Supplemental Online Content

Lindquist A, Hastie R, Kennedy A, et al. Developmental outcomes for children after elective birth at 39 weeks' gestation. *JAMA Pediatr*. Published online May 9, 2022. doi:10.1001/jamapediatrics.2022.1165

eAppendix. A priori statistical analysis plan
eTable 1. Primary analyses using complete (nonmissing) case data
eTable 2. Primary analyses using alternative modelling techniques
eTable 3. Secondary analysis of mode of birth at 39 weeks' gestation and developmental vulnerability using complete case (nonmissing) data

This supplemental material has been provided by the authors to give readers additional information about their work.

eAppendix. A priori statistical analysis plan

Exploring associations between obstetric interventions and childhood health outcomes

School-age educational and developmental outcomes for singleton children following elective birth at term (39 weeks) vs expectant management

STATISTICAL ANALYSIS PLAN

Version 11.0, Sept 2021

RESEARCH TEAM – roles, responsibilities and signatures

Dr Anthea Lindquist
Perinatal Epidemiologist, FRANZCOG (Mercy Perinatal/University of Melbourne)
Dr Roxanne Hastie
Perinatal Epidemiologist, PhD (Mercy Perinatal/University of Melbourne)
Professor Stephen Tong
Clinician scientist, FRANZCOG (Mercy Perinatal/University of Melbourne)
Professor Susan Walker
Maternal fetal medicine specialist, FRANZCOG (Mercy Perinatal/University of Melbourne)
Professor Jeanie Cheong
Academic neonatologist, FRACP (Royal Women's Hospital, University of Melbourne)
Dr Jon Quach
Educationalist academic, PhD (Department of Education, University of Melbourne)
Dr Richard Hiscock
Statistician and anaesthetist, ANZCA (Mercy Hospital for Women)
Dr Lyle Gurrin
Biostatistician, PhD (Department of Public Health, University of Melbourne)
Dr Amber Kennedy
Obstetrician and IVF specialist, PhD candidate (Mercy Perinatal/University of Melbourne)

BACKGROUND

Induction of labour is one of the most common obstetric interventions. In 2017, 43.1% of all mothers were already being induced at term gestation in Australia (1). This is likely to increase in light of the findings of the landmark clinical trial published in 2018: the ARRIVE trial (2).

The ARRIVE trial. This US based study randomised 6000 low risk nulliparous women at term (randomised between 38+0-38+6). It found induction of labour at 39 weeks gestation (compared with deferring the induction until 41 weeks gestation to give women the chance to go into labour spontaneously) reduced the relative risk of a caesarean section by about 16% (18.6% in the induction group vs 22.2%, p<0.001) (2). They also showed that being induced may reduce the risk of severe perinatal complications including ….might be useful to list the main ones (relative risk 0.80 [95% confidence interval 0.64-1.00]). Furthermore, in a survey of participants, the trial found that those who were induced felt 'more in control' than those randomised to awaiting spontaneous labour (2). It is entirely feasible that in coming years, more than 50% of all pregnancies will be induced simply because they have reached 39 weeks, without any further obstetric indication.

As a caution against this increasing practice, there is evidence suggesting that the risk of special education needs may be significantly higher for those who delivered at 39 weeks gestation, compared with those who are delivered later (40 - 41 weeks) (3). However, past studies investigating this association have relied upon only severe assessments of childhood outcome (eg. a diagnosis of special education needs) or have examined outcomes across the full spectrum of gestational age (4). The impact of elective delivery at 39 weeks on childhood developmental and educational outcomes needs to be examined from a clinical perspective, in a contemporaneous Australian population. This will help inform practice and guide both clinicans and patients in point-of-care decision-making.

PROJECT OVERVIEW

There are three study aims:

1) To describe the characteristics of mothers and children for the overall cohort, and each gestational age exposure group, in addition to the distribution of outcome data.
    a. This will be performed as a descriptive analysis.
2) To determine the *causal effect* of elective birth at 39 weeks on school age outcomes.
    a. Using a casual inference framework, we will attempt to replicate the conditions of an RCT and potential recruitment of women at the point of decision making in antenatal care – 39 weeks. *Do we offer elective birth (induction or caesarean section) to women who have reached 39 weeks or do we encourage them to continue pregnancy (and labour spontaneously, be induced or have a caesarean section later in pregnancy)?*
    b. Elective birth at 39 weeks (induction of labour or elective caesarean section) will be compared with all births after 39 weeks (by any mode of birth).
3) To examine for a potential gradient in childhood outcomes with increasing week of term gestational age (37-42 weeks).
    a) Traditional regression will be used to examine for a gradient of change in childhood outcome metrics (Odds Ratio or Relative Risk) by change in week of gestational age (37, 38, 39, 40, 41 and >=42 weeks gestation).

DATA SOURCES

1) Victorian Perinatal Data Collection
    a) Approximately 1,000,000 Victorian births from 2005-2014
    b) Routinely-collected Birth Outcome System (BOS) data
2) Australian Early Developmental Census
    a) Performed *every three years* across Australia. The data are collected by prep school teachers who conduct an observational assessment for every child in their first year of full-time primary school.
    b) Birth data to be linked with corresponding AEDC data from 2009 (unlikely to link with any of our births), 2012, 2015, 2018.
    c) Five domains of assessment:
        i) Physical health and wellbeing – looks at absence from school due to illness, independence with toileting, coordination, fine motor skills and physical ability to climb stairs, and other aspects.
        ii) Social competence – assesses ability to get along with peers, respect for adults and others, curiosity about the world and eagerness to play and learn, willingness to help other children who are hurt or having difficulty with an activity and problem-solving day-to-day tasks.
        iii) Emotional maturity – aims to identify impulsive or bullying behaviour, as well as the child's ability to make decisions, be patient and attentive, share and wait their turn in games.
        iv) Language and cognitive skills – assesses ability to use English effectively; listen and follow instructions in addition to more in-depth assessments of reading, writing and numeracy.

v) Communication skills and general knowledge – looks at the child's ability to tell a story, participate in imaginative play, communicate their own needs to their peers and to adults and assesses their knowledge about the world (eg. the seasons, types of fruit, different animals etc).

d) From the teacher's assessment, an individual score is calculated for each child. Outcome measures include *both* a **score** (out of 10) and a **centile** for each domain for every child. Centile scores are used to define:
   i) *developmentally "vulnerable" = <10th centile*
   ii) developmentally "at risk" = 10-25th centile
   iii) developmentally "on track" = > 25th centile

PART ONE: DESCRIPTVE ANALYSIS

1. Flow chart of study participants:
Total number in cohort > Number of linked records > Number with matched AEDC outcome data > Cohort following exclusions > Gestational age cohorts

Exclusion criteria: stillbirth, multiple pregnancy, babies with known congenital anomalies

2. Population characteristics to be described for:
a) Overall cohort
b) Children born at 37, 38, 39, 40, 41 and >=42 *completed* weeks gestation

