	44.2	64	39.5
Bachelor degree or above	12,928	21.1	43	26.7
Health care of mother				
Private	13,754	22.4	45	27.7
Public	47,529	77.6	117	72.3
Hospital category				
Private	11,027	18.0	34	21.0
Public	50,256	82.0	128	79.0
Mother's birth region				
Oceania and Antarctica	54,519	89.0	143	88.4
Europe	4,149	6.8	**	**
Africa, Asia, Americas	2,615	4.3	**	**
Mother's ethnicity				
Caucasian	56,609	92.4	157	96.8
Aboriginal or Torres Strait Islander	1,934	3.2	**	**
Asian or other	2,739	4.5	**	**

Maternal smoking during pregnancy				
No	44,248	72.2	138	85.3
Yes	17,035	27.8	24	14.7
IRSAD[c]				
Most disadvantaged (1)	17,346	28.3	45	27.8
2nd quintile	13,701	22.4	34	20.8
3rd quintile	10,077	16.4	30	18.3
4th quintile	11,385	18.6	32	19.8
Most advantaged (5)	8774	14.3	21	13.3
Remoteness				
Major cities	41666	68.0	119	73.7
Inner regional	5985	9.8	9	5.8
Outer regional/remote	13631	22.2	33	20.5

[a]SD (Standard Deviation)
[b]BWGAZ (birthweight for gestational age z-score)
[c]IRSAD (Index of Relative Socioeconomic Advantage and Disadvantage)
**Cells with n<5 and adjacent cells have been retracted as per ethics requirements

Table 7-2: Crude mean (SD) NAPLAN[a] scale scores and average treatment effect (ATE) of type 1 diabetes on educational outcomes in year 5 (N = 61,445)

	Type 1 diabetes n = 162		No type 1 diabetes n = 61,283				
	Crude mean	**SD[b]**	**Crude Mean**	**SD**	**Difference in crude mean**	**ATE**	**(95% CI)[c]**
Reading	482.8	78.9	475.5	74.3	7.3	6.8	(-6.3–19.9)
Writing	458.8	75.1	455.6	74.3	3.2	1.9	(-11.0–14.8)
Spelling	478.3	73.8	472.8	74.3	5.5	3.8	(-9.5–17.0)
Grammar	481.9	87.8	475.1	99.0	6.8	0.7	(-13.5–15.1)
Numeracy	463.2	67.5	463.6	74.3	-0.4	-5.4	(-14.9–4.0)

[a]NAPLAN (National Assessment Program-Literacy and Numeracy)
[b]SD (Standard Deviation)
[c]CI (Confidence interval)

Table 7-3: Crude mean (SD) NAPLAN[d] scale scores and average treatment effect (ATE) of recent diagnosis (≤2 years since diagnosis) and 3-10 years exposure to type 1 diabetes (N = 61,445)

	No type 1 diabetes n = 61,283		**[c]Type 1 diabetes (≤2 years) n = 52**		**[d]Type 1 diabetes (3-10 years) n = 110**		**Type 1 diabetes (≤2 years) Vs no T1D**		**Type 1 diabetes (3-10 years) Vs no T1D**	
	Mean[a]	**SD[b]**	**Mean**	**SD**	**Mean**	**SD**	**ATE[e]**	**95%CI[f]**	**ATE**	**95%CI**
Reading	475.5	81.5	488.2	83.0	480.2	75.4	6.3	(-29.8–42.3)	2.6	(-10.9–16.2)
Writing	455.6	75.5	462.1	81.6	457.3	71.1	-0.2	(-33.2–32.9)	2.3	(-12.4–17.0)
Spelling	472.8	76.9	488.4	73.7	473.5	73.5	9.1	(-17.2–35.4)	-0.2	(-15.3–14.9)
Grammar	475.1	87.2	492.1	103.5	477.0	78.5	9.6	(-26.2–45.4)	-3.7	(-18.8–11.4)
Numeracy	463.6	68.7	464.3	69.3	462.7	67.9	-10.2	(-29.3–8.9)	-6.0	(-18.3–6.4)

[a]Crude mean (scale scores)
[b]SD (Standard Deviation)
[c]≤2 years since T1D diagnosis at the time of year 5 NAPLAN
[d]3-10 years since T1D diagnosis at the time of year 5 NAPLAN
[e]ATE (Average treatment effect)
[f]CI (Confidence interval)
[d]NAPLAN (National Assessment Program-Literacy and Numeracy)

Supplementary Tables

Table 7-4: (Supplementary Table) Characteristics of children with and without type 1 diabetes (complete-case analysis)

	No type 1 diabetes		**Type 1 diabetes**	
	N	**%**	**N**	**%**
Mother's age (years)	**61,277**		**145**	
Mean (SD)[a]	28.6	(5.8)	28.6	(6.3)
Father's age (years)	**58,379**		**139**	
Mean (SD)	32.0	(6.5)	31.9	(7.4)
Number of antenatal visits	**54,175**		**129**	
Mean (SD)	10.6	(3.1)	10.7	(2.6)
BWGAZ[b]	**61,059**		**145**	
Mean (SD)	-0.02	(1.0)	0.12	(1.1)
Maternal diabetes	**61,059**		**145**	
No	58,547	95.9	134	92.4
Yes	2,512	4.1	11	7.6
Hypertension	**61,059**		**145**	
No	55,271	90.5	123	84.8
Yes	5,788	9.5	22	15.2
Parents' highest occupation	**61,134**		**145**	
Managers, administrators	19,130	31.3	46	31.7
Para-Professional, tradespersons	24,806	40.6	69	47.6
Plant machine operators	9,313	15.2	14	9.7
Students, pensioners	7,885	12.9	16	11.0
Parents' highest education	**42,250**		**107**	
School only	14,400	34.1	38	**35.5**
Certificate/Diploma	18,757	44.4	40	**37.4**
Bachelor degree or above	9,093	21.5	29	27.1
Health care of mother	**61,059**		**145**	
Private	13,674	22.4	38	26.2
Public	47,385	77.6	107	73.8
Hospital category	**61,059**		**145**	
Private	10,961	18.0	29	20.0
Public	50,098	82.1	116	80.0
Mother's birth region	**61,056**		**145**	
Oceania and Antarctica	54,332	89.0	129	89.0
Europe	4,124	6.8	**	**

Africa, Asia, Americas	2,600	4.3	**	**
Mother's ethnicity	**61,059**		**145**	
Caucasian	56,405	92.4	**	**
Aboriginal or Torres Strait Islander	1,928	3.2	**	**
Asian or other	2,726	4.5	**	**
Maternal smoking during pregnancy	**60,252**		**143**	
No	43,485	72.2	122	85.3
Yes	16,767	27.8	21	14.7
IRSAD[c]	**61,183**		**145**	
Most disadvantaged (1)	17,310	28.3	41	28.3
2nd quintile	13,677	22.4	31	21.4
3rd quintile	10,060	16.4	27	18.6
4th quintile	11,370	18.6	29	20.0
Most advantaged (5)	8,766	14.3	17	11.7
Remoteness	**61,263**		**145**	
Major cities	41,653	68.0	107	73.8
Inner regional	5,983	9.8	**	**
Outer regional/remote	13,627	22.2	**	**

[a]SD (Standard Deviation)

[b]BWGAZ (birthweight for gestational age z-score)

[c]IRSAD (Index of Relative Socioeconomic Advantage and Disadvantage)

**Cells with <5 and adjacent cells have been retracted as per ethics requirements

Table 7-5: (Supplementary Table) Average treatment effect (ATE) of type 1 diabetes on NAPLAN[f] scale scores in year 5 (complete case analysis)

	Type 1 diabetes			No type 1 diabetes					
	N	**Mean[a]**	**SD[b]**	**N**	**Mean**	**SD**	**ATE[c]**	**95% CI[d]**	**N**
Reading	151	483.0	74.3	57744	476.7	78.6	2.90	(-10.67–16.48)	33,692
Writing	151	458.9	71.1	57657	457.0	73.2	-2.19	(-23.59–19.21)	33,648
Spelling	150	479.4	68.1	57851	473.8	74.0	0.33	(-16.33–17.00)	33,734
Grammar	150	483.0	82.2	57851	476.4	85.4	-3.48	(-19.37–12.40)	33,734
Numeracy	144	466.2	62.7	57433	464.7	66.5	-0.93	(-12.43–10.58)	33,495

[a]Crude mean

[b]SD (Standard Deviation)

[c]ATE (Average treatment effect)

[d]CI (Confidence interval)

[f]NAPLAN (National Assessment Program-Literacy and Numeracy)

Supplementary File 7-6. Stata Syntax for computing augmented inverse probability weight (AIPW) on imputed data for binary exposure

```
matrix TEmat=J(20,2,0)
forvalues j=1/20{
 qui mi xeq `j' : teffects aipw (Outcome-variable  confounders, linear) ///
(Exposure-variable  predictors/covariates, logit),  vce(robust)

matrix coeffv=e(b)
matrix var=e(V)
matrix TEmat [`j',1]=coeffv[1,1]
*within imputed data variance
matrix TEmat [`j',2]=var[1,1]
}

svmat TEmat
sum TEmat1
sca TEmat1bar=r(mean)
*between imputed data variance
sca TEmat1_var=(r(sd))^2
sum TEmat2
sca TEmat2bar=r(mean)
sca TV=TEmat2bar +(1+1/20)*TEmat1_var
sca lci=TEmat1bar-1.96*sqrt(TV)
sca uci=TEmat1bar+1.96*sqrt(TV)
matrix final=J(1,3,0)
matrix colnames final= TEmat1bar lci uci
matrix final[1,1]=TEmat1bar
matrix final[1,2]=lci
matrix final[1,3]=uci
matlist final
*********
```

CHAPTER 8 DISCUSSION

My doctoral thesis has brought together descriptive epidemiology on the incidence of type 1 diabetes in South Australia, with an examination of potential risk factors (caesarean birth and maternal smoking in pregnancy), and an exploration of the sequelae of type 1 diabetes in terms of children's educational outcomes in year 5. This project has utilised a wide range of information on all children in South Australia born from 1999-2013, sourced from multiple linked administrative datasets. The SA ECDP[256] used in this thesis has been recognized federally as an innovative data platform for improving children's wellbeing, with the principle of using public data for public good.[258] In this thesis, rigorous methodological approaches were deployed to address the systematic biases that are commonly seen in observational studies. To estimate the risk of bias due to unmeasured confounding negative-control outcome designs and E-value analyses were employed,[285, 288] to reduce risk of bias due to missing data multiple imputation was used,[290] and to improve causal inference doubly robust methods were used.[250, 276] By using multiple linked administrative datasets that hold different information, I was able to adjust for a wide range of potentially confounding variables. Both primary and 25 additional ICD-10-AM diagnoses codes from hospital admissions records were used to identify children with type 1 diabetes, which helped increased case ascertainment. Furthermore, being a population-wide data source, inclusion of children in the study was independent of the exposure and the outcome, thereby reducing the risk of selection bias. In this final chapter I interpret the key findings, describe the

contribution of this dissertation to the field, discuss issues of variability in the effect estimates, limitations of this body of work, and potential areas for future research.

8.1 Key findings: interpretation and contribution

Type 1 diabetes is estimated to affect ~0.3% of the Australian population aged <20 years.[41] Despite the low prevalence, it has large individual (e.g. low life expectancy,[15] psychosocial challenges[14]) and economic implications (e.g. high healthcare and hospitalization costs, poor educational outcomes).[12, 36, 358] As the incidence of type 1 diabetes has doubled in the last four decades in many countries,[7, 175, 267] a major research effort has formed to understand the aetiology, but many divergent findings have been described.[20-22, 25, 29, 30, 36, 37] The research in this thesis uses many progressive epidemiological methods to minimise possible sources of bias and better understand these divergent findings.

The first descriptive study (Chapter 4) confirmed that type 1 diabetes incidence in South Australian data varied depending on the measure of socioeconomic position used. Individual-level socioeconomic indicators demonstrated higher type 1 diabetes incidence among more advantaged children, however, there was no clear area-level socioeconomic patterning of type 1 diabetes incidence. This postulates that area and individual socioeconomic measures have different associations with type 1 diabetes.[154] The area-level measure of socioeconomic position is the principal component of multiple items including income, education,

occupation, and housing of the area. Each socioeconomic variable included in the IRSAD score has a different loading on the principal component and it is the sum of these that make up the final score. Each variable in the area-level score (IRSAD) might have a different association with type 1 diabetes. If we consider this in the context of the hygiene hypothesis, there may be a negligible association between education and type 1 diabetes when the basic level of hygiene is the same. However, housing may have a greater impact, as household crowding has been associated with lower risk of type 1 diabetes. Area-level measures of socioeconomic position[73, 95, 149] may also misclassify an individual's socioeconomic position.[154, 359] Studying socioeconomic inequalities in type 1 diabetes incidence is important for understanding why type 1 diabetes is increasing. The countries with the top ten highest published childhood type 1 diabetes incidence are developed countries, including Australia.[54, 175] This suggests a role of lifestyle changes driven by economic development for type 1 diabetes pathogenesis. One such lifestyle change could be involvement of both parents in the workforce, leading to time constraints for making healthy meals and reliance on ready-to-eat or processed foods.[360, 361] Parental employment has been linked with higher risk of childhood obesity[315, 361], which is a risk factor of type 1 diabetes[362] according to the accelerator or beta-cell stress hypothesis.[106] Some upper-middle income countries (Brazil, China) have reported low type 1 diabetes incidence.[8, 150] However, there is a dearth of type 1 diabetes data from lower-middle and low-income countries,[54] as they do not have diabetes registers, or routine record collection systems. This makes it more difficult to draw inferences about the association between socioeconomic position and type 1 diabetes incidence globally.