VARIABLE	DESCRIPTOR
Child baseline data	
Sex	(% female)
Language-background other than English	(%)
Aboriginal/Torres Strait Islander	(%)
Parents overseas born	(%)
Birthweight	Mean, SD and centile
Gestational age of delivery	Median and IQR
Mode of birth – elective caesarean section, emergency caesarean section, instrumental birth (forceps and vacuum) and spontaneous vaginal birth	(%)
Maternal baseline data	
Maternal age	Median (IQR)
BMI	Mean (SD)
IVF/ART	%
Perinatal complications (PET, suspected SGA etc)	N (%)
Marital status: -Married/de facto -Single -Divorced -Widowed	(% of each category)
Level of maternal education: -Year 10 or below -High school	(% of each category)

Trade/TAFE/diploma/certificate -Undergraduate Uni degree or above	
Maternal Occupation: -Senior management, -Other business manager, -Tradesman/woman, clerks, sales and service staff, -Machine operators, -Not in paid work, -Not Stated/Unknown	(% of each category)
SEIFA quintile: -1 (most disadvantaged) -2 -3 -4 -5 (least disadvantaged)	(%) for each category
Second parent baseline data	
Level of second parent education: -Level of maternal education -Year 10 or below -High school Trade/TAFE/diploma/certificate -Undergraduate Uni degree or above	(% of each category)
Second Parent Occupation: -Senior management, -Other business manager, -Tradesman/woman, clerks, sales and service staff, -Machine operators, -Not in paid work, -Not Stated/Unknown	(% of each category)
Childhood outcome data	
AEDC	
Individual score	Median (IQR)
Developmental vulnerability ($<10^{th}$ in ≥2 domains)	%
Developmental vulnerability in any of the five domains ($<10^{th}$ centile) – physical, social, emotional, language and cognitive (assessed separately)	%

3. Figures

- Graphs will be generated to describe a) the proportion of developmentally vulnerable children ($<10^{th}$ centile ≥2 domains) and b) the proportion of children 'on track' amongst the overall cohort, and amongst each completed week of gestational age (37-42 weeks).

PART TWO: CAUSAL ANALYSIS

Combining the power of a large state-wide dataset and record-linkage methodology, we will examine whether being born electively at 39 weeks (induction or elective caesarean) is associated with poorer childhood developmental and educational outcomes compared with those born after 39 weeks gestation by any mode of birth. This component of the study seeks to replicate the conditions of an RCT ie. The women that could reasonably be rectuited to a trial at 39 weeks are those not in labour but who can be 'booked' or planned for an intervention (induction or caesarean section).

P – all eligible Victorian births 2005-2014 that reached at least 39 weeks gestation
I – Children born after induction or elective caesarean section at 39 completed weeks gestational age
C – Children born at 40, 41 and >=42 weeks by any mode of delivery
O – Developmental and educational outcomes – AEDC scores
T – The nominal age at which children are assessed for AEDC (foundation school year: 4-6 years of age)

Null Hypothesis

The population distribution of school-aged developmental and educational outcomes for children born after induction of labour or elective caesarean section at 39 weeks gestation is the same as those of all other children who reach at least 39 weeks gestation and go on to be born by any mode of birth.

Statistical analysis

Causal research question: *What is the causal effect of gestational age at birth on the AEDC-defined outcomes for children born electively at 39 weeks compared with those born later (40, 41 and >=42 weeks)?*
Exposure: Completed week of gestational age.
Primary outcome: A score of <10th centile in "2 or more" domains (Binary yes/no).
Secondary outcomes: A score in the lowest decile in each domain: 1) physical health and wellbeing, 2) social competence, 3) emotional maturity, 4) language and cognitive skills, and 5) communication skills and general knowledge.
Inclusion criteria: Children born during the study time period
Exclusion criteria: Preterm gestation, multiple pregnancies, stillbirths, babies with known congenital anomaly.

Anticipated sample size: All births in Victoria during 2005-2014 will be matched to AEDC childhood outcome data. AEDC is conducted every 3 years and as a result, our total cohort will include matched perinatal-childhood data for approximately 224,973 women. This sample size has been estimated from state-wide, published birth reports for the relevant years of the study. After exclusion of pre-term births (~8.3%), multiple pregnancies (~3.3%), babies with known congenital anomalies (~3.0%) and babies born after 42 weeks gestation (0.5%) we estimate that *our sample size will include approximately 190,905 matched mother-child data pairs.*

The baseline risk of developmental vulnerability in children born at term using AEDC assessment is 20% (5). Based on our study numbers, we will be able to detect with 95% power, a 0.8% difference in risk between groups.

Multilevel structure of cohort: Typical of perinatal data, we expect the number of maternal clusters will be large (~90,000) and small in size (1 to < 3 children per cluster). Exposure and outcome are at level 1 (children) with confounders at both level 1 and level 2 (maternal). For valid inference analysis methods must assess the impact of this clustering, and account for it if required.

Causal Inference methodology

The steps to ***causal inference****:*

1. *Think carefully about the precise nature of the causal question to be addressed, usually around a currently humanly feasible intervention that we wish to apply to the whole of the target population, and compare the outcome (averaged over the whole of the target population) to the corresponding average outcome that occurs when no one in the target population gets the intervention;*
2. *Convert the causal question in Step 1 to a precise quantity to be estimated (a causal estimand), usually using the notation of the potential outcomes approach to causal inference.*
3. *At least state and, if possible, be rigorous and honest about the the assumptions (e.g. consistency, conditional exchangeability, positivity) under which the estimand in Step 2 can be identified from the data at hand;*
4. *Find (or develop) estimation strategies (i.e. statistical methods) that are valid under the assumptions in Step 3. Apply these to the data in Step 3;*
5. *Use statistical tools to assess quantitatively the sensitivity of the results to plausible departures from the assumptions in Step 3.*

1) Develop a causal model using directed acyclic graphs (DAGs) to both identify and summarise causal pathways.
    a. Potential confounders – demographic and perinatal – will be determined by the above-named authorship team (a quorum of specialists with expert knowledge):
        i. Childs gender (binary)
        ii. Child's age at testing
        iii. Documented special education needs (SEN)
        iv. Birthweight (continuous)
        v. Maternal age (continuous)
        vi. BMI
        vii. Level of maternal education (categorical)
        viii. SEIFA socio-economic quintile (categorical)
        ix. Obstetric complications such as preeclampsia, diabetes and suspected fetal growth restriction
        x. Mode of birth (categorical – elective caesrean section, emergency caesarean section, instrumental birth, spontaneous vaginal birth)
        xi. Auxillary covariates (for imputation only):
            1. Aboriginal/Torres Strait Islander background (binary)
            2. Language-background other than English (binary)
            3. IVF (or ART) conception (binary)
            4. Parents overseas born
            5. Second parent level of education
            6. Parity

b. AEDC scores will be standardised for child's age at assessment
c. The list of covariates to be included in the propensity score model will be finalised after consideration of potential open backdoor causal paths and conditioning effects of collider nodes (as documented in the DAG, below). The list above will be used as a start. Distinction will be made between the standard epidemiological definition of a confounder and the causal pathway definition.

   Causal pathway and confounding *(excerpt from Causal Inference , What if. M Hernan & J Robins p85: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/)*
   "..the bias has the same structure: it is due to the presence of a cause (known covariate L or unknown covariare U) that is shared by the treatment A and the outcome Y, which results in an open backdoor path between A and Y. We refer to the bias caused by shared causes of treatment and outcome as confounding,…"

These potential confounders (SM_covariates) will then be used in the construction of a propensity score model for exposure assignment (vide infra).

2) Handling of missing data using Multiple Imputation

Overview: Imputation of missing data must account for the maternal clustering of this dataset. Further, the imputation model should be congenial to the planned substantive causal models. Detailed discussion of these issues and the approach to imputation and causal models for this analysis are presented in Gestation-AEDC analysis notes[RJH 16 December 2020] – attached. In summary:

a. Consideration of missing data will be confined to all the covariates (SM_covariates) identified for the substantive causal model. When imputing these missing covariate values the imputation model will include all SM covariates as well as additional (auxillary) covariates from the CCOPMM database/AEDC databases.
b. To account for multilevel structure multiple imputation will be performed using either joint multivariate model or chained equations that account for maternal clustering, and the imputed datsets will then be used for substantive analysis. Subjects with missingness greater than 50% of SM covariates will be excluded from analysis, this proportion will be assessed and if substantial, may result in changes to the SM_covariate list used for ongoing analysis. Imputation model will include outcome, exposure, substative model covariates and a limited number of auxillary variables to improve imputation. For each variable datatype used will be consistent with that used in the analysis model, see analysis notes for details relating to covariate transformations (including ratio variables), categorized continuous variables and interactions.
c. Handling of clustering within mothers in the imputation models. Given the large number of clusters and small cluster size the substantive analysis model may be performed using both marginal models and random effect models. To account for clustered data, imputation will be performed using appropriate models in either JOMO and MICE packages in R. Given the large number of clusters and their small cluster size both imputation and analysis models will account for clustering using cluster adjusted (robust) SE.
d. Diagnostics will then be performed to assess the quality of the imputed datasets.
e. For each of these imputed datasets the outcome metric, average point estimates and standard errors (95% CI) appropriately adjusted, given the imputation framework, will

be obtained. The number of imputed datasets will be set at a minimum of 20 with the number increased to equal to the highest percentage of missingness in the raw SM_covariate list (ceiling of 50% by design). Details of imputation and assessment of adequacy will be formally presented.

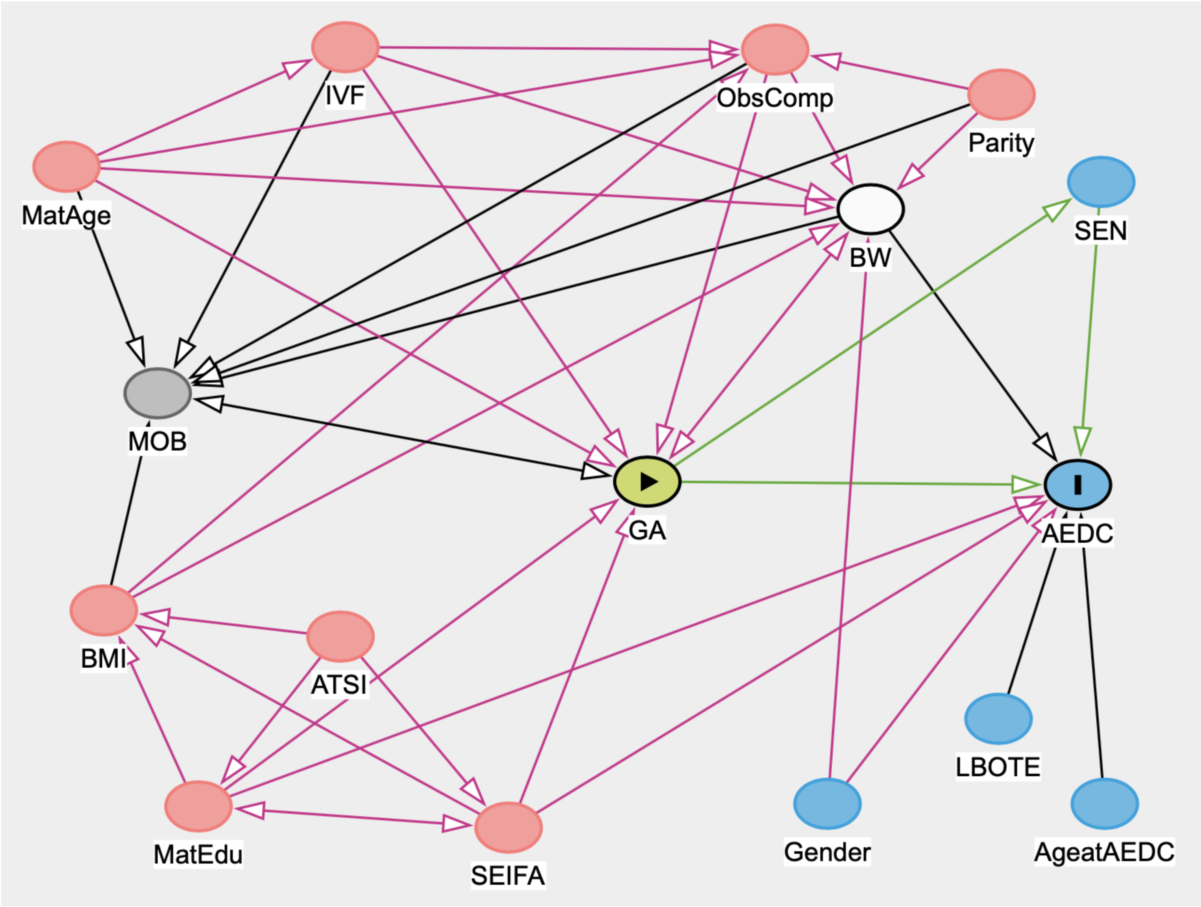
- For causal analysis the positivity assumption is met if there is overlap in the distribution of IPW between exposure arms. That is, the analysis cohort must exclude those observations where IPW don't overlap. The distribution of IPW can be transformed (stabilized weights) to improve both overlap and covariate balance. These will be explored before defining the analysis weights.