Socioeconomic position is a major determinant of health behaviour and healthcare choices, as socioeconomic patterning has been seen in both type 1 diabetes incidence and in its risk factors. For example, socioeconomically advantaged women are more likely to have caesarean births and less likely to smoke during pregnancy.[253]

Caesarean birth and maternal smoking in pregnancy are two potential causes of type 1 diabetes that are controversial and debated. Study 2 (Chapter 5) demonstrated negligible (5%) increased type 1 diabetes incidence for children born by caesarean, compared with normal vaginal delivery, with wide confidence intervals including the null. Contrary to the study hypothesis, intrapartum caesarean birth, where the foetus may have some exposure to vaginal microbiota was not associated with a lower type 1 diabetes incidence. This suggested that caesarean-induced changes in neonatal microbiota composition probably do not effect type 1 diabetes risk. In Australia, 35% of births were by caesarean in 2017,[253] and WHO has indicated that globally 6.2 million caesarean births are performed each year without medical indication,[308, 363] suggesting an overuse of this medical intervention. Caesarean birth has been linked with many other adverse childhood health outcomes such as asthma and obesity,[363, 364] although its long terms implications on child health have not been rigorously investigated.[363] However, the large majority of studies on caesarean have reported short term positive neonatal and child health outcomes.[363, 365, 366] On balance, the findings of previous large population-based studies,[21-23] and this thesis suggest that caesarean birth is not a risk factor for childhood type 1 diabetes.

Study 3 (Chapter 6) demonstrated that maternal smoking in pregnancy was associated with lower childhood type 1 diabetes incidence compared with unexposed children, which was also supported by the meta-analytic estimates. The negative control outcome and E-value analyses indicated the potential for some unmeasured confounding. This suggests that previous studies that did not adjust for important confounders might have overestimated the protective effect of maternal smoking in pregnancy for childhood type 1 diabetes. In this doctoral thesis, even without adjusting for father's type 1 diabetes, the observed effect of maternal smoking in pregnancy on childhood type 1 diabetes was smaller (16-24% lower risk) than previously reported estimates (25-60% reduced risk).[29, 30, 32, 33, 310] This may be related to adjustment for a more comprehensive set of confounders. Maternal smoking in pregnancy has many short and long term consequences for the child,[329, 330, 367-369] however, it has been linked with reduced risk of some diseases with an immune component, such as preeclampsia, ulcerative colitis and Parkinson's disease.[334, 336, 338, 370-372] Further studies among smokers might be helpful to understand the epigenetics or immune suppression mechanisms linking maternal smoking with childhood type 1 diabetes. In addition, smoking cessation data could be linked with the broader SA ECDP, to study whether children born to women who quit smoking had the same type 1 diabetes risk as those who continuously smoke throughout pregnancy. Furthermore, it could be used to study whether the association between smoking and type 1 diabetes is causal, or not.

One of the encouraging results from this thesis is that children with type 1 diabetes are not disadvantaged in terms of their educational outcomes compared to children without type 1

diabetes in South Australia (Chapter 7). This implies that the efforts taken in Australia for children with type 1 diabetes including insulin and insulin-pump subsidization, and involvement of schools in type 1 diabetes management, among other things, have been beneficial in helping children reach their full potential for learning.[233, 352, 354] Perhaps, due to improved type 1 diabetes management in Australia[13] and Denmark[37] children with type 1 diabetes are achieving the same educational outcomes as children without type 1 diabetes. This is not observed globally as a recent Swedish study reported a negative effect of type 1 diabetes on high school outcomes measured from 1998-2010.[246] This is surprising given that Sweden has a universal health care system but suggests that inequalities in educational outcomes for children with type 1 diabetes may remain in some settings even when there are advancements in care that could ameliorate such inequalities. The finding that there were no discernible differences in educational outcomes of children with and without type 1 diabetes are especially reassuring for patients and their families to know that children can reach their learning capability if high quality care is provided and accessible. Education outcomes are key to human capital. Studies like this could be used to advocate for type 1 diabetes care in settings where children with type 1 diabetes have poorer educational outcomes than their non-type 1 diabetes peers.

8.2 Variability in effect estimates

One of the curious observations from my thesis is that the estimates of the effect of caesarean birth (Chapter 5) and maternal smoking on childhood type 1 diabetes (Chapter 6) and educational outcomes of children with and without type 1 diabetes (Chapter 7) were smaller than previously reported findings. In the following sub-sections, detail about small effect estimates in my findings and previous studies will be discussed; including adjusting for a wide range of potential confounders, larger numbers of children with type 1 diabetes than many previous studies, and use of recent whole-of-population data instead of data in high risk populations.

8.2.1 Differences in confounding adjustment

Exchangeability is a term used in modern causal inference corresponding to comparability between the treated and untreated group,[276] and is one of the important criteria for obtaining unbiased effect estimates. However, in observational studies the exposed and unexposed are not generally exchangeable, and this lack of comparability gives rise to confounding bias.[373] Confounding is a major threat to the internal validity of the observed effect of the exposure on the outcome. In this thesis, efforts were made to make the exposed and unexposed exchangeable by conditioning on the measured confounders, in order to achieve conditional

exchangeability.[276] Analytical methods can only successfully control confounding to the extent that confounders are accurately measured (i.e. little to no error in the measurement of confounders) and are included in the analysis.[275, 373] However, as mentioned in Chapters 5, 6, and 7, many previous studies did not adjust for some important confounders making them at risk of bias due to residual confounding and potentially produced biased estimates. Adjusting for a wide range of confounders could be one reason for the smaller and closer to the null findings reported in this thesis than other studies.[25 30]

8.2.2 Small number of children with type 1 diabetes

Type 1 diabetes is a rare disease, therefore, having a sufficient number of children with type 1 diabetes who are exposed to the risk factors under investigation has always been an issue for research. Even the largest nationwide studies can have small numbers of exposed cases (for some exposures). Many population-based cohort studies conducted in at high risk populations had a relatively small number of children with type 1 diabetes exposed to caesarean birth and maternal smoking in pregnancy.[27, 29, 30] Studies with small numbers of children with type 1 diabetes (who are exposed to the risk factor under study) may have lower reproducibility, reduced statistical power, a lower chance of detecting a true effect, and are less likely to yield reliable or precise effect estimates.[291, 374, 375] The use of multiple linked population-wide administrative datasets and follow up of children in successive birth years

for 15 years, enabled this study to have a large number of children (n = 286,058). However, even having larger numbers of children compared to most previous studies, the confidence intervals in this study were wide, due to the small number of exposed cases. Although Australia is in the top ten countries for published type 1 diabetes incidence, the absolute numbers in the South Australian population are still quite small. One option to overcome small exposed samples in research might be to combine data from multiple studies in meta-analyses, as attempted in study 3 of this doctoral thesis (Chapter 7). South Australia is a relatively small state and including data from other states would have increased the number of children with type 1 diabetes and improved precision. However, data linkage across Australian jurisdictions is highly complex due to jurisdictional laws and ethics approvals required for accessing the data. Furthermore, not all jurisdictions have established data linkage systems. Thus, data linkage including multiple jurisdictions in Australia at this point is challenging. Other possibilities to have big studies with a sufficient number of children with type 1 diabetes could be global consortia where researchers from different jurisdictions in Australia and other countries could pool data (for meta-analysis) or combine analysed data. Large multicentre cohort studies focusing on at risk participants are potential areas to get sufficient numbers of children with type 1 diabetes such as ENDIA and TEDDY,[212, 333] again these studies have their own challenges in terms of recruitment and follow up which is discussed further below (8.2.3).

8.2.3 Variability in study design, data sources and follow-up time

In this section I will explain potential reasons for small estimates in my thesis compared to previous studies, given the use of different data sources and years of follow up. Population-based cohort studies conducted in high risk populations (having a first degree relative with type 1 diabetes, or HLA genotype) such as ENDIA, DAISY and TEDDY[27, 333, 376, 377] have been investigating risk factors of type 1 diabetes. These cohort studies prospectively collect data; and may be at lower risk of confounding if all potential confounders are properly measured and used. However, cohort studies are costly and can take a long time to detect the outcome, with loss to follow-up potentially leading to attrition and selection bias.[378] In addition, other cohort studies followed children for a short time, reported preclinical type 1 diabetes as their outcome, and had a very small number of children with clinical type 1 diabetes which impacted their precision.[291, 374] This depicts the challenges faced by the whole field of studying type 1 diabetes or other conditions that are rare. One way to deal with the issue of small numbers is the setup of multicentre cohort studies in at risk populations to investigate the aetiology of type 1 diabetes. Linkage of routinely collected administrative datasets of children born in successive birth years may be less costly, and is an efficient way of exploring the risk factors and outcomes of children with type 1 diabetes, and are less prone to have selection bias, power and precision problems. However, population-wide linked data can have issues of unmeasured confounding, because these data are not collected with a particular research question in mind.

Most previous studies that investigated the risk factors of type 1 diabetes were case-control studies and some of them did not adjust for individual-level socioeconomic information and important confounders.[31, 33, 310] This may explain why the case-control studies, and the meta-analyses consisting of mostly case-control studies may have high risk of confounding.[25, 31, 33, 310] In addition, some matched case-control studies did not adjust for the matching variables,[26] which may introduce bias.

8.3 Limitations and suggestions for future research

In Chapter 5, 6, 7 the limitations of each individual study have been discussed. The thesis limitations and future research directions are discussion in this section.

In study 4 (Chapter 7) I looked at the educational outcomes of children in year 5 (age ~ 10 years), but I could not look at the educational outcomes at 10-14 years (due to small numbers), during which the highest incidence of type 1 diabetes has been reported in most countries, including Sweden,[59] Norway,[66] US,[70] Northern Ireland,[71] and Australia.[57, 63, 64] In addition, older children and young adults take their own responsibility for managing blood glucose, it may only be at older ages when inequalities become evident.[355, 356, 379] Studies have shown better glycaemic control among younger children compared with adolescents.[355, 356] When the transition occurs from more parental involvement to self-management of type 1 diabetes, deteriorating adherence to treatment resulting in poor glycaemic control during

adolescence is often reported.[380] This highlights the importance of continuous surveillance and the routine collection of data to investigate long term implications of type 1 diabetes to inform policy decisions.

I will discuss disparities in accessing healthcare that could impact children's access to insulin, insulin pumps and consumables and can have implications for their type 1 diabetes management and metabolic control, which further increases the risk of complications and poorer educational outcomes. There is a lack of universal healthcare coverage in many countries, and insulin costs have tripled in the last decade,[381, 382] making it even more unaffordable for disadvantaged people, and those without health insurance.[383-385] High insulin costs have impacted access to insulin all over the world,[381, 382, 385-387] with about 2.8% of households in high-income and 63% of households in low-income countries unable to afford insulin.[386] In the US the cost of insulin is around 23 times more than in Australia.[388] The inequalities in affordability and accessibility of insulin in high, middle and low income countries could have lasting health, educational, and socioeconomic consequences for people with type 1 diabetes.[382] It was recently reported that the high insulin cost has been catastrophic for young adults in the US leading to insulin under-dosing; 25% of young adults reported underusing insulin, jeopardizing their health and survival.[379] In addition, in the US 11% people had no health insurance in 2018,[384] and HbA1c levels were higher for people without health insurance (HbA1c level 8.6%) than people with insurance (HbA1c level 7.5%).[383] This depicts a large socioeconomic disparity in glycaemic control for people with and without health insurance.[383] Such circumstances across the globe not only have

implications for maintaining glycaemic control, but can also have consequences for health, educational outcomes and even for the survival of people with type 1 diabetes. Therefore, future studies could look at the long-term health and educational outcomes of people with type 1 diabetes with and without access to health insurance, or in settings where people have to pay out-of-pocket for insulin. According to the WHO, every year 100 million people are pushed to extreme poverty because of out-of-pocket spending on health.[389]

Studies have suggested an increasing role of the environment for type 1 diabetes pathogenies.[82] Despite this, a genetic risk score has been developed to predict the risk of type 1 diabetes.[82, 86] The utility of an environmental risk score could be investigated to explore the impact of the pre-and-postnatal environmental factors on the risk of type 1 diabetes.

In my thesis paternal information on age, education, employment and occupation were available but there was no information on other paternal characteristics. In research more generally there has been a call for greater information on fathers[390] as this has been a long-overlooked contributor to children's health. For data linkage, it is unlikely that such information will be introduced to perinatal data collections in the near future because of the cost of data collection and the fact that it is unlikely to be needed by government. However, birth cohorts are making headway on collecting information from fathers, such as TEDDY.[212] Information from such studies could have been applied to my effect estimates in a quantitative bias analysis, however it was beyond the scope of the project. This might be a

way forward to get better quality evidence, to combine numbers from large linkage studies, with the rich data from cohort studies.

Next, I reflect on what I learned from my doctoral studies and consider possible next steps for research. My research on caesarean section and childhood type 1 diabetes, as well as other similar studies has shown that birth method induced variation in the neonatal microbiome is unlikely to be involved in type 1 diabetes pathogenesis. Therefore, clinical studies should focus on other areas for exploration of the potential causes of type 1 diabetes rather than the neonatal microbiome. The maternal smoking and type 1 diabetes study suggests that the prenatal environment might a have a role in type 1 diabetes pathogenesis. Epidemiological data has the limitation of not having detailed clinical information, therefore, clinical studies are needed to explore the mechanisms involved. A powerful way to make such progress in research would be for clinical and epidemiological experts to combine skills and resources as I believe there is a need for more epidemiologically rigorous research to inform clinical work, and *vice versa*. Additionally, further inroads might be made if diabetes register data was freely available for linkage with other administrative datasets following ethics approval. Currently the Australian Diabetes Register data, particularly unit-record level data, is only available to researchers after going through ethics approval and relevant data custodian agreement in all states and territories, which makes the process costly and time consuming. Even there is a cost to access the register data for research. In addition, clinical trials should routinely seek consent from participants for linkage with administrative and registry data, to be able to explore more complex questions and longer-term outcomes

by bringing together the strengths of both clinical and epidemiological data. As demonstrated in my maternal smoking and type 1 diabetes study, residual confounding is a common problem in observational studies and has been highlighted previously. The linkage of information from clinical trials and birth cohorts (when detailed clinical information is carefully observed) with administrative datasets and registries can help produce higher-quality evidence.