3) Substantive causal models
   a. All regression modelling will be performed within the multiple imputation frame-work outlined and be compatible with the imputation model used.
   b. We plan to model exposure as binary (39 vs >39 weeks). The metric used will be compatible with the presentation of exposure in the imputation model.
   c. The outcome metric is considered to be absolute risk difference and relative risk between those born electively at 39 weeks and those who reached 39 weeks but who were not induced or delivered by elective caesarean. A propensity score model for this binary exposure will be built using the covariates listed above and informed by the causal diagram.
   d. Initial modelling will incorporate propensity score (PS) covariate balance using IPW. The PS will be formed using all SM_covariates in a regression model that accounts for maternal clustering by using either robust variance estimate or a random intercept model. We will combine both regression adjustment and augmented IPW in a doubly robust regression model. Sensitivity analyses will be performed using estimates derived from Targeted Maximal Likelihood Estimation (TMLE). Pooled estimates will be combined across imputed datsets using Rubin's rules.

   Software: analysis will be performed using Stata v16 (StataCorp. 2019. *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LLC. Causal inference will be performed using teffects commands (Treatment-effects estimation for observational data - see TE -teffects) run within within Stata's multiple imputation suite; and the TMLE command in R.

```
dag {
bb="-5.784,-5.126,6.677,8.388"
AEDC [outcome,pos="5.221,1.861"]
ATSI [pos="-2.236,3.982"]
AgeatAEDC [pos="5.493,6.320"]
BMI [pos="-4.629,3.617"]
BW [adjusted,pos="3.126,-1.993"]
GA [exposure,pos="0.867,1.815"]
Gender [pos="2.689,6.320"]
IVF [pos="-2.186,-4.259"]
LBOTE [pos="4.418,5.132"]
MOB [pos="-4.084,0.581"]
MatAge [pos="-5.005,-2.587"]
MatEdu [pos="-3.671,6.356"]
ObsComp [pos="2.160,-4.225"]
Parity [pos="4.449,-3.597"]
SEIFA [pos="-0.536,6.653"]
```

```
SEN [pos="5.457,-2.375"]
ATSI -> BMI
ATSI -> MatEdu
ATSI -> SEIFA
AgeatAEDC -> AEDC
BMI -> BW
BMI -> MOB
BMI -> ObsComp
BW -> AEDC
BW -> MOB
BW <-> GA
GA -> AEDC
GA -> SEN
GA <-> MOB
Gender -> AEDC
Gender -> BW
IVF -> BW
IVF -> GA
IVF -> MOB
IVF -> ObsComp
LBOTE -> AEDC
MatAge -> BW
MatAge -> GA
MatAge -> IVF
MatAge -> MOB
MatAge -> ObsComp
MatEdu -> AEDC
MatEdu -> BMI
MatEdu -> GA
MatEdu <-> SEIFA
ObsComp -> BW
ObsComp -> GA
ObsComp -> MOB
Parity -> BW
Parity -> MOB
Parity -> ObsComp
SEIFA -> AEDC
SEIFA -> BMI
SEIFA -> GA
SEN -> AEDC
}
```



PART THREE: TRADITIONAL REGRESSION ANALYSIS

Null Hypothesis

The population distribution of school-aged developmental and educational outcomes is the same for all children born across the range of term gestation at 37, 38, 39, 40, 41 or >=42 completed weeks by any mode of birth.

Statistical analysis

Research question: Is there a gradient across the range of term gestational age (37-42 weeks gestation) in AEDC-defined school outcomes for singleton children, born by elective caesarean section or induction of labour?
Exposure: Completed week of term gestational age.
Primary outcome: A score of <10th centile in "2 or more" domains (Binary yes/no).
Secondary outcomes: A score in the lowest decile in each domain: 1) physical health and wellbeing, 2) social competence, 3) emotional maturity, 4) language and cognitive skills, and 5) communication skills and general knowledge.
Inclusion criteria: Live-born children free of lethal anomalies born between 37 and 42 completed weeks of gestation
Exclusion criteria: Preterm gestation, multiple pregnancies, stillbirths, babies with known congenital anomaly.

Potential confounders selected a priori and based on descriptive analysis, included in regression model. Standard logistic regression modelling will be performed to assess for the change in OR/RR for each completed categorical week of gestational age (37 through to 42 weeks). In keeping with the previous study aim, this component of the study will include only children born electively at each gestational age (elective caesarean section and induction of labour).

CONTINGENT ANALYSES

Depending on the findings of the above-described AEDC analysis and at the discretion of the research team, the anlaysis may be repeated using NAPLAN school-aged outcomes.

Exploring associations between obstetric interventions and global childhood health in the first 2000 days of life

School-age educational and developmental outcomes for singleton children following induction of labour at 39 weeks vs planned caesarean section.

STATISTICAL ANALYSIS PLAN

Version 11.0, September 2021

RESEARCH TEAM

Dr Roxanne Hastie
Perinatal Epidemiologist, PhD (Mercy Perinatal/University of Melbourne)
Dr Anthea Lindquist
Perinatal Epidemiologist, FRANZCOG (Mercy Perinatal/University of Melbourne)
Professor Stephen Tong
Clinician scientist, FRANZCOG (Mercy Perinatal/University of Melbourne)
Professor Susan Walker
Maternal fetal medicine specialist, FRANZCOG (Mercy Perinatal/University of Melbourne)
Professor Jeanie Cheong
Academic neonatologist, FRACP (Royal Women's Hospital, University of Melbourne)
Dr Jon Quach
Educationalist academic, PhD (Department of Education, University of Melbourne)
Dr Richard Hiscock
Statistician and anaesthetist, ANZCA (Mercy Hospital for Women)
Dr Lyle Gurrin
Biostatistician, PhD (Department of Public Health, University of Melbourne)
Dr Amber Kennedy
Obstetrician and IVF specialist and PhD candidate (Mercy Perinatal/University of Melbourne)

BACKGROUND

Planned births at 39weeks gestation are increasing in Australia and other Western countries(1). Women planning to birth at 39 weeks can do so via induction of labour or elective caesarean section.

Although the short-term risk and benefits for both mother and baby for each mode of birth have been well described, there is little available evidence of longer-term outcomes for offspring. Elucidating the medium- and longer-term outcomes for offspring born via induction of labour compared with planned caesarean section will allow women and their clinical team make more informed decisions whilst planning for birth.

Furthermore, those women who attempt a vaginal birth may ultimately birth via instrumental vaginal birth (birth with the assistance of forceps or ventouse/vacuum) or in-labour emergency caesarean section. These interventions are often performed to expedite birth either because there are concerns that the fetus may be in peril, or because the fetus has not birthed after a long period of active pushing (for instrumental birth). Again, the short-term risks associated with mode of birth are well documented, however the longer-term risks beyond the neonatal period are almost unknown.

The use of forceps or ventouse applies significant forces to the fetal head (between 80-146 kPa) (2) and that compared to vaginal birth, instrumental birth is associated with increased risks of major scalp trauma, scalp fractures, intracerebral bleeding and neonatal seizures (3). It is therefore plausible that the significant pressures applied to the fetal head by instrumental births may be associated with adverse effects on longer-term global childhood health outcomes. Given there is an alternative management option - caesarean section - whether instrumental birth is associated with adverse childhood health outcomes merits investigation.

Combining the power of a large state-wide dataset and record-linkage methodology, we will examine, among women undergoing planned birth at 39^{+0} – 39^{+6} weeks, whether- being born via induction of labour compared with caesarean section is associated with poorer childhood developmental and educational outcomes at age of school entry. *This analysis will use statistical methods that attempt to simulate clinical trial conditions (casual inference pathways). Hence, the cohort is split at a point where a clinical decision is made (book a caesarean vs an induction of labour) rather than the final mode of birth.*

Additionally, among women who have an induction of labour at 39 weeks, we will assess childhood developmental and educational outcomes by mode of birth, comparing those women who have an unassisted vaginal birth to i) instrumental birth and ii) in labour caesarean section**.** *The second analysis does not use casual inference pathway analysis but seeks to compare outcomes after different modes of birth.*

PROJECT OVERVIEW

This study will be performed in three parts:

4) Descriptive analysis for the overall cohort, those who birth undergo an induction of labour and those who have a planned caesarean section.
5) Determine the *causal effect* of induction of labour at 39 weeks gestation on school age outcomes:
    a. A causal inference approach will be used to mimic a randomised controlled trial and the potential recruitment of women at 39 weeks gestation who are planning birth to either 1) induction of labour or 2) caesarean section. For planned birth at 39 weeks gestation, which mode/onset of birth should be offered, induction of labour or caesarean section
6) Investigate *the associations* between mode of birth and school aged developmental outcome.
    a. Following on from the cohort identified in Part 2, *women undergoing induction of labour at 39 weeks* will be used to further examine the associations between women having an unassisted vaginal birth and i) instrumental vaginal birth and ii) emergency in-labour CS and school aged developmental outcomes.

DATA SOURCES

Perinatal data for every birth in Victoria will be obtained from the Victorian Perinatal Data Collection Unit (VPDC). Outcomes measures will be assessed using the Australian Early Developmental Census (AEDC) data metric.

3) Victorian Perinatal Data Collection
   a) Approximately 1,000,000 Victorian births from 2005-2014
   b) Routinely collected Birth Outcome System (BOS) data

4) Australian Early Developmental Census
   a) Performed *every three years* across Australia. The data are collected by prep school teachers who conduct an observational assessment for every child in their first year of full-time primary school.
   b) Birth data to be linked with corresponding AEDC data from 2009, **2012, 2015, 2018.**
   c) **Five domains of assessment:**
      i) Physical health and wellbeing – looks at absence from school due to illness, independence with toileting, coordination, fine motor skills and physical ability to climb stairs, and other aspects.
      ii) Social competence – assesses ability to get along with peers, respect for adults and others, curiosity about the world and eagerness to play and learn, willingness to help other children who are hurt or having difficulty with an activity and problem-solving day-to-day tasks.
      iii) Emotional maturity – aims to identify impulsive or bullying behaviour, as well as the child's ability to make decisions, be patient and attentive, share and wait their turn in games.
      iv) Language and cognitive skills – assess ability to use English effectively; listen and follow instructions in addition to more in-depth assessments of reading, writing and numeracy.
      v) Communication skills and general knowledge – looks at the child's ability to tell a story, participate in imaginative play, communicate their own needs to their peers and to adults and assesses their knowledge about the world (eg. the seasons, types of fruit, different animals etc).
   d) From the teacher's assessment, an individual score is calculated for each child. Outcome measures include *both* a **score** (out of 10) and a **centile** for each domain for every child. Centile scores are used to define:
      i) *developmentally "vulnerable" = <10th centile*
      ii) developmentally "at risk" = 10-25th centile
      iii) developmentally "on track" = > 25th centile

PART ONE: DESCRIPTIVE ANALYSIS

1. Flow chart of study participants:
Total number in cohort > Number of linked records > Number with matched AEDC outcome data

Exclusion criteria: stillbirth, multiple pregnancy, babies with known congenital anomalies, cases of placenta praevia or placental abruption, breech birth, missing gestational age, missing mode of birth, missing birth onset, and those born before 39+0 weeks and after 40+0 weeks.

2.Population characteristics to be described for:

a) Overall cohort

b) Children born by induction of labour and elective caesarean section, in addition to IOL and unassisted vaginal birth, instrumental vaginal birth, in-labour caesarean section.

VARIABLE	DESCRIPTOR
Child baseline data	
Gender	(% female)
Year of testing	(%)
Language-background other than English	(%)
Aboriginal/Torres Strait Islander	(%)
Birthweight	Mean and SD
Gestational age of delivery	Median and IQR
Mode of delivery – elective caesarean section, emergency caesarean section, instrumental birth (forceps and vacuum) and spontaneous vaginal birth	(%)
Maternal baseline data	
Maternal age	Median and IQR
Maternal BMI	Median and IQR
Parity -Nulliparous -Multiparous	(% of each category)
Marital status: -Married/de facto -Single -Divorced -Widowed	(% of each category)
Level of maternal education: -Year 10 or below -High school Trade/TAFE/diploma/certificate -Undergraduate Uni degree or above	(% of each category)
SEIFA quintile: -1 (most disadvantaged) -2 -3 -4 -5 (least disadvantaged)	(% of each category)

Maternal medical and obstetric conditions: -Hypertensive disorders -Type 1 diabetes -Type 2 diabetes -Renal disease -Circulatory diseases -Gestational diabetes -Suspected SGA -APH -PROM -GBS positive	(% of each category)
Second parent baseline data	
Level of second parent education: -Level of maternal education -Year 10 or below -High school Trade/TAFE/diploma/certificate -Undergraduate Uni degree or above	(% of each category)
Childhood outcome data	
AEDC	
Individual score	Median (IQR)
Developmental vulnerability (<10th in ≥2 domains)	%

3. Figures

- Graphs will be generated to describe a) the proportion of developmentally vulnerable children (<10th centile ≥2 domains) and b) the proportion of children 'on track' amongst the overall cohort, and amongst IOL and elective caesarean section group

PART TWO: CAUSAL ANALYSIS

Among women birthing at 39 weeks gestation, what are the causal effects of induction of labour on the school-age developmental and educational outcomes for children, compared with those born via elective caesarean section?

P – all eligible Victorian births 2005-2014
I – Children born by induction of labour at 39^{+0} – 39^{+6} gestational weeks
C – Children born by elective pre-labour caesarean section
O – Developmental and educational outcomes – AEDC scores
T – The nominal age at which children are measured for AEDC (foundation school year: 4-6 years)

NULL HYPOTHESIS

The population distribution of school-aged developmental and educational outcomes for children born at 39 weeks by induction of labour are the same as those born via elective caesarean section at 39 weeks.