8.4 Conclusion

During this doctoral candidature my epidemiological expertise has matured from the use of descriptive epidemiology to causal thinking using directed acyclic graphs, gradually applying more sophisticated thinking and methods about causality, and then moving beyond conventional regression to a potential outcomes approach. This whole-of-population routinely collected administrative linked data study of children in South Australia, born from 1999-2013, demonstrated that type 1 diabetes incidence varied for individual and area-level socioeconomic position, and that the hygiene hypothesis was only supported by individual level socioeconomic patterning of type 1 diabetes incidence in South Australia. The involvement of birth method induced variation in neonatal microbiota in type 1 diabetes was not supported by the caesarean and childhood type 1 diabetes study. The negative control and E-value analyses indicated residual confounding in the estimate of maternal smoking in

pregnancy on childhood type 1 diabetes, suggesting that previously reported large protective effects were probably confounded. However, triangulation of the evidence suggested a small reduced risk of type 1 diabetes for children exposed to maternal smoking in pregnancy, highlighting the need to explore the mechanisms involved; such as the effect of smoking cessation on DNA methylation or transfer of nicotine to the foetus. My findings of similar education outcomes for children with and without type 1 diabetes are heartening and highlight the importance of improvement in type 1 diabetes management.

CHAPTER 9 REFRENCES

1. Dayan CM, Korah M, Tatovic D, Bundy BN, Herold KC. Changing the landscape for type 1 diabetes: the first step to prevention. *Lancet (London, England)*. 2019;394(10205).

2. Bluestone JA, Herold K, Eisenbarth G. Genetics, pathogenesis and clinical interventions in type[thinsp]1 diabetes. *Nature*. 2010;464(7293):1293-1300.

3. Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. *Lancet*. 2014;383(9911):69-82.

4. Bate KL, Jerums G. Preventing complications of diabetes. Med J Aust 2003; 179 (9): 498-503

5. Misra S, Oliver NS. Diabetic ketoacidosis in adults. *BMJ* 2015;351:h5660.

6. DIAMOND Project Group (2006) Incidence and trends of childhood type 1 diabetes worldwide 1990-1999. *Diabet Med*.23(8):857-866.

7. Patterson CC, Harjutsalo V, Rosenbauer J, Neu A, Cinek O, Skrivarhaug T, et al. Trends and cyclical variation in the incidence of childhood type 1 diabetes in 26 European centres in the 25 year period 1989-2013: a multicentre prospective registration study. *Diabetologia*. 2019;62(3):408-417.

8. Weng J, Zhou Z, Guo L, Zhu D, Ji L, Luo X, et al. Incidence of type 1 diabetes in China, 2010-13: population based study. *BMJ*. 2018;360.

9. Harjutsalo V, Sund R, Knip M, Groop P-H. Incidence of Type 1 Diabetes in Finland. Letters. *JAMA*. 2013;310(4):427-428.

10. Podar T, Solntsev A, Karvonen M, Padaiga Z, Brigis G, Urbonaite B, et al. Increasing incidence of childhood-onset type I diabetes in 3 Baltic countries and Finland 1983-1998. *Diabetologia*. 2001;44 Suppl 3:B17-20.

11. Redondo MJ, Yu L, Hawa M, Mackenzie T, Pyke DA, Eisenbarth GS, et al. Heterogeneity of type I diabetes: analysis of monozygotic twins in Great Britain and the United States. *Diabetologia*. 2001;44(3):354-362.

12. Icks A, Rosenbauer J, Holl RW, Grabert M, Rathmann W, Giani G. Hospitalization Among Diabetic Children and Adolescents and the General Population in Germany. *Diabetes Care*. 2001;24(3):435-440.

13. Cooper MN, McNamara KA, de Klerk NH, Davis EA, Jones TW. School performance in children with type 1 diabetes: a contemporary population-based study. *Pediatr Diabetes*. 2016;17(2):101-111.

14. Northam EA, Lin A, Finch S, Werther GA, Cameron FJ. Psychosocial Well-Being and Functional Outcomes in Youth With Type 1 Diabetes 12 years After Disease Onset. *Diabetes Care*. 2010;33(7):1430-1437.

15. Huo L, Harding JL, Peeters A, Shaw JE, Magliano DJ. Life expectancy of type 1 diabetic patients during 1997–2010: a national Australian registry-based cohort study. *Diabetologia*. 2016;59(6):1177-1185.

16. BAKER IDI Heart and Diabetes Institute. Diabetes silent pandemic and its impact on Australia. BAKER IDI Heart and Diabetes Institute, 2012.

17. Colagiuri S. *DiabCo$t Australia : Type 1: assessing the burden of type 1 diabetes in Australia*. A.C.T: Diabetes Australia; 2009.

18. Betrán AP, Ye J, Moller A-B, Zhang J, Gülmezoglu AM, Torloni MR. The Increasing Trend in Caesarean Section Rates: Global, Regional and National Estimates: 1990-2014. *PLoS ONE*. 2016;11(2):e0148343.

19. Reitan T, Callinan S. Changes in smoking rates among pregnant women and the general female population in Australia, Finland, Norway, and Sweden. *Nicotine Tob Res*. 2017;19(3):282-289.

20. Khashan AS, Kenny LC, Lundholm C, Kearney PM, Gong T, Almqvist C. Mode of obstetrical delivery and type 1 diabetes: a sibling design study. *Pediatrics*. 2014;134(3):e806-813.

21. Clausen TD, Bergholt T, Eriksson F, Rasmussen S, Keiding N, Løkkegaard EC. Prelabor Cesarean Section and Risk of Childhood Type 1 Diabetes: A Nationwide Register-based Cohort Study. *Epidemiology (Cambridge, Mass)*. 2016;27(4):547.

22. Stene LC, Magnus P, Lie RT, Sovik O, Joner G. No association between preeclampsia or cesarean section and incidence of type 1 diabetes among children: a large, population-based cohort study. *Pediatr Res*. 2003;54(4):487-490.

23. Samuelsson U, Lindell N, Bladh M, Åkesson K, Carlsson A, Josefsson A. Caesarean section per se does not increase the risk of offspring developing type 1 diabetes: a Swedish population-based study. *Diabetologia*. 2015;58(11):2517-2524.

24. Waernbaum I, Dahlquist G, Lind T. Perinatal risk factors for type 1 diabetes revisited: a population-based register study. *Diabetologia*. 2019;62(7):1173-1184.

25. Cardwell CR, Stene LC, Joner G, Cinek O, Svensson J, Goldacre MJ, et al. Caesarean section is associated with an increased risk of childhood-onset type 1 diabetes mellitus: a meta-analysis of observational studies. *Diabetologia*. 2008;51(5):726-735.

26. Mattsson K, Jonsson I, Malmqvist E, Larsson HE, Rylander L. Maternal smoking during pregnancy and offspring type 1 diabetes mellitus risk: accounting for HLA haplotype. *Eur J Epidemiol*. 2015;30(3):231-238.

27. Frederiksen B, Kroehl M, Lamb MM, Seifert J, Barriga K, Eisenbarth GS, et al. Infant exposures and development of type 1 diabetes mellitus: The Diabetes Autoimmunity Study in the Young (DAISY). *JAMA Pediatr*. 2013;167(9):808-815.

28. Svensson J, Carstensen B, Mortensen HB, Borch-Johnsen K. Early childhood risk factors associated with type 1 diabetes--is gender important? *Eur J Epidemiol*. 2005;20(5):429-434.

29. Magnus MC, Tapia G, Olsen SF, Granstrom C, Mårild K, Ueland PM, et al. Parental smoking and risk of childhood-onset type 1 diabetes. *Epidemiology 29(6):848–856*. 2018.

30. Haynes A, Cooper MN, Bower C, Jones TW, Davis EA. Maternal smoking during pregnancy and the risk of childhood type 1 diabetes in Western Australia. *Diabetologia*. 2014;57(3):469-472.

31. Robertson L, Harrild K. Maternal and neonatal risk factors for childhood type 1 diabetes: a matched case-control study. *BMC Public Health*. 2010;10(1):281.

32. D'Angeli MA, Merzon E, Valbuena LF, Tirschwell D, Paris CA, Mueller BA. Environmental factors associated with childhood-onset type 1 diabetes mellitus: an exploration of the hygiene and overload hypotheses. *Arch Pediatr Adolesc Med*. 2010;164(8):732-738.

33. Marshall AL, Chetwynd A, Morris JA, Placzek M, Smith C, Olabi A, et al. Type 1 diabetes mellitus in childhood: a matched case control study in Lancashire and Cumbria, UK. *Diabet Med*. 2004;21(9):1035-1040.

34. Ievins R, Roberts SE, Goldacre MJ. Perinatal factors associated with subsequent diabetes mellitus in the child: record linkage study. *Diabet Med*. 2007;24(6):664-670.

35. Dahlquist G, Kallen B, Swedish Childhood Diabetes Study G. School performance in children with type 1 diabetes--a population-based register study. *Diabetologia*. 2007;50(5):957-964.

36. Persson S, Dahlquist G, Gerdtham UG, Steen Carlsson K. Impact of childhood-onset type 1 diabetes on schooling: a population-based register study. *Diabetologia*. 2013;56(6):1254-1262.

37. Skipper N, Gaulke A, Sildorf SM, Eriksen TM, Nielsen NF, Svensson J. Association of Type 1 Diabetes With Standardized Test Scores of Danish School children. *JAMA*. 2019;321(5):484-492.

38. American Diabetes Association. Standards of Medical Care in Diabetes—2019. Abridged for Primary Care Providers. *Diabetes Care*. 2019;42(Supplement 1):S13-S28.

39. Regnell S, Lernmark Å. Early prediction of autoimmune (type 1) diabetes. *Diabetologia*. 2017;60(8):1370-1381.

40. Ziegler AG, Rewers M, Simell O, Simell T, Lempainen J, Steck A, et al. Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. *JAMA*. 2013;309(23):2473-2479.

41. Couper JJ, Haller MJ, Greenbaum CJ, Ziegler A-G, Wherrett DK, Knip M, et al. ISPAD Clinical Practice Consensus Guidelines 2018: Stages of type 1 diabetes in children and adolescents. *Pediatric Diabetes*. 2018;19(S27):20-27.

42. Insel RA, Dunne JL, Atkinson MA, Chiang JL, Dabelea D, Gottlieb PA, et al. Staging presymptomatic type 1 diabetes: a scientific statement of JDRF, the Endocrine Society, and the American Diabetes Association. *Diabetes Care*. 2015;38(10):1964-1974.

43. Ziegler AG, Bonifacio E, the B-BSG. Age-related islet autoantibody incidence in offspring of patients with type 1 diabetes. *Diabetologia*. 2012;55(7):1937-1943.

44. Ilonen J, Hammais A, Laine AP, Lempainen J, Vaarala O, Veijola R, et al. Patterns of beta-cell autoantibody appearance and genetic associations during the first years of life. *Diabetes*. 2013;62(10):3636-3640.

45. Krischer J, Lynch K, Schatz D, Ilonen J, Lernmark Å, Hagopian W, et al. The 6 year incidence of diabetes-associated autoantibodies in genetically at-risk children: the TEDDY study. *Diabetologia*. 2015;58(5):980-987.

46. Ziegler AG, Pflueger M, Winkler C, Achenbach P, Akolkar B, Krischer JP, et al. Accelerated progression from islet autoimmunity to diabetes is causing the escalating incidence of type 1 diabetes in young children. *J Autoimmun*. 2011;37(1):3-7.

47. Mayer-Davis EJ, Kahkoska AR, Jefferies C, Dabelea D, Balde N, Gong CX, Aschner P and Craig M. Definition, epidemiology and classification of diabetes in children and adolescents. ISPAD Clinical Practice Consensus Guidelines 2018.

48. Thomas NJ, Jones SE, Weedon MN, Shields BM, Oram RA, Hattersley AT. Frequency and phenotype of type 1 diabetes in the first six decades of life: a cross-sectional, genetically stratified survival analysis from UK Biobank.(Author abstract). *Lancet Diabetes Endocrinol*. 2018;6(2):122.

49. Weets I, De Leeuw IH, Du Caju MVL, Rooman R, Keymeulen B, Mathieu C, et al. The incidence of type 1 diabetes in the age group 0–39 years has not increased in Antwerp (Belgium) between 1989 and 2000: evidence for earlier disease manifestation. *Diabetes Care*. 2002;25(5):840-846.

50. Maahs DM, West NA, Lawrence JM, Mayer-Davis EJ. Epidemiology of type 1 diabetes. *Endocrinol Metab Clin North Am*. 2010;39(3):481-497.

51. Australian Institute of Health and Welfare 2019. Incidence of insulin-treated diabetes in Australia. Cat. no. CDK 11. Canberra: AIHW, Viewed 29 October, 2019.

52. International Diabetes Federation. IDF Diabetes ATLAS. Eighth edition. Belgium: International Diabetes Federation, 2017.

53. Australian Institute of Health and Welfare. Prevalence of type 1 diabetes among children aged 0–14 in Australia 2013. Canberra: Australian Institute of Health and Welfare, 2015.

54. Patterson C, Guariguata L, Dahlquist G, Soltesz G, Ogle G, Silink M. Diabetes in the young - a global view and worldwide estimates of numbers of children with type 1 diabetes. *Diabetes Res Clin Pract*. 2014;103(2):161-175.