STATISTICAL ANALYSIS

Causal research question: What is the causal effect of IOL on the AEDC-defined outcomes for children compared to those born by elective caesarean section?
Exposure: IOL at 39+0 – 39+6.
Primary outcome: A score of $<10^{th}$ centile in “2 or more” domains (Binary yes/no).
Secondary outcomes: A score in the lowest decile in each domain: 1) physical health and wellbeing, 2) social competence, 3) emotional maturity, 4) language and cognitive skills, and 5) communication skills and general knowledge.
Inclusion criteria: Children born during the study period in Victoria.
Exclusion criteria: Stillbirth, multiple pregnancy, babies with known congenital anomalies, cases of placenta praevia, placental abruption, breech birth (as these will have needed to be born by caesarean section), missing gestational age, missing mode of birth, missing birth onset, and those born before 39+0 weeks and after 40+0 weeks.

Anticipated sample size: All births in Victoria during 2005-2014 will be matched to AEDC childhood outcome data. AEDC is conducted every 3 years and as a result, our total cohort will include matched perinatal-childhood data for approximately 224,973 women. This sample size has been estimated from state-wide, published birth reports for the relevant years of the study. After exclusion of births </> 39 weeks gestation, modes of birth other than planned caesarean section and induction of labour, multiple pregnancies (~3.3%) and babies with known congenital anomalies (~3.0%) we estimate that ***our sample size will include approximately 18,000 matched mother-child data pairs.***
The baseline risk of developmental vulnerability in children born at term using AEDC assessment is 20% (4).

Causal Inference methodology

*The steps to **causal inference**:*

6. *Think carefully about the precise nature of the causal question to be addressed, usually around a currently humanly feasible intervention that we wish to apply to the whole of the target population, and compare the outcome (averaged over the whole of the target population) to the corresponding average outcome that occurs when no one in the target population gets the intervention;*
7. *Convert the causal question in Step 1 to a precise quantity to be estimated (a causal estimated), usually using the notation of the potential outcomes approach to causal inference.*
8. *At least state and, if possible, be rigorous and honest about the the assumptions (e.g. consistency, conditional exchangeability, positivity) under which the estimated in Step 2 can be identified from the data at hand;*

> 9. *Find (or develop!) estimation strategies (i.e. statistical methods) that are valid under the assumptions in Step 3. Apply these to the data in Step 3;*
> 10. *Use statistical tools to assess quantitatively the sensitivity of the results to plausible departures from the assumptions in Step 3.*

4) Develop a causal model using directed acyclic graphs (DAGs) to both identify and summarise causal pathways.

a. Potential confounders – demographic and perinatal – will be determined by the above-named authorship team (a quorum of specialists with expert knowledge):

i. Maternal age (continuous)
ii. Maternal BMI (continuous)
iii. Parity (categorical)
iv. ART (categorical)
v. Pre-existing maternal disease (categorical)
vi. Obstetric complications (categorical)
vii. SEIFA socio-economic quintile (categorical)
viii. Birthweight (continuous)
ix. Childs age at testing (binary)
x. Childs gender (binary)
xi. Year of testing
xii. Auxiliary variables for imputation

1. Marital status
2. Maternal Height
3. Language-background other than English (binary)
4. Level of maternal education (categorical)
5. Parent 2 education
6. Maternal country of birth
7. Remoteness
8. Language base other than English
9. Aboriginal/Torres Strait Islander background
10. Hospital setting (public vs private)

b. AEDC scores to be standardized by age of child at assessment before analysis

c. The list of covariates to be included in the propensity score model will be finalised after consideration of potential open backdoor causal paths and conditioning effects of collider nodes. The list above will be used as start. Distinction will be made between the standard epidemiological definition of a confounder and the causal pathway definition. **These potential confounders (PS_covariates) will then be used in the construction of a propensity score model for exposure assignment (vide infra).**

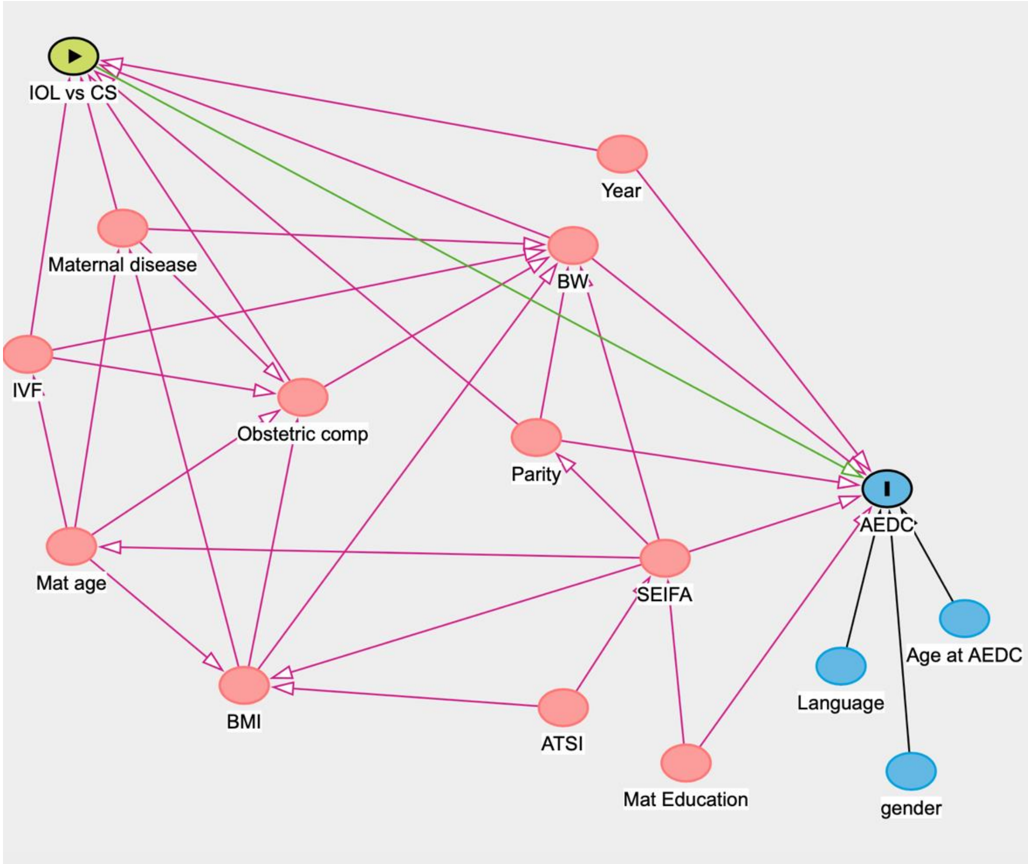
5) Handling of missing data using Multiple Imputation

a. Consideration of missing data will be confined to all the covariates identified for the construction of a (propensity score) model of the exposure status. When imputing these missing covariate values the imputation model will include all those identified for inclusion in the propensity score model as well as additional covariates* from the BOS database/AEDC databases. *These co-variates will be selected once the data are available for review.

b. Multiple imputation will be performed using chained equations, and the imputed datasets will then be used for propensity score construction. Subjects with missingness greater than 50% of PS_covariates will be excluded from analysis, this proportion will be assessed and if substantial, may result in changes to the PS_covariate list used for ongoing analysis.
c Diagnostics will then be performed to assess the quality of the imputed datasets.
d. For each of these imputed datasets Inverse Probability Weights (IPW, vide infra) will be calculated and used in weighted regression analyses. For each outcome metric, average point estimates and standard errors (95% CI) appropriately adjusted, given the imputation framework, will be obtained. The number of imputed datasets will be set at a minimum of 20 with the number increased to equal to the highest percentage of missingness in the raw PS_covariate list (ceiling of 50% by design). Details of imputation and assessment of adequacy will be formally presented.

6) Development of Inverse Probability Weights.
a) For each imputed dataset, logistic regression with robust standard errors will be performed for the binary exposure against all the PS_covariates. Inverse probability weights will be calculated for everyone given their exposure and these will be used to weight the adjusted regression model.
b) For each imputed dataset the distribution of IPW by exposure status will be examined to ensure: (i) that there is common support (overlap) of the distributions adequacy of adjustment; and (ii) that the weighting was effective in equalizing the covariate distributions in the 'pseudopopulation'. This will be performed using standardized difference (StD) constructed from the imputed mean minus the raw data mean for all PS_covariate, with an StD of less than 25% being considered adequate. Forrest plots comparing unadjusted and adjusted covariate distributions will also be performed.
c) For causal analysis the positivity assumption is met if there is overlap in the distribution of IPW between exposure arms. That is the analysis cohort must exclude those observations where IPW don't overlap. The distribution of IPW can be transformed (stabilized weights) to improve both overlap and covariate balance. These will be explored before defining the analysis weights.

7) Regression modelling
a. All causal regression modelling will be performed within the multiple imputation framework using the doubly robust regression adjustment weighted propensity score-based model.
b. The outcome metric is absolute risk difference (RD) and relative risk (RR) between IOL and elective C/S. This will be performed using the doubly robust regression model (above) with sensitivity analysis performed using estimates derived from Targeted Maximal Likelihood Estimation (TMLE). Pooled estimates will be combined across imputed datsets using Rubin's rules.
c. Software: analysis will be performed using Stata v16 (StataCorp. 2019. *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LLC. Causal inference will be performed using teffects commands (Treatment-effects estimation for observational data - see TE -teffects) run within Stata's multiple imputation suite and TMLE command in R.



Minimal sufficient adjustment sets for estimating the total effect of MOB IOL on AEDC:

- ATSI, BW, Parity, SEIFA, Year
- BMI, BW, Mat age, Parity, SEIFA, Year
- BW, IVF, Maternal disease, Obstetric comp, Parity, Year
- BW, Mat Education, Parity, SEIFA, Year

Code: dag {
bb="0,0,1,1"
"Age at AEDC" [pos="0.930,0.690"]
"IOL vs CS" [exposure,pos="0.079,0.075"]
"Mat Education" [pos="0.664,0.848"]
"Mat age" [pos="0.077,0.611"]
"Maternal disease" [pos="0.126,0.263"]
"Obstetric comp" [pos="0.298,0.448"]
AEDC [outcome,pos="0.856,0.548"]
ATSI [pos="0.547,0.787"]
BMI [pos="0.242,0.763"]
BW [pos="0.556,0.282"]
IVF [pos="0.035,0.401"]
Language [pos="0.812,0.742"]
Parity [pos="0.521,0.492"]
SEIFA [pos="0.644,0.624"]
Year [pos="0.603,0.182"]
gender [pos="0.879,0.857"]

```
"Age at AEDC" -> AEDC
"IOL vs CS" -> AEDC
"Mat Education" -> AEDC
"Mat Education" -> SEIFA
"Mat age" -> "Maternal disease"
"Mat age" -> "Obstetric comp"
"Mat age" -> BMI
"Mat age" -> IVF
"Maternal disease" -> "IOL vs CS"
"Maternal disease" -> "Obstetric comp"
"Maternal disease" -> BW
"Obstetric comp" -> "IOL vs CS"
"Obstetric comp" -> BW
ATSI -> BMI
ATSI -> SEIFA
BMI -> "Maternal disease"
BMI -> "Obstetric comp"
BMI -> BW
BW -> "IOL vs CS"
BW -> AEDC
IVF -> "IOL vs CS"
IVF -> "Obstetric comp"
IVF -> BW
Language -> AEDC
Parity -> "IOL vs CS"
Parity -> AEDC
Parity -> BW
SEIFA -> "Mat age"
SEIFA -> AEDC
SEIFA -> BMI
SEIFA -> BW
SEIFA -> Parity
Year -> "IOL vs CS"
Year -> AEDC
gender -> AEDC
}
```

PART THREE: SECONDARY ANALYSIS

Following on from the cohort developed in Part 2, we will also examine, among women who have an induction of labour at 39 weeks, what are the associations between i) instrumental vaginal birth and ii) in-labour caesarean section compared with unassisted vaginal birth and offspring school-age developmental and educational outcomes?

P –eligible Victorian induction of labour births 2005-2014

I – a) Children born by instrumental vaginal birth (forceps or vacuum) and b) in-labour caesarean section

C – Children born via unassisted vaginal birth
O – Developmental and educational outcomes – AEDC scores
T – The nominal age at which children are measured for AEDC (foundation school year: 4-6 years)

NULL HYPOTHESIS

The population distribution of school-aged developmental and educational outcomes for children born at via instrumental vaginal birth and in-labour caesarean section are the same as those born via unassisted vaginal birth

STATISTICAL ANALYSIS

Planned comparisons: unassisted vaginal birth (ref)
1. Instrumental vaginal birth
2. Emergency in-labour

Traditional regression modelling will be used to investigate the association between mode of birth and offspring AEDC outcomes.

Adjusted analysis
To adjust for potential confounding and baseline differences traditional multivariate regression modelling will be performed.

A. Potential confounders to be considered (final included covariates determined by the authorship team and informed by direct acyclic graphs)
    i. Maternal age (continuous)
    ii. Maternal BMI (continuous)
    iii. Parity (categorical)
    iv. Pre-existing maternal disease (categorical)
    v. Obstetric complications (categorical)
    vi. Indication for induction (categorical)
    vii. SEIFA socio-economic quintile (categorical)
    viii. Birthweight centile (continuous)
    ix. Year

Contingent Analyses

1. Depending on the results of the primary AEDC analysis, the same research question may be explored using NAPLAN outcome data, specifically if the primary analysis suggests that AEDC outcomes do differ by mode of birth

1. Giles G, Herkes R. The Australian Atlas of Healthcare Variation series. 2018.
2. O'Brien SM, Winter C, Burden CA, Boulvain M, Draycott TJ, Crofts JF. Pressure and traction on a model fetal head and neck associated with the use of forceps, Kiwi ventouse and the BD Odon Device in operative vaginal birth: a simulation study. BJOG. 2017;124 Suppl 4:19-25.
3. Doumouchtsis SK, Arulkumaran S. Head trauma after instrumental births. Clin Perinatol. 2008;35(1):69-83, viii.
4. Smithers LG, Searle AK, Chittleborough CR, Scheil W, Brinkman SA, Lynch JW. A whole-of-population study of term and post-term gestational age at birth and children's development. BJOG. 2015;122(10):1303-11.