55. Gale EAM. The Rise of Childhood Type 1 Diabetes in the 20th Century. *Diabetes*. 2002;51(12):3353-3361.

56. Skrivarhaug T, Stene LC, Drivvoll AK, Strom H, Joner G. Incidence of type 1 diabetes in Norway among children aged 0-14 years between 1989 and 2012: has the incidence stopped rising? Results from the Norwegian Childhood Diabetes Registry. *Diabetologia*. 2014;57(1):57-62.

57. Haynes A, Bulsara MK, Bower C, Jones TW, Davis EA. Regular peaks and troughs in the Australian incidence of childhood type 1 diabetes mellitus (2000-2011). *Diabetologia*. 2015;58(11):2513-2516.

58. Harjutsalo V, Sjöberg L, Tuomilehto J. Time trends in the incidence of type 1 diabetes in Finnish children: a cohort study. *The Lancet*. 2008;371(9626):1777-1782.

59. Berhan Y, Waernbaum I, Lind T, Möllsten A, Dahlquist G. Thirty Years of Prospective Nationwide Incidence of Childhood Type 1 Diabetes. The Accelerating Increase by Time Tends to Level Off in Sweden. *Diabetes*. 2011;60(2):577-581.

60. Sutton DL, Lyle DM, Pierce JP. Incidence and prevalence of insulin-dependent diabetes mellitus in the zero- to 19-years' age-group in Sydney. *Med J Aust*. 1989;151(3):140-141, 144-146.

61. Kelly HA, Byrne GC. Incidence of IDDM in Western Australia in children aged 0-14 yr from 1985 to 1989. *Diabetes Care*. 1992;15(4):515-517.

62. Australian Institute of Health and Welfare. Incidence of type 1 diabetes in Australia 2000–2013. Canberra: Australian Institute of Health and Welfare, 2015.

63. Haynes A, Bulsara MK, Jones TW, Davis EA. Incidence of childhood onset type 1 diabetes in Western Australia from 1985 to 2016: Evidence for a plateau. *Pediatr Diabetes*. 2018;19(4):690-692.

64. Taplin CE, Craig ME, Lloyd M, Taylor C, Crock P, Silink M, et al. The rising incidence of childhood type 1 diabetes in New South Wales, 1990-2002. *Med J Aust*. 2005;183(5):243-246.

65. Australian Institute of Health and Welfare 2018. Diabetes snapshot. Canberra: AIHW. Viewed 08 August 2019. https://www.aihw.gov.au/reports/diabetes/diabetes-snapshot.

66. Dzidzonu DK, Skrivarhaug T, Joner G, Moger TA. Ethnic differences in the incidence of type 1 diabetes in Norway: a register-based study using data from the period 2002-2009. *Pediatr Diabetes*. 2016;17(5):337-341.

67. Ji J, Hemminki K, Sundquist J, Sundquist K. Ethnic differences in incidence of type 1 diabetes among second-generation immigrants and adoptees from abroad. *J Clin Endocrinol Metab*. 2010;95(2):847.

68. Bruno G, Maule M, Merletti F, Novelli G, Falorni A, Iannilli A, et al. Age-period-cohort analysis of 1990-2003 incidence time trends of childhood diabetes in Italy: the RIDI study. *Diabetes*. 2010;59(9):2281-2287.

69. Jarosz-Chobot P, Polanska J, Szadkowska A, Kretowski A, Bandurska-Stankiewicz E, Ciechanowska M, et al. Rapid increase in the incidence of type 1 diabetes in Polish children from 1989 to 2004, and predictions for 2010 to 2025. *Diabetologia*. 2011;54(3):508-515.

70. Rogers MAM, Kim C, Banerjee T, Lee JM. Fluctuations in the incidence of type 1 diabetes in the United States from 2001 to 2015: a longitudinal study. *BMC Medicine*. 2017;15(1):199.

71. Cardwell CR, Carson DJ, Patterson CC. Secular trends, disease maps and ecological analyses of the incidence of childhood onset Type 1 diabetes in Northern Ireland, 1989-2003. *Diabet Med*. 2007;24(3):289-295.

72. Dahlquist GG, Nyström L, Patterson CC, Swedish Childhood Diabetes Study G, Diabetes Incidence in Sweden Study G. Incidence of type 1 diabetes in Sweden among individuals aged 0-34 years, 1983-2007: an analysis of time trends. *Diabetes care*. 2011;34(8):1754-1759.

73. Haynes A, Bulsara MK, Bower C, Codde JP, Jones TW, Davis EA. Independent effects of socioeconomic status and place of residence on the incidence of childhood type 1 diabetes in Western Australia. *Pediatr Diabetes*. 2006;7(2):94-100.

74. Rytkönen M, Moltchanova E, Ranta J, Taskinen O, Tuomilehto J, Karvonen M. The incidence of type 1 diabetes among children in Finland—rural–urban difference. *Health & Place*. 2003;9(4):315-325.

75. Ram R, Morahan G. Effects of type 1 diabetes risk alleles on immune cell gene expression. *Genes* 2017;8(6):167.

76. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet*. 2007;39(7):857-864.

77. Steck AK, Rewers MJ. Genetics of type 1 diabetes. *Clin Chem*. 2011;57(2):176-185.

78. Noble JA, Valdes AM, Cook M, Klitz W, Thomson G, Erlich HA. The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. *Am J Hum Genet*. 1996;59(5):1134-1148.

79. Pociot F, Akolkar B, Concannon P, Erlich HA, Julier C, Morahan G, et al. Genetics of type 1 diabetes: what's next? *Diabetes*. 2010;59(7):1561-1571.

80. Noble JA, Erlich HA. Genetics of type 1 diabetes. *Cold Spring Harb Perspect Med*. 2012;2(1):a007732.

81. Noble JA, Valdes AM. Genetics of the HLA region in the prediction of type 1 diabetes. *Curr Diab Rep*. 2011;11(6):533-542.

82. Craig ME, Kim KW, Isaacs SR, Penno MA, Hamilton-Williams EE, Couper JJ, et al. Early-life factors contributing to type 1 diabetes. *Diabetologia*. 2019;62(10):1823-1834.

83. Nguyen C, Varney MD, Harrison LC, Morahan G. Definition of High-Risk Type 1 Diabetes HLA-DR and HLA-DQ Types Using Only Three Single Nucleotide Polymorphisms. *Diabetes*. 2013;62(6):2135-2140.

84. Steck AK, Bugawan TL, Valdes AM, Emery LM, Blair A, Norris JM, et al. Association of Non-HLA Genes With Type 1 Diabetes Autoimmunity. *Diabetes*. 2005;54(8):2482-2486.

85. Ilonen J, Sjoroos M, Knip M, Veijola R, Simell O, Akerblom HK, et al. Estimation of genetic risk for type 1 diabetes. *Am J Med Genet*. 2002;115(1):30-36.

86. Bonifacio E, Beyerlein A, Hippich M, Winkler C, Vehik K, Weedon MN, et al. Genetic scores to stratify risk of developing multiple islet autoantibodies and type 1 diabetes: A prospective study in children. *PLoS Med*. 2018;15(4):e1002548.

87. Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, et al. DNA methylation in newborns and maternal smoking in pregnancy: Genome-wide consortium meta-analysis. Am J Hum Genet. 2016;98(4):680-696.

88. Bonifacio E, Hummel M, Walter M, Schmid S, Ziegler AG. IDDM1 and multiple family history of type 1 diabetes combine to identify neonates at high risk for type 1 diabetes. *Diabetes Care*. 2004;27(11):2695-2700.

89. Steck AK, Barriga KJ, Emery LM, Fiallo-Scharer RV, Gottlieb PA, Rewers MJ. Secondary Attack Rate of Type 1 Diabetes in Colorado Families. *Diabetes Care*. 2005;28(2):296-300.

90. Parkkola A, Harkonen T, Ryhanen SJ, Ilonen J, Knip M. Extended family history of type 1 diabetes and phenotype and genotype of newly diagnosed children. *Diabetes Care*. 2013;36(2):348-354.

91. Steck AK, Armstrong TK, Babu SR, Eisenbarth GS. Stepwise or linear decrease in penetrance of type 1 diabetes with lower-risk HLA genotypes over the past 40 years. *Diabetes*. 2011;60(3):1045-1049.

92. Bach J, Chatenoud L. The Hygiene Hypothesis: An Explanation for the Increased Frequency of Insulin-Dependent Diabetes. *Cold Spring Harb Perspect Med*. 2012;2(2).

93. Gale EA. A missing link in the hygiene hypothesis? *Diabetologia*. 2002;45:588-594.

94. Knip M, Luopajärvi K, Härkönen T. Early life origin of type 1 diabetes. *Semin Immunopathol*. 2017;39(6):653-667.

95. Patterson CC, Carson DJ, Hadden DR. Epidemiology of childhood IDDM in Northern Ireland 1989–1994: low incidence in areas with highest population density and most household crowding. *Diabetologia*. 1996;39(9):1063-1069.

96. Liese AD, Puett RC, Lamichhane AP, Nichols MD, Dabelea D, Lawson AB, et al. Neighborhood level risk factors for type 1 diabetes in youth: the SEARCH case-control study. *Int J Health Geogr*. 2012;11(1):1.

97. Staines A, Bodansky HJ, Alexander FE. Small area variation in the incidence of childhood insulin dependent diabetes mellitus in Yorkshire, UK: Links with overcrowding and population density *Int J Epidemiol*.Dec;26(6):1307-13.

98. Crow YJ, Alberti KG, Parkin JM. Insulin dependent diabetes in childhood and material deprivation in northern England, 1977-86. *BMJ*. 1991;303(6795):158-160.

99. Baumer JH, Hunt LP, Shield JP. Social disadvantage, family composition, and diabetes mellitus: prevalence and outcome. *Arch Dis Child*. 1998;79(5):427-430.

100. Joner G, Sovik O. Increasing incidence of diabetes mellitus in Norwegian children 0-14 years of age 1973-1982. *Diabetologia*. 1989;32(2):79-83.

101. Green A, Andersen PK, Svendsen AJ, Mortensen K. Increasing incidence of early onset Type 1 (insulin-dependent) diabetes mellitus: a study of Danish male birth cohorts. *Diabetologia*. 1992;35(2):178-182.

102. Songini M, Loche M, Muntoni S, Stabilini M, Coppola A, Dessi G, et al. Increasing prevalence of juvenile onset Type 1 (insulin-dependent) diabetes mellitus in Sardinia: the military service approach. *Diabetologia*. 1993;36(6):547-552.

103. Mackenbach JP. The epidemiologic transition theory. *J Epidemiol Community Health*. 1994;48(4):329-331.

104. Omran AR. The epidemiologic transition; a theory of the epidemiology of population change. Milbank Memorial Fund Quarterly 1971;49:509-38.

105. Jarosz-Chobot P, Polanska J, Polanski A. Does social-economical transformation influence the incidence of type 1 diabetes mellitus? A Polish example. *Pediatr Diabetes*. 2008;9(3 Pt 1):202-207.

106. Rewers M, Ludvigsson J. Environmental risk factors for type 1 diabetes. *Lancet*. 2016;387(10035):2340-2348.

107. Strachan DP. Hay fever, hygiene, and household size. *BMJ*. 1989;299(6710):1259-1260.

108. Kolb H, Elliott RB. Increasing incidence of IDDM a consequence of improved hygiene? *Diabetologia*. 1994;37(7):729-729.

109. Viskari H, Ludvigsson J, Uibo R, Salur L, Marciulionyte D, Hermann R, et al. Relationship between the incidence of type 1 diabetes and maternal enterovirus antibodies: time trends and geographical variation. *Diabetologia*. 2005;48(7):1280-1287.

110. Viskari H, Knip M, Tauriainen S, Huhtala H, Veijola R, Ilonen J, et al. Maternal enterovirus infection as a risk factor for type 1 diabetes in the exposed offspring. *Diabetes Care*. 2012;35(6):1328-1332.

111. Knip M, Simell O. Environmental triggers of type 1 diabetes. *Cold Spring Harb Perspect Med*. 2012;2(7):a007690.

112. Lynch KF, Lee H-S, Törn C, Vehik K, Krischer JP, Larsson HE, et al. Gestational respiratory infections interacting with offspring HLA and CTLA-4 modifies incident β-cell autoantibodies. *J Autoimmun*. 2018;86:93-103.

113. Beyerlein A, Donnachie E, Jergens S, Ziegler A-G. Infections in Early Life and Development of Type 1 Diabetes. Letters. *JAMA*. 2016;315(17):1899-1901.

114. Cardwell CR, Carson DJ, Patterson CC. No association between routinely recorded infections in early life and subsequent risk of childhood-onset Type 1 diabetes: a

matched case-control study using the UK General Practice Research Database. *Diabet Med*. 2008;25(3):261-267.

115. Ponsonby A-L, Pezic A, Cochrane J, Cameron FJ, Pascoe M, Kemp A, et al. Infant anthropometry, early life infection, and subsequent risk of type 1 diabetes mellitus: a prospective birth cohort study. *Pediatric Diabetes*. 2011;12(4pt1):313-321.

116. Sioofy-Khojine A-B, Lehtonen J, Nurminen N, Laitinen OH, Oikarinen S, Huhtala H, et al. Coxsackievirus B1 infections are associated with the initiation of insulin-driven autoimmunity that progresses to type 1 diabetes. *Diabetologia*. 2018;61(5):1193-1202.

117. Viner RM, Segal TY, Lichtarowicz-Krynska E, Hindmarsh P. Prevalence of the insulin resistance syndrome in obesity. *Arch Dis Child*. 2005;90(1):10-14.

118. Campos C. Chronic Hyperglycemia and Glucose Toxicity: Pathology and Clinical Sequelae. *Postgrad Med*. 2012;124(6):90-97.

119. Wilkin TJ. The accelerator hypothesis: weight gain as the missing link between Type I and Type II diabetes. *Diabetologia*. 2001;44(7):914-922.

120. Beyerlein A, Thiering E, Pflueger M, Bidlingmaier M, Stock J, Knopff A, et al. Early infant growth is associated with the risk of islet autoimmunity in genetically susceptible children. *Pediatr Diabetes*. 2014;15(7):534-542.

121. Magnus MC, Olsen SF, Granstrom C, Joner G, Skrivarhaug T, Svensson J, et al. Infant Growth and Risk of Childhood-Onset Type 1 Diabetes in Children From 2 Scandinavian Birth Cohorts. *JAMA Pediatr*. 2015;169(12):e153759.

122. Couper JJ, Beresford S, Hirte C, Baghurst PA, Pollard A, Tait BD, et al. Weight gain in early life predicts risk of islet autoimmunity in children with a first-degree relative with type 1 diabetes. *Diabetes Care*. 2009;32(1):94-99.

123. Hyppönen E, Virtanen SM, Kenward MG, Knip M, Akerblom HK. Obesity, increased linear growth, and risk of type 1 diabetes in children. *Diabetes Care*. 2000;23(12):1755-1760.

124. Goldacre RR. Associations between birthweight, gestational age at birth and subsequent type 1 diabetes in children under 12: a retrospective cohort study in England, 1998–2012. *Diabetologia*. 2017.

125. Stene LC, Magnus P, Lie RT, Sovik O, Joner G. Birth weight and childhood onset type 1 diabetes: population based cohort study. *BMJ*. 2001;322(7291):889-892.

126. Khashan AS, Kenny LC, Lundholm C, Kearney PM, Gong T, McNamee R, et al. Gestational Age and Birth Weight and the Risk of Childhood Type 1 Diabetes: A Population-Based Cohort and Sibling Design Study. *Diabetes Care*. 2015;38(12):2308-2315.

127. Harder T, Roepke K, Diller N, Stechling Y, Dudenhausen JW, Plagemann A. Birth weight, early weight gain, and subsequent risk of type 1 diabetes: systematic review and meta-analysis. *Am J Epidemiol*. 2009;169(12):1428-1436.

128. Cardwell CR, Stene LC, Joner G, Davis EA, Cinek O, Rosenbauer J, et al. Birthweight and the risk of childhood-onset type 1 diabetes: a meta-analysis of observational studies using individual patient data. *Diabetologia*. 2010;53(4):641-651.

129. Atkinson MA, Chervonsky A. Does the gut microbiota have a role in type 1 diabetes? Early evidence from humans and animal models of the disease. *Diabetologia*. 2012;55(11):2868-2877.

130. Knip M, Siljander H. The role of the intestinal microbiota in type 1 diabetes mellitus. *Nat Rev Endocrinol*. 2016;12(3):154-167.

131. Siljander H, Honkanen J, Knip M. Microbiome and type 1 diabetes. *EBioMedicine*. 2019;46:512-521.

132. Han H, Li Y, Fang J, Liu G, Yin J, Li T, et al. Gut Microbiota and Type 1 Diabetes. *Int J Mol Sci*. 2018;19(4):995.

133. Tanaka M, Nakayama J. Development of the gut microbiota in infancy and its impact on health in later life. *Allergol Int*. 2017;66(4):515-522.

134. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A*. 2010;107(26):11971-11975.

135. Langdon A, Crook N, Dantas G. The effects of antibiotics on the microbiome throughout development and alternative approaches for therapeutic modulation. *Genome Med*. 2016;8(1):39-39.

136. Francino MP. Antibiotics and the Human Gut Microbiome: Dysbioses and Accumulation of Resistances. *Front Microbiol*. 2016;6(1543).

137. Matsuyama M, Gomez-Arango LF, Fukuma NM, Morrison M, Davies PSW, Hill RJ. Breastfeeding: a key modulator of gut microbiota characteristics in late infancy. *J Dev Orig Health Dis*. 2019;10(2):206-213.

138. Backhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host Microbe*. 2015;17(5):690-703.

139. Azad MB, Konya T, Maughan H, Guttman DS, Field CJ, Chari RS. Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. *Can Med Assoc J*. 2013;185.

140. Rooks MG, Garrett WS. Gut microbiota, metabolites and host immunity. *Nat Rev Immunol*. 2016;16(6):341-352.

141. Belkaid Y, Hand Timothy W. Role of the Microbiota in Immunity and Inflammation. *Cell*. 2014;157(1):121-141.

142. Murri M, Leiva I, Gomez-Zumaquero JM, Tinahones FJ, Cardona F, Soriguer F, et al. Gut microbiota in children with type 1 diabetes differs from that in healthy children: a case-control study. *BMC Medicine*. 2013;11:46-46.

143. Stewart CJ, Ajami NJ, O'Brien JL, Hutchinson DS, Smith DP, Wong MC, et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature*. 2018;562(7728):583-588.

144. Giongo A, Gano KA, Crabb DB, Mukherjee N, Novelo LL, Casella G, et al. Toward defining the autoimmune microbiome for type 1 diabetes. *ISME J*. 2011;5(1):82-91.

145. Stene LC, Gale EAM. The prenatal environment and type 1 diabetes. *Diabetologia*. 2013;56(9):1888-1897.

146. Kondrashova A, Reunanen A, Romanov A, Karvonen A, Viskari H, Vesikari T, et al. A six-fold gradient in the incidence of type 1 diabetes at the eastern border of Finland. *Ann Med*. 2005;37(1):67-72.

147. Soderstrom U, Aman J, Hjern A. Being born in Sweden increases the risk for type 1 diabetes - a study of migration of children to Sweden as a natural experiment. *Acta Paediatr*. 2012;101(1):73-77.

148. Oilinki T, Otonkoski T, Ilonen J, Knip M, Miettinen PJ. Prevalence and characteristics of diabetes among Somali children and adolescents living in Helsinki, Finland. *Pediatr Diabetes*. 2012;13(2):176-180.

149. Patterson CC, Dahlquist G, Soltesz G, Green A. Is childhood-onset type I diabetes a wealth-related disease? An ecological analysis of European incidence rates. *Diabetologia*. 2001;44 Suppl 3:B9-16.

150. Negrato CA, Lauris JRP, Saggioro IB, Corradini MCM, Borges PR, Crês MC, et al. Increasing incidence of type 1 diabetes between 1986 and 2015 in Bauru, Brazil. *Diabetes Res Clin Pract*. 2017;127:198-204.

151. Feltbower RG, Manda SO, Gilthorpe MS, Greaves MF, Parslow RC, Kinsey SE, et al. Detecting small-area similarities in the epidemiology of childhood acute lymphoblastic leukemia and diabetes mellitus, type 1: a Bayesian approach. *Am J Epidemiol*. 2005;161(12):1168-1180.

152. Haynes A, Bower C, Bulsara MK, Finn J, Jones TW, Davis EA. Perinatal risk factors for childhood type 1 diabetes in Western Australia--a population-based study (1980-2002). *Diabet Med*. 2007;24(5):564-570.

153. Australian Bureau of Statistics. Technical Paper Socio-Economic Indexes for Areas (SEIFA), 2016. ABS Catalogue No. 2033.0.55.001. Australian Bureau of Statistics, 2016. .

154. Australian Bureau of Statistics. Socio-Economic Indexes For Areas: Getting a handle on individual diversity within areas. Canberra: Australian Bureau of Statistics, 2011.

155. Bruno G, Spadea T, Picariello R, Gruden G, Barutta F, Cerutti F, et al. Early life socioeconomic indicators and risk of type 1 diabetes in children and young adults. *J Pediatr* 2013;162(3):600-605.

156. Stokholm J, Thorsen J, Chawes BL, Schjørring S, Krogfelt KA, Bønnelykke K, et al. Cesarean section changes neonatal gut colonization. *J Allergy Clin Immunol*. 2016;138(3):881-889.e882.

157. de Goffau MC, Fuentes S, van den Bogert B, Honkanen H, de Vos WM, Welling GW, et al. Aberrant gut microbiota composition at the onset of type 1 diabetes in young children. *Diabetologia*. 2014;57(8):1569-1577.

158. Vatanen T, Franzosa EA, Schwager R, Tripathi S, Arthur TD, Vehik K, et al. The human gut microbiome in early-onset type 1 diabetes from the TEDDY study. *Nature*. 2018;562(7728):589-594.

159. Patterson C, Guariguata L, Dahlquist G, Soltész G, Ogle G, Silink M. Diabetes in the young – a global view and worldwide estimates of numbers of children with type 1 diabetes. *Diabetes Res Clin Pract*. 2014;103(2):161-175.

160. World Health Organization. WHO Statement on Caesarean Section Rates, Geneva: HO 2015.

161. Rebelo F, da Rocha CM, Cortes TR, Dutra CL, Kac G. High cesarean prevalence in a national population-based study in Brazil: the role of private practice. *Acta Obstet Gynecol Scand*. 2010;89(7):903-908.

162. Lumbiganon P, Laopaiboon M, Gulmezoglu AM, Souza JP, Taneepanichskul S, Ruyan P, et al. Method of delivery and pregnancy outcomes in Asia: the WHO global survey on maternal and perinatal health 2007-08. *Lancet*. 2010;375(9713):490-499.

163. Lee HY, Lu CL, Chen HF, Su HF, Li CY. Perinatal and childhood risk factors for early-onset type 1 diabetes: a population-based case-control study in Taiwan. *Eur J Public Health*. 2015;25(6):1024-1029.

164. Algert CS, McElduff A, Morris JM, Roberts CL. Perinatal risk factors for early onset of Type 1 diabetes in a 2000-2005 birth cohort. *Diabet Med*. 2009;26(12):1193-1197.

165. The Lancet, Diabetes & Endocrinology (2019). Type 1 diabetes research: poised for progress. *Lancet Diabetes Endocrinol*.7(1):1-1.

166. Schübeler D. Function and information content of DNA methylation. *Nature*. 2015;517:321.

167. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*. 2012;13:484.

168. Berlin I, Heilbronner C, Georgieu S, Meier C, Spreux-Varoquaux O. Newborns' cord blood plasma cotinine concentrations are similar to that of their delivering smoking mothers. *Drug Alcohol Depend*. 2010;107(2):250-252.

169. Sopori M. Effects of cigarette smoke on the immune system. *Nat Rev Immunol*. 2002;2(5):372-377.

170. Paul DS, Teschendorff AE, Dang MAN, Lowe R, Hawa MI, Ecker S, et al. Increased DNA methylation variability in type 1 diabetes across three immune effector cell types. *Nat Commun*. 2016;7:13555.

171. Mattsson K, Kallen K, Rignell-Hydbom A, Lindh CH, Jonsson BA, Gustafsson P, et al. Cotinine validation of self-reported smoking during pregnancy in the Swedish Medical Birth Register. *Nicotine Tob Res*. 2016;18(1):79-83.

172. Qiu F, Liang CL, Liu H, Zeng YQ, Hou S, Huang S, et al. Impacts of cigarette smoking on immune responsiveness: up and down or upside down? *Oncotarget*. 2017;8(1):268-284.

173. Imkampe A-K, Gulliford MC. Trends in type 1 diabetes incidence in the UK in 0- to 14-year-olds and in 15- to 34-year-olds, 1991–2008. *Diabet Med*. 2011;28(7):811-814.

174. Hansen AR, Akomolafe TO, McGalliard Z, Belle-Isle L, Zhang J. Striving to Meet Healthy People 2020 Objectives: Trend Analysis of Maternal Smoking. *Public Health Rep*. 2018;133(6):644-649.

175. Australian Institute of Health and Welfare (2019). Trends in incidence of type 1 diabetes by age, 2000-2016. Data tables: Diabetes 2019. Available from www.aihw.gov.au/reports/diabetes/diabetes-snapshot/data. Accessed 1 May 2019

176. Australian Institute of Health and Welfare, Laws PJ, Hilder L (2008). Australia's mothers and babies 2006. Perinatal statistics series no. 22. Cat. no. PER 46. Available from www.aihw.gov.au/reports/mothers-babies/australias-mothers-babies-2006/contents/table-of-contents. Accessed 1 May 2019.

177. Australian Institute of Health and Welfare (2018). Australia's mothers and babies 2016—in brief. Perinatal Statistics Series no. 34. Catalogue no. PER 97. Canberra: Australian Institute of Health and Welfare, 2018. Available from www.aihw.gov.au/reports/mothers-babies/australias-mothers-babies-2016-in-brief/contents/table-of-contents. Accessed 3 May 2019.

178. Cardwell CR, Carson DJ, Patterson CC. Parental age at delivery, birth order, birth weight and gestational age are associated with the risk of childhood Type 1 diabetes: a UK regional retrospective cohort study. *Diabetic Medicine*. 2005;22(2):200-206.

179. Stene LC, Magnus P, Lie RT, Søvik O, Joner G. Maternal and paternal age at delivery, birth order, and risk of childhood onset type 1 diabetes: population based cohort study. *BMJ*. 2001;323(7309):369.

180. Cardwell CR, Stene LC, Joner G, Bulsara MK, Cinek O, Rosenbauer J, et al. Maternal age at birth and childhood type 1 diabetes: a pooled analysis of 30 observational studies. *Diabetes*. 2010;59(2):486-494.

181. Pearce N. Analysis of matched case-control studies. *BMJ*. 2016;352:i969.

182. Mansournia MA, Hernán MA, Greenland S. Matched designs and causal diagrams. *Int J Epidemiol*. 2013;42(3):860-869.

183. Lindell N, Carlsson A, Josefsson A, Samuelsson U. Maternal obesity as a risk factor for early childhood type 1 diabetes: a nationwide, prospective, population-based case-control study. *Diabetologia*. 2018;61(1):130-137.

184. Rasmussen T, Stene LC, Samuelsen SO, Cinek O, Wetlesen T, Torjesen PA, et al. Maternal BMI before pregnancy, maternal weight gain during pregnancy, and risk of persistent positivity for multiple diabetes-associated autoantibodies in children with the high-risk HLA genotype: the MIDIA study. *Diabetes care*. 2009;32(10):1904.

185. Magnus MC, Olsen SF, Granstrom C, Lund-Blix NA, Svensson J, Johannesen J, et al. Paternal and maternal obesity but not gestational weight gain is associated with type 1 diabetes. *Int J Epidemiol*. 2018;47(2):417-426.

186. Hidayat K, Zou S-Y, Shi B-M. The influence of maternal body mass index, maternal diabetes mellitus, and maternal smoking during pregnancy on the risk of childhood-onset type 1 diabetes mellitus in the offspring: Systematic review and meta-analysis of observational studies. *Obes Rev*. 2019;20(8):1106-1120.

187. Arkkola T, Kautiainen S, Takkinen HM, Kenward MG, Nevalainen J, Uusitalo U, et al. Relationship of maternal weight status and weight gain rate during pregnancy to the development of advanced beta cell autoimmunity in the offspring: a prospective birth cohort study. *Pediatr Diabetes*. 2011;12(5):478-484.

188. Antvorskov JC, Halldorsson TI, Josefsen K, Svensson J, Granström C, Roep BO, et al. Association between maternal gluten intake and type 1 diabetes in offspring: National prospective cohort study in Denmark. *BMJ (Online)*. 2018;362.

189. Lamb MM, Myers MA, Barriga K, Zimmet PZ, Rewers M, Norris JM. Maternal diet during pregnancy and islet autoimmunity in offspring. *Pediatr Diabetes*. 2008;9(2):135-141.

190. Virtanen SM, Uusitalo L, Kenward MG, Nevalainen J, Uusitalo U, Kronberg-Kippila C, et al. Maternal food consumption during pregnancy and risk of advanced beta-cell autoimmunity in the offspring. *Pediatr Diabetes*. 2011;12(2):95-99.

191. The EURODIAB Substudy 2 Study Group . Vitamin D supplement in early childhood and risk for Type I (insulin-dependent) diabetes mellitus *Diabetologia*. 1999;42(1):51-54.

192. Sorensen IM, Joner G, Jenum PA, Eskild A, Torjesen PA, Stene LC. Maternal serum levels of 25-hydroxy-vitamin D during pregnancy and risk of type 1 diabetes in the offspring. *Diabetes*. 2012;61(1):175-178.

193. Sorensen IM, Joner G, Jenum PA, Eskild A, Brunborg C, Torjesen PA, et al. Vitamin D-binding protein and 25-hydroxyvitamin D during pregnancy in mothers whose children later developed type 1 diabetes. *Diabetes Metab Res Rev*. 2016;32(8):883-890.

194. Dong J-Y, Zhang W, Chen J, Zhang Z-L, Han S-F, Qin L-Q. Vitamin D Intake and Risk of Type 1 Diabetes: A Meta-Analysis of Observational Studies. *Nutrients*. 2013;5(9):3551-3562.

195. Thorsen SU, Marild K, Olsen SF, Holst KK, Tapia G, Granstrom C, et al. Lack of Association Between Maternal or Neonatal Vitamin D Status and Risk of Childhood Type 1 Diabetes: A Scandinavian Case-Cohort Study. *Am J Epidemiol*. 2018;187(6):1174-1181.

196. Niinisto S, Takkinen HM, Uusitalo L, Rautanen J, Nevalainen J, Kenward MG, et al. Maternal dietary fatty acid intake during pregnancy and the risk of preclinical and clinical type 1 diabetes in the offspring. *Br J Nutr*. 2014;111(5):895-903.

197. Sorensen IM, Joner G, Jenum PA, Eskild A, Stene LC. Serum long chain n-3 fatty acids (EPA and DHA) in the pregnant mother are independent of risk of type 1 diabetes in the offspring. *Diabetes Metab Res Rev*. 2012;28(5):431-438.

198. Niinisto S, Takkinen HM, Uusitalo L, Rautanen J, Vainio N, Ahonen S, et al. Maternal intake of fatty acids and their food sources during lactation and the risk of preclinical and clinical type 1 diabetes in the offspring. *Acta Diabetol*. 2015;52(4):763-772.

199. Thorsen SU, Halldorsson TI, Bjerregaard AA, Olsen SF, Svensson J. Maternal and early life iron intake and risk of childhood type 1 diabetes: A Danish case-cohort study. *Nutrients*. 2019;11(4).

200. Størdal K, McArdle HJ, Hayes H, Tapia G, Viken MK, Lund-Blix NA, et al. Prenatal iron exposure and childhood type 1 diabetes. *Sci Rep*. 2018;8(1).

201. Kyvsgaard J, Overgaard AJ, Thorsen SU, Hansen TH, Pipper CB, Mortensen HB, et al. High neonatal blood iron content is associated with the risk of childhood type 1 diabetes mellitus. *Nutrients*. 2017;9(11).

202. Field CJ. The Immunological Components of Human Milk and Their Effect on Immune Development in Infants. *J Nutr*. 2005;135(1):1-4.

203. Hosea Blewett HJ, Cicalo MC, Holland CD, Field CJ. The immunological components of human milk. *Adv Food Nutr Res*. 2008;54:45-80.

204. Lund-Blix NA, Dydensborg Sander S, Størdal K, Nybo Andersen A-M, Rønningen KS, Joner G, et al. Infant Feeding and Risk of Type 1 Diabetes in Two Large Scandinavian Birth Cohorts. *Diabetes Care*. 2017:dc170016.

205. Güngör D, Nadaud P, LaPergola CC, Dreibelbis C, Wong YP, Terry N, et al. Infant milk-feeding practices and diabetes outcomes in offspring: a systematic review. *Am J Clin Nutr*. 2019;109(Supplement_1):817S-837S.

206. Cardwell CR, Stene LC, Ludvigsson J, Rosenbauer J, Cinek O, Svensson J, et al. Breast-feeding and childhood-onset type 1 diabetes: a pooled analysis of individual participant data from 43 observational studies. *Diabetes Care*. 2012;35(11):2215-2225.

207. Hakola L, Takkinen H-M, Niinistö S, Ahonen S, Nevalainen J, Veijola R, et al. Infant Feeding in Relation to the Risk of Advanced Islet Autoimmunity and Type 1 Diabetes in Children With Increased Genetic Susceptibility: A Cohort Study. *Am J Epidemiol*. 2018;187(1):34-44.

208. Chmiel R, Beyerlein A, Knopff A, Hummel S, Ziegler AG, Winkler C. Early infant feeding and risk of developing islet autoimmunity and type 1 diabetes. *Acta Diabetol*. 2015;52(3):621-624.

209. O'Sullivan A, Farver M, Smilowitz JT. The Influence of Early Infant-Feeding Practices on the Intestinal Microbiome and Body Composition in Infants. *Nutr Metab Insights*. 2015;8(Suppl 1):1-9.

210. Forbes JD, Azad MB, Vehling L, Tun HM, Konya TB, Guttman DS, et al. Association of Exposure to Formula in the Hospital and Subsequent Infant Feeding Practices With Gut Microbiota and Risk of Overweight in the First Year of Life. *JAMA Pediatr*. 2018;172(7):e181161.

211. Norris JM, Barriga K, Klingensmith G, Hoffman M, Eisenbarth GS, Erlich HA, et al. Timing of initial cereal exposure in infancy and risk of islet autoimmunity. *JAMA*. 2003;290(13):1713-1720.

212. Uusitalo U, Lee HS, Andren Aronsson C, Vehik K, Yang J, Hummel S, et al. Early Infant Diet and Islet Autoimmunity in the TEDDY Study. *Diabetes Care*. 2018;41(3):522-530.

213. Piescik-Lech M, Chmielewska A, Shamir R, Szajewska H. Systematic Review: Early Infant Feeding and the Risk of Type 1 Diabetes. *J Pediatr Gastroenterol Nutr*. 2017;64(3):454-459.

214. Ziegler AG, Schmid S, Huber D, Hummel M, Bonifacio E. Early infant feeding and risk of developing type 1 diabetes-associated autoantibodies. *JAMA*. 2003;290(13):1721-1728.

215. Ashraf AP, Eason NB, Kabagambe EK, Haritha J, Meleth S, McCormick KL. Dietary iron intake in the first 4 months of infancy and the development of type 1 diabetes: a pilot study. *Diabetol Metab Syndr*. 2010;2:58.

216. Tapia G, Stordal K, Marild K, Kahrs CR, Skrivarhaug T, Njolstad PR, et al. Antibiotics, acetaminophen and infections during prenatal and early life in relation to type 1 diabetes. *Int J Epidemiol*. 2018;47(5):1538-1548.

217. Clausen TD, Bergholt T, Bouaziz O, Arpi M, Eriksson F, Rasmussen S, et al. Broad-Spectrum Antibiotic Treatment and Subsequent Childhood Type 1 Diabetes: A Nationwide Danish Cohort Study. *PLoS One*. 2016;11(8):e0161654.

218. Hviid A, Svanstrom H. Antibiotic use and type 1 diabetes in childhood. *Am J Epidemiol*. 2009;169(9):1079-1084.

219. Cardwell CR, Svensson J, Waldhoer T, Ludvigsson J, Sadauskaite-Kuehne V, Roberts CL, et al. Interbirth interval is associated with childhood type 1 diabetes risk. *Diabetes*. 2012;61(3):702-707.

220. Cardwell CR, Stene LC, Joner G, Bulsara MK, Cinek O, Rosenbauer J, et al. Birth order and childhood type 1 diabetes risk: a pooled analysis of 31 observational studies. *Int J Epidemiol*. 2011;40(2):363-374.

221. Morgan E, Halliday SR, Campbell GR, Cardwell CR, Patterson CC. Vaccinations and childhood type 1 diabetes mellitus: a meta-analysis of observational studies. *Diabetologia*. 2016;59(2):237-243.

222. Aiello LP. Diabetic retinopathy and other ocular findings in the diabetes control and complications trial/epidemiology of diabetes interventions and complications study. *Diabetes Care*. 2014;37(1):17-23.

223. de Boer IH. Kidney disease and related findings in the diabetes control and complications trial/epidemiology of diabetes interventions and complications study. *Diabetes Care*. 2014;37(1):24-30.

224. Orchard TJ, Nathan DM, Zinman B, Cleary P, Brillon D, Backlund JY, et al. Association between 7 years of intensive treatment of type 1 diabetes and long-term mortality. *JAMA*. 2015;313(1):45-53.

225. Beck RW, Bergenstal RM, Laffel LM, Pickup JC. Advances in technology for management of type 1 diabetes. *Lancet (London, England)*. 2019;394(10205).

226. Katz M, Laffel L. Mortality in Type 1 Diabetes in the Current Era: Two Steps Forward, One Step Backward. Editorial. *JAMA*. 2015;313(1):35-36.

227. Marks HH. Longevity and mortality of diabetics. *Am J Public Health Nations Health*. 1965;55:416-423.

228. Stene LC. Gaps in life expectancy for people with type 1 diabetes. *Diabetologia*. 2016;59(6):1150-1152.

229. Beck RW, Tamborlane WV, Bergenstal RM, Miller KM, DuBose SN, Hall CA. The T1D Exchange clinic registry. *J Clin Endocrinol Metab*. 2012;97(12):4383-4389.

230. Rydén A, Sörstadius E, Bergenheim K, Romanovschi A, Thorén F, Witt EA, et al. The Humanistic Burden of Type 1 Diabetes Mellitus in Europe: Examining Health Outcomes and the Role of Complications. *PloS one*. 2016;11(11):e0164977-e0164977.

231. Rechenberg K, Whittemore R, Holland M, Grey M. General and diabetes-specific stress in adolescents with type 1 diabetes. *Diabetes Res Clin Pract*. 2017;130:1-8.

232. Mitchell SJ, Hilliard ME, Mednick L, Henderson C, Cogen FR, Streisand R. Stress among fathers of young children with type 1 diabetes. *Fam Syst Health*. 2009;27(4):314-324.

233. Australian Government, Departmnet of Health. Australian Government subsidy of Continuous Glucose Monitors. accessed 8/04/20119 http://www.health.gov.au/internet/main/publishing.nsf/Content/MC15-004873-continuous-glucose-monitors.

234. Diabetes Victoria. Understanding insulin pumps. Information for people with type 1 diabetes. Melbourne Victoria, 2015.

235. Australian Institute of Health and Welfare. Insulin pump use in Australia. Diabetes series no.18. Cat. no. CVD 58. Canberra: Australian Institute of Health and Welfare, 2012.

236. Diabetes Australia 2014. Insulin Pump Therapy in Australia . https://www.diabetesaustralia.com.au/news/13542?type=articles. Access date: 23-10-2019.

237. Hannonen R, Komulainen J, Riikonen R, Ahonen T, Eklund K, Tolvanen A, et al. Academic skills in children with early-onset type 1 diabetes: the effects of diabetes-related risk factors. *Dev Med Child Neurol*. 2012;54(5):457-463.

238. Cameron FJ, Scratch SE, Nadebaum C, Northam EA, Koves I, Jennings J, et al. Neurological consequences of diabetic ketoacidosis at initial presentation of type 1 diabetes in a prospective cohort study of children.(Report). *Diabetes Care*. 2014;37(6):1554.

239. Glaab LA, Brown R, Daneman D. School attendance in children with Type 1 diabetes. *Diabet Med*. 2005;22(4):421-426.

240. Duca LM, Wang B, Rewers M, Rewers A. Diabetic Ketoacidosis at Diagnosis of Type 1 Diabetes Predicts Poor Long-term Glycemic Control. *Diabetes Care*. 2017;40(9):1249-1255.

241. Semenkovich K, Bischoff A, Doty T, Nelson S, Siller AF, Hershey T, et al. Clinical presentation and memory function in youth with type 1 diabetes. *Pediatr Diabetes*. 2016;17(7):492-499.

242. Ghetti S, Lee JK, Sims CE, DeMaster DM, Glaser NS. Diabetic Ketoacidosis and Memory Dysfunction in Children with Type 1 Diabetes. *J Pediatr*. 2010;156(1):109-114.

243. Rosenthal JM, Amiel SA, Yágüez L, Bullmore E, Hopkins D, Evans M, et al. The effect of acute hypoglycemia on brain function and activation: a functional magnetic resonance imaging study. *Diabetes*. 2001;50(7):1618.

244. Semenkovich K, Patel PP, Pollock AB, Beach KA, Nelson S, Masterson JJ, et al. Academic abilities and glycaemic control in children and young people with Type 1 diabetes mellitus. *Diabet Med*. 2016;33(5):668-673.

245. Hannonen R, Komulainen J, Eklund K, Tolvanen A, Riikonen R, Ahonen T. Verbal and academic skills in children with early-onset type 1 diabetes. *Dev Med Child Neurol*. 2010;52(7):e143-e147.

246. Persson E, Persson S, Gerdtham U-G, Steen Carlsson K. Effect of type 1 diabetes on school performance in a dynamic world: new analysis exploring Swedish register data. *Applied Economics*. 2019;51(24):2606-2622.

247. Houng B, Justman M. NAPLAN Scores as Predictors of Access to Higher Education in Victoria. Melbourne Institute Working Paper Series Working Paper No. 22/14. University of Melbourne, 2014.

248. Goss P, Sonnemann J, Chisholm C, Nelson L. Widening gaps: what NAPLAN tells us about student progress, Grattan Institute, Australia, 2016. (ISBN: 978-1-925015-82-9).

249. Cassells R, Duncan A, Abello A, D'Souza G, Nepal B. Smart Australians: education and innovation in Australia, AMP NATSEM, The University of Canberra, 2012.

250. Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA, Davidian M. Doubly Robust Estimation of Causal Effects. *Am J Epidemiol*. 2011;173(7):761-767.

251. Australian Bureau of Statistics . Census of Population and Housing: Reflecting Australia - Stories from the Census, 2016. Aboriginal and Torres Strait Islander Population – South Australia. access date 22/11/2019 https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/2071.0~2016~Main%20Features~Aboriginal%20and%20Torres%20Strait%20Islander%20Population%20-%20South%20Australia~10004

252. Australian Institute of Health and Welfare 2017. Health-adjusted life expectancy in Australia: expected years lived in full health 2011. Australian Burden of Disease Study series no.16. BOD 17. Canberra: AIHW.

253. Australian Institute of Health and Welfare 2019. Australia's mothers and babies 2017—in brief. Perinatal statistics series no. 35. Cat. no. PER 100. Canberra: AIHW.

254. Australian Bureau of Statistics. 2016 Census Data Summary, Snapshot of Australia

255. Australian Institute of Health and Welfare 2018. Australia's health 2018. Australia's health series no. 16. AUS 221. Canberra: AIHW.

256. Nuske T, Pilkington R, Gialamas A, Chittleborough C, Smithers L, Lynch J (2016). The SA Early Childhood Data Project. Research Series. School of Public Health, The University of Adelaide, 2016. Available from https://health.adelaide.edu.au/betterstart/publications/reports/sa-ecdp-report-2016.pdf Accessed 4 May 2019.

257. Pilkington R, Grant J, Chittleborough C, Gialamas A, Montgomerie A, Lynch J. 2017. Child Protection in South Australia. BetterStart Child Health and Development Research Group, School of Public Health, The University of Adelaide.

258. Pilkington R, Montgomerie A, Grant J, Gialamas A, Malvaso C, Smithers L, Chittlebourough C, Lynch J. (2019). An innovative linked data platform to improve the wellbeing of children - the South Australian Early Childhood Data Project. In

Australian Institute of Health and Welfare (Ed.), Australia's welfare 2019: data insights. Australia's welfare series no. 14. Cat. no. AUS 226. Canberra.

259. Scheil W, Jolly K, Scott J, Catcheside B, Sage L, Kennare R (2017). Pregnancy outcome in South Australia 2015. Available from www.sahealth.sa.gov.au/wps/wcm/connect/public+content/sa+health+internet/about+us/health+statistics/pregnancy+outcome+statistics. Accessed 2 April 2018.

260. South Australia, Health Care Regulations 2008, under the Health Care Act 2008. Version: 1.7.2019.

261. McLean A. *Validation of the 1994 South Australian perinatal data collection form*. Adelaide, SA: Pregnancy Outcome Unit, Epidemiology Branch, Department of Human Services; 2001.

262. Chan A, Scott J, Nguyen A, Keane R. Pregnancy outome in South Australia 2001. Pregnancy outcome unit. Epidemiology Branch, Department Of Human Services Issn 0819-3835. Governmnet of South Australia. .

263. Governmnt of South Australia, SA Health. https://www.sahealth.sa.gov.au/wps/wcm/connect/public+content/sa+health+internet/about+us/our+performance/our+data+collections/admitted+patient+care.

264. Australian Consortium for Classification Development (2017). The International Statistical Classification of Diseases and Related Health Problems, tenth revision, Australian modification (ICD-10-AM/ACHI/ACS), 10th edn. Available from www.ihpa.gov.au/what-we-do/icd-10-am-achi-acs-previous-editions. Accessed 2 May 2017.

265. Srinivasan S, Craig ME, Beeney L, Hayes R, Harkin N, Ambler GR, et al. An ambulatory stabilisation program for children with newly diagnosed type 1 diabetes. *Med J Aust*. 2004;180(6):277-280.

266. Australian Institute of Health and Welfare. The National (insulin-treated) Diabetes Register 2016; Quality Statement. AIHW Data Quality Statements, Endorsed 29/01/2018.

267. Haynes A, Bower C, Bulsara MK, Jones TW, Davis EA. Continued increase in the incidence of childhood type 1 diabetes in a population-based Australian sample (1985-2002). *Diabetologia*. 2004;47(5):866-870.

268. Australian Curriculum, Assessment and Reporting Authority 2017, NAPLAN Achievement in Reading, Writing, Language Conventions and Numeracy: National Report for 2017, ACARA, Sydney, 2017.

269. Centre for Health Record Linkage (2012). Quality assurance; Available from www.cherel.org.au/quality-assurance. Accessed 7 February 2018.

270. National Research Infrastructure of Australia. Big data population based research, accessing health and human services data through SA NT Datalink. NCRIS, an Austraian Government initiative

271. Holman CD, Bass AJ, Rouse IL, Hobbs MS. Population-based linkage of health records in Western Australia: development of a health services research linked database. *Aust N Z J Public Health*. 1999;23(5):453-459.

272. Greenland S, Morgenstern H. Confounding in health research. *Annu Rev Public Health*. 2001;22:189-212.

273. McNamee R. Confounding and confounders. *Occup Environ Med*. 2003;60(3):227-234.

274. Greenland S, Robins JM. Identifiability, Exchangeability, and Epidemiological Confounding. *Int J Epidemiol*. 1986;15(3):413-419.

275. Jager KJ, Zoccali C, MacLeod A, Dekker FW. Confounding: What it is and how to deal with it. *Kidney Int*. 2008;73(3):256-260.

276. Hernan MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006;60(7):578-586.

277. DAGitty. http://www.dagitty.net/.

278. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10(1):37-48.

279. Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. 3rd ed. ed. Philadelphia

Sydney: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008.

280. Suttorp MM, Siegerink B, Jager KJ, Zoccali C, Dekker FW. Graphical presentation of confounding in directed acyclic graphs. *Nephrol Dial Transplant*. 2014;30(9):1418-1423.

281. Hernan MA, Hernandez-Diaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol*. 2002;155(2):176-184.

282. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology (Cambridge, Mass)*. 2009;20(4):488-495.

283. Hernan MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol*. 2016;183(8):758-764.

284. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006;60(7):578-586.

285. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010;21(3):383-388.

286. Arnold BF, Ercumen A. Negative Control Outcomes: A Tool to Detect Bias in Randomized Trials. *JAMA*. 2016;316(24):2597-2598.

287. Haneuse S, VanderWeele TJ, Arterburn D. Using the E-Value to Assess the Potential Effect of Unmeasured Confounding in Observational Studies. *JAMA*. 2019;321(6):602-603.

288. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: Introducing the E-value. *Ann Intern Med*. 2017;167(4):268-274.

289. StataCorp. 2011. Stata: Release 12. Statistical Software. College Station,TX: StataCorp LP.

290. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338:b2393.

291. Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*. 2013;14(5):365-376.

292. Bruin, J. 2006. newtest: command to compute new test. UCLA: Statistical Consulting Group. https://stats.idre.ucla.edu/stata/ado/analysis/.

293. Enders CK. Multiple imputation as a flexible tool for missing data handling in clinical research. *Behav Res Ther*. 2017;98:4-18.

294. White IR, Carlin JB. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Stat Med*. 2010;29(28):2920-2931.

295. Machekano RN, Dorsey G, Hubbard A. Efficacy studies of malaria treatments in Africa: efficient estimation with missing indicators of failure. *Stat Methods Med Res*. 2007;17(2):191-206.

296. Yoke CW, Khalid ZM. Comparison of multiple imputation and complete-case in a simulated longitudinal data with missing covariate. *AIP Conference Proceedings*. 2014;1605(1):918-922.

297. Rubin DB. Inference and missing data. *Biometrika*. 1976;63(3):581-592.

298. Rubin DB (1987) Multiple imputation for nonresponse in surveys Wiley, Hoboken,NJ.

299. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med*. 2011;30(4):377-399.

300. Hughes RA, Heron J, Sterne JAC, Tilling K. Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *International Journal of Epidemiology*. 2019;48(4):1294-1304.

301. Bhaskaran K, Smeeth L. What is the difference between missing completely at random and missing at random? *Int J Epidemiol*. 2014;43(4):1336-1339.

302. Adler NE, Newman K. Socioeconomic Disparities In Health: Pathways And Policies. *Health Aff (Millwood)*. 2002;21(2):60-76.

303. Holmqvist BM, Lofman O, Samuelsson U. A low incidence of type 1 diabetes between 1977 and 2001 in south-eastern Sweden in areas with high population density and which are more deprived. *Diabet Med*. 2008;25(3):255-260.

304. du Prel J-B, Icks A, Grabert M, Holl RW, Giani G, Rosenbauer J. Socioeconomic conditions and type 1 diabetes in childhood in North Rhine–Westphalia, Germany. *Diabetologia*. 2007;50(4):720-728.

305. Australian Bureau of Statistics, Technical paper, Socio-Economic Indexes for Areas (SEIFA) 2011. ABS Catalogue no. 2033.0.55.001. Canberra: Australian Bureau of Statistics, 2013.

306. Hess KR. Graphical methods for assessing violations of the proportional hazards assumption in Cox regression. *Stat Med*. 1995;14(15):1707-1723.

307. Schoenfeld D. Partial Residuals for The Proportional Hazards Regression Model. *Biometrika*. 1982;69(1):239-241.

308. World Health Organization. The Global Numbers and Costs of Additionally Needed and Unnecessary Caesarean Sections Performed per Year: Overuse as a Barrier to Universal Coverage. World Health Report (2010) Background Paper, No 30.

309. Adlercreutz EH, Wingren CJ, Vincente RP, Merlo J, Agardh D. Perinatal risk factors increase the risk of being affected by both type 1 diabetes and coeliac disease. *Acta Paediatr*. 2015;104(2):178-184.

310. Svensson J, Carstensen B, Mortensen HB, Borch-Johnsen K. Early childhood risk factors associated with type 1 diabetes - Is gender important? *Eur J Epidemiol*. 2005;20(5):429-434.

311. International Diabetes Federation. Diabetes ATLAS. Seventh edition. Belgium: International Diabetes Federation, 2015.

312. *Breslow NE, Day NE. Statistical methods in cancer research*. Lyon, France International Agency for Research on Cancer 1987.

313. Wasserstein RL. Lazar NA. The ASA's Statement on p-Values: Context, Process, and Purpose. *The American Statistician*. 2016;70(2):129-133.

314. Begum M, Pilkington R, Chittleborough C, Lynch J, Penno M, Smithers L. Caesarean section and risk of type 1 diabetes: whole-of-population study. *Diabet Med*. 2019.

315. Hope S, Pearce A, Whitehead M, Law C. Parental employment during early childhood and overweight at 7-years: findings from the UK Millennium Cohort Study. *BMC Obes*. 2015;2(1):33.

316. Lynch KF, Subramanian SV, Ohlsson H, Chaix B, Lernmark A, Merlo J. Context and disease when disease risk is low: the case of type 1 diabetes in Sweden. *J Epidemiol Community Health* 2010;64:789--795.

317. Australian Health Ministers' Advisory Council. Aboriginal and Torres Strait Islander Health Performance Framework 2017 Report. Canberra: AHMAC, 2017.

318. Australian Institute of Health and Welfare. Australia's mothers and babies 2015—in brief. Perinatal statistics series no. 33. Canberra: Australian Institute of Health and Welfare , 2017.

319. Appropriate technology for birth. *Lancet*. 1985;2(8452):436-437.

320. Maria GD-B, Elizabeth KC, Monica C, Magda M, Glida H, Noah F, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci*. 2010;107(26):11971.

321. Dobbins TA, Simpson JM, Sullivan EA, Roberts CL. Australian national birthweight percentiles by sex and gestational age, 1998-2007. *MJA*. 2012;197(5):291-294.

322. National Key Centre for Social Applications of Geographical Information Systems. The University of Adelaide, Australia. Department of Health and Aged Care. Occasional papers. Canberra. 2001.

323. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338.

324. Wasserstein RL, Lazar NA. The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician*. 2016;70(2):129-133.

325. Ziegler A-G, Nepom GT. Prediction and Pathogenesis in Type 1 Diabetes. *Immunity*. 2010;32(4):468-478.

326. Department for Health and Ageing. Prelabour rupture of the membranes (PROM) ≥ 37 weeks Clinical Guideline, 2015. South Australian Perinatal Practice Guidelines. Department for Health and Ageing, Government of South Australia. ISBN: 978-1-74243-746-0.

327. Marshall JK, Thabane M, Garg AX, Clark W, Meddings J, Collins SM. Intestinal permeability in patients with irritable bowel syndrome after a waterborne outbreak of acute gastroenteritis in Walkerton, Ontario. *Aliment Pharmacol Ther*. 2004;20(11-12):1317-1322.

328. White IR, Royston P. Imputing missing covariate values for the Cox model. *Stat Med*. 2009;28(15):1982-1998.

329. Banderali G, Martelli A, Landi M, Moretti F, Betti F, Radaelli G, et al. Short and long term health effects of parental tobacco smoking during pregnancy and lactation: a descriptive review. *J Transl Med*. 2015;13(1):327.

330. Kiechl-Kohlendorfer U, Ralser E, Pupp Peglow U, Reiter G, Griesmaier E, Trawoger R. Smoking in pregnancy: a risk factor for adverse neurodevelopmental outcome in preterm infants? *Acta Paediatr*. 2010;99(7):1016-1019.

331. Australian Institute of Health and Welfare (2019). National Core Maternity Indicators. Smoking during pregnancy. Available from www.aihw.gov.au/reports/mothers-babies/ncmi-data-visualisations/contents/antenatal-period-indicators/smoking-during-pregnancy. Accessed 16 September 2019.

332. Ekblad M, Gissler M, Korkeila J, Lehtonen L. Trends and risk groups for smoking during pregnancy in Finland and other Nordic countries. *Eur J Public Health*. 2014;24(4):544-551.

333. Penno MA, Couper JJ, Craig ME, Colman PG, Rawlinson WD, Cotterill AM, et al. Environmental determinants of islet autoimmunity (ENDIA): a pregnancy to early life cohort study in children at-risk of type 1 diabetes. *BMC Pediatr*. 2013;13:124.

334. England L, Zhang J. Smoking and risk of preeclampsia: a systematic review *Front Biosci*. 2007;12:2471-2483.

335. Li X, Li W, Liu G, Shen X, Tang Y. Association between cigarette smoking and Parkinson's disease: a meta-analysis. Arch Gerontol Geriatr. 2015;61(3):510-516.

336. Berkowitz L, Schultz BM, Salazar GA, Pardo-Roa C, Sebastián VP, Álvarez-Lobos MM, et al. Impact of cigarette smoking on the gastrointestinal tract inflammation: Opposing effects in crohn's disease and ulcerative colitis. Front Immunol. 2018;9:74-74.

337. Ungprasert P, Crowson CS, Matteson EL. Smoking, obesity and risk of sarcoidosis: A population-based nested case-control study. Respir Med. 2016;120:87-90.

338. Xia Y, Kellems RE. Is preeclampsia an autoimmune disease? . *Clin Immunol*. 2009;133(1):1-12.

339. Cleynen I, Boucher G, Jostins L, Schumm LP, Zeissig S, Ahmad T, et al. Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *Lancet* 2016;387(10014):156-167.

340. He J, Ryder AG, Li S, Liu W, Zhu X. Glycemic extremes are related to cognitive dysfunction in children with type 1 diabetes: A meta-analysis. *Journal of Diabetes Investigation*. 2018;9(6):1342-1353.

341. Hwang JJ, Jiang L, Sanchez Rangel E, Fan X, Ding Y, Lam W, et al. Glycemic Variability and Brain Glucose Levels in Type 1 Diabetes. *Diabetes*. 2019;68(1):163-171.

342. Ly TT, Gallego PH, Davis EA, Jones TW. Impaired awareness of hypoglycemia in a population-based sample of children and adolescents with type 1 diabetes. *Diabetes care*. 2009;32(10):1802.

343. McCrimmon RJ, Ryan CM, Frier BM. Diabetes and cognitive dysfunction. *Lancet*. 2012;379(9833):2291-2299.

344. Lin A, Northam EA, Rankins D, Werther GA, Cameron FJ. Neuropsychological profiles of young people with type 1 diabetes 12 yr after disease onset. *Pediatr Diabetes*. 2010;11(4):235-243.

345. Gaudieri PA, Chen R, Greer TF, Holmes CS. Cognitive Function in Children With Type 1 Diabetes. *A meta-analysis*. 2008;31(9):1892-1897.

346. Fleming M, Fitton CA, Steiner MFC, McLay JS, Clark D, King A, et al. Educational and Health Outcomes of Children Treated for Type 1 Diabetes: Scotland-Wide Record Linkage Study of 766,047 Children. *Diabetes Care*. 2019:dc182423.

347. Diabetes Victoria and National Diabetes Service Scheme-NDSS. Mastering diabetes in Preschools and Schools. Access date 14/08/2019. ndss.com.au.

348. Karges B, Schwandt A, Heidtmann B, Kordonouri O, Binder E, Schierloh U, et al. Association of Insulin Pump Therapy vs Insulin Injection Therapy With Severe Hypoglycemia, Ketoacidosis, and Glycemic Control Among Children, Adolescents, and Young Adults With Type 1 Diabetes. *JAMA*. 2017;318(14):1358-1366.

349. Stata Manual. Teffects aipw — Augmented inverse-probability weighting. stata.com.

350. Stata Manual. Teffects multivalued — Multivalued treatment effects. stata.com.

351. NAPLAN score equivalence tables 2017. https://www.nap.edu.au/results-and-reports/how-to-interpret/score-equivalence-tables accessed on 4/4/2019.

352. South Australia Department of Education. Managing health in education and care. access date 14/08/2019 https://www.education.sa.gov.au/supporting-students/health-e-safety-and-wellbeing/health-support-planning/managing-health-education-and-care/physical-health/diabetes.

353. Birkebaek NH, Drivvoll AK, Aakeson K, Bjarnason R, Johansen A, Samuelsson U, et al. Incidence of severe hypoglycemia in children with type 1 diabetes in the Nordic countries in the period 2008-2012: association with hemoglobin A 1c and treatment modality. *BMJ Open Diabetes Res Care*. 2017;5(1):e000377.

354. Australian Government Department of Social Services . Disability and Carers. Guide to the List of Recognised Disabilities. Access date 14/18/2019 https://www.dss.gov.au/our-responsibilities/disability-and-carers/benefits-payments/carer-allowance/guide-to-the-list-of-recognised-disabilities.

355. Noorani M, Ramaiya K, Manji K. Glycaemic control in type 1 diabetes mellitus among children and adolescents in a resource limited setting in Dar es Salaam - Tanzania. *BMC Endocr Disord*. 2016;16(1):29.

356. McKinney PA, Feltbower RG, Stephenson CR, Reynolds C. Children and young people with diabetes in Yorkshire: a population-based clinical audit of patient data 2005/2006. *Diabet Med*. 2008;25(11):1276-1282.

357. Perry LB, Lubienski C, Ladwig J. How do learning environments vary by school sector and socioeconomic composition? Evidence from Australian students. *Aust J Educ*. 2016;60(3):175-190.

358. Persson S, Gerdtham UG, Steen Carlsson K, Swedish Childhood Diabetes Study G. Labor market consequences of childhood onset type 1 diabetes. *Econ Hum Biol*. 2016;23:180-192.

359. Galobardes B. Socioeconomic inequalities in health: individual or area level; does it matter? *BMC Public Health*. 2012;12(1):171.

360. Cawley J, Liu F. Maternal employment and childhood obesity: a search for mechanisms in time use data. *Econ Hum Biol*. 2012;10(4):352-364.

361. Anderson PM. Parental employment, family routines and childhood obesity. *Econ Hum Biol*. 2012;10(4):340-351.

362. Censin JC, Nowak C, Cooper N, Bergsten P, Todd JA, Fall T. Childhood adiposity and risk of type 1 diabetes: A Mendelian randomization study. *PLoS Med*. 2017;14(8):e1002362-e1002362.

363. Sandall J, Tribe RM, Avery L, Mola G, Visser GHA, Homer CSE, et al. Short-term and long-term effects of caesarean section on the health of women and children. *The Lancet*. 2018;392(10155):1349-1357.

364. Keag OE, Norman JE, Stock SJ. Long-term risks and benefits associated with cesarean delivery for mother, baby, and subsequent pregnancies: Systematic review and meta-analysis. *PLoS Med*. 2018;15(1):e1002494.

365. Smithers LG, Mol BW, Jamieson L, Lynch JW. Cesarean birth is not associated with early childhood body mass index. *Pediatr Obes*. 2017;12 Suppl 1:120-124.

366. Molina G, Weiser TG, Lipsitz SR, Esquivel MM, Uribe-Leitz T, Azad T, et al. Relationship Between Cesarean Delivery Rate and Maternal and Neonatal Mortality. *JAMA*. 2015;314(21):2263-2270.

367. Pineles BL, Park E, Samet JM. Systematic review and meta-analysis of miscarriage and maternal exposure to tobacco smoke during pregnancy. *Am J Epidemiol*. 2014;179(7):807-823.

368. Shah NR, Bracken MB. A systematic review and meta-analysis of prospective studies on the association between maternal cigarette smoking and preterm delivery. *Am J Obstet Gynecol*. 2000;182(2):465-472.

369. Rayfield S, Plugge E. Systematic review and meta-analysis of the association between maternal smoking in pregnancy and childhood overweight and obesity. *J Epidemiol Community Health*. 2017;71(2):162-173.

370. Jiang T, Li G, Xu J, Gao S, Chen X. The Challenge of the Pathogenesis of Parkinson's Disease: Is Autoimmunity the Culprit? *Front Immunol*. 2018;9:2047-2047.

371. Garretti F, Agalliu D, Lindestam Arlehamn CS, Sette A, Sulzer D. Autoimmunity in Parkinson's Disease: The Role of α-Synuclein-Specific T Cells. *Front Immunol*. 2019;10(303).

372. Wei J, Liu C-X, Gong T-T, Wu Q-J, Wu L. Cigarette smoking during pregnancy and preeclampsia risk: a systematic review and meta-analysis of prospective studies. *Oncotarget*. 2015;6(41):43667.

373. Greenland S, Morgenstern H. Confounding in Health Research. *Annu Rev Public Health*. 2001;22(1):189-212.

374. Hackshaw A. Small studies: strengths and limitations. *Eur Respir J*. 2008;32(5):1141-1143.

375. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124.

376. Group TS. The Environmental Determinants of Diabetes in the Young (TEDDY) Study. *Ann N Y Acad Sci*. 2008;1150:1-13.

377. The Environmental Determinants of Diabetes in the Young (TEDDY) Study. *Ann N Y Acad Sci*. 2008;1150.

378. Szklo M. Population-based cohort studies. *Epidemiol Rev*. 1998;20(1):81-90.

379. Herkert D, Vijayakumar P, Luo J, Schwartz JI, Rabin TL, DeFilippo E, et al. Cost-Related Insulin Underuse Among Patients With Diabetes. *JAMA Internal Medicine*. 2019;179(1):112-114.

380. Young MT, Lord JH, Patel NJ, Gruhn MA, Jaser SS. Good cop, bad cop: quality of parental involvement in type 1 diabetes management in youth. *Curr Diab Rep*. 2014;14(11):546-546.

381. Hua X, Carvalho N, Tew M, Huang ES, Herman WH, Clarke P. Expenditures and Prices of Antihyperglycemic Medications in the United States: 2002-2013. *JAMA*. 2016;315(13):1400-1402.

382. Ewen M, Joosse H-J, Beran D, Laing R. Insulin prices, availability and affordability in 13 low-income and middle-income countries. *BMJ Global Health*. 2019;4(3):e001410.

383. Mainous AG, 3rd, Diaz VA, Saxena S, Baker R, Everett CJ, Koopman RJ, et al. Diabetes management in the USA and England: comparative analysis of national surveys. *J R Soc Med*. 2006;99(9):463-469.

384. National Health Interview Survey. long Term Trends in Health Insurance Coverage. U.S. Department of Health and Human Services ,Centers for Disease Control and Prevention , National Center for Health Statistics. Released 07/2019. .

385. Cefalu WT, Dawes DE, Gavlak G, Goldman D, Herman WH, Van Nuys K, et al. Insulin Access and Affordability Working Group: Conclusions and Recommendations. *Diabetes Care*. 2018;41(6):1299-1311.

386. Chow CK, Ramasundarahettige C, Hu W, Alhabib KF, Avezum A, Cheng X, et al. Availability and affordability of essential medicines for diabetes across high-income, middle-income, and low-income countries: a prospective epidemiological study. *Lancet Diabetes Endocrinol*. 2018;6(10):798-808.

387. Moucheraud C, Lenz C, Latkovic M, Wirtz VJ. The costs of diabetes treatment in low- and middle-income countries: a systematic review. *BMJ Global Health*. 2019;4(1):e001258.

388. U.S. House of Representatives. Prices of Diabetes Drugs for Seniors and the Uninsured in the United States and Abroad. California's 27th Congressional District Staff Report, Committee on Oversight and Reform, U.S. House of Representatives, March, 2019.

389. Xu K, Soucat A & Kutzin J et al. Public Spending on Health: A Closer Look at Global Trends. Geneva: World Health Organization; 2018. (WHO/HIS/HGF/HFWorkingPaper/18.3). Licence: CC BY-NC-SA 3.0 IGO.

390. Davison KK, Charles JN, Khandpur N, Nelson TJ. Fathers' Perceived Reasons for Their Underrepresentation in Child Health Research and Strategies to Increase Their Involvement. *Matern Child Health J*. 2017;21(2):267-274.