1. Australian Institute of Health and Welfare. National Core Maternity Indicators., [Website]. 2019 [Available from: https://www.aihw.gov.au/reports/mothers-babies/ncmi-data-visualisations/contents/labour-and-birth-indicators/induction-of-labour.
2. Grobman WA, Rice MM, Reddy UM, Tita ATN, Silver RM, Mallett G, et al. Labor Induction versus Expectant Management in Low-Risk Nulliparous Women. N Engl J Med. 2018;379(6):513-23.
3. MacKay DF, Smith GC, Dobbie R, Pell JP. Gestational age at delivery and special educational need: retrospective cohort study of 407,503 schoolchildren. PLoS Med. 2010;7(6):e1000289.
4. Dhamrait GK, Christian H, O'Donnell M, Pereira G. Gestational age and child development at school entry. Sci Rep. 2021;11(1):14522.
5. Smithers LG, Searle AK, Chittleborough CR, Scheil W, Brinkman SA, Lynch JW. A whole-of-population study of term and post-term gestational age at birth and children's development. BJOG. 2015;122(10):1303-11.

eTable1. Primary analyses using complete (nonmissing) case data

Elective birth at 39 weeks' vs expectant management (40 – 42 weeks' gestation)				
	Unadjusted		**Adjusted**	
	Relative risk (95% CI)	Percentage risk difference (95% CI)	Relative risk (95% CI)	Percentage risk difference (95% CI)
Developmental vulnerability in 2 or more domains	1.03 (0.98, 1.10)	0.28 (-0.21, 0.76)	1.04 (0.94, 1.14)	0.31 (-0.48, 1.10)
Physical health and wellbeing	0.94 (0.88, 1.03)	-0.43 (-0.88, 0.03)	0.95 (0.86, 1.06)	-0.34 (-1.1, 0.43)
Social competence	1.05 (0.99, 1.12)	0.38 (-0.08, 0.84)	1.08 (0.97, 1.21)	0.64 (-0.13, 1.42)
Emotional maturity	1.08 (1.02, 1.16)	0.58 (0.13, 1.02)	1.01 (0.91, 1.13)	0.11 (-0.71, 0.92)
Language and cognitive skills (school-based)	1.01 (0.94, 1.09)	0.07 (-0.32, 0.46)	1.03 (0.89, 1.18)	0.13 (-0.57, 0.83)
Communication skills and general knowledge	1.01 (0.94, 1.08)	0.03 (-0.38, 0.44)	1.05 (0.92, 1.19)	0.25 (-0.43, 0.92)
Elective birth at 39 weeks'; Induction of labor vs planned cesarean section				
Developmental vulnerability in 2 or more domains	1.06 (0.95, 1.18)	0.46 (-0.41, 1.34)	0.97 (0.82, 1.14)	-0.28 (-1.68, 1.12)
Physical health and wellbeing	1.08 (0.96, 1.21)	0.55 (-0.28, 1.38)	1.14 (0.97, 1.34)	0.95 (-0.19, 2.01)
Social competence	1.04 (0.93, 1.17)	0.30 (-0.54, 1.14)	0.91 (0.76, 1.09)	-0.72 (-2.05, 0.62)
Emotional maturity	1.15 (1.02., 1.30)	0.97 (0.16, 1.77)	1.02 (0.84, 1.23)	0.12 (-1.19, 1.44)
Language and cognitive skills (school-based)	1.10 (0.96, 1.26)	0.48 (-0.22, 1.18)	1.05 (0.86, 1.29)	0.26 (-0.77, 1.29)
Communication skills and general knowledge	1.00 (0.88, 1.14)	0.007 (-0.73 to. 0.75)	0.82 (0.66, 1.00)	-1.15 (-2.22, 0.00)

eTable2. Primary analyses using alternative modelling techniques

Elective birth at 39 weeks' vs expectant management (40 – 42 weeks' gestation)		
	Relative risk (95% CI)	Percentage risk difference (95% CI)
Regression adjustment modelling	1.04 (0.96, 1.13)	0.25 (-0.33, 0.83)
Inverse probability weighted	1.03 (0.96, 1.12)	0.28 (-0.38, 0.94)
Inverse probability weighted regression adjusted	1.03 (0.96, 1.12)	0.28 (-0.38, 0.94)
Targeted Maximum Likelihood Estimation	1.05 (0.99, 1.13)	0.50 (-0.002, 0.99)
Elective birth at 39 weeks'; Induction of labor vs planned cesarean section		
Regression adjustment modelling	0.92 (0.81, 1.05)	-0.23 (-1.17, 0.71)
Inverse probability weighted	0.96 (0.82, 1.13)	-0.35 (-1.74, 1.04)
Inverse probability weighted regression adjusted	0.96 (0.82, 1.12)	-0.34 (-1.71, 1.03)
Targeted Maximum Likelihood Estimation	0.96 (0.85, 1.06)	-0.38 (-1.27, 0.51)

eTable3. Secondary analysis of mode of birth at 39 weeks' gestation and developmental vulnerability using complete case (nonmissing) data

	Complete case – unadjusted		Complete case – adjusted	
	Relative risk (95% CI)	Percentage risk difference (95% CI)	Relative risk (95% CI)	Percentage risk difference (95% CI)
Developmental vulnerability in 2 or more domains				
Instrumental vaginal birth	0.79 (0.64, 0.98)	-1.80 (-3.33, -0.26)	0.93 (0.61, 1.43)	-0.73 (-1.91, 3.36)
Unplanned cesarean section	1.04 (0.86, 1.26)	0.38 (-1.36, 2.11)	1.15 (0.67, 1.97)	1.36 (-4.06, 6.77)
Physical health and wellbeing				
Instrumental vaginal birth	0.69 (0.55, 0.87)	-2.55 (-3.99, -1.11)	0.79 (0.51, 1.23)	-0.60 (-2.97, 1.77)
Unplanned cesarean section	0.82 (0.66, 1.02)	-1.51 (-3.07, 0.04)	0.73 (0.38, 1.38)	-1.67 (-5.45, 2.11)
Social competence				
Instrumental vaginal birth	0.87 (0.70, 1.08)	-1.00 (-2.49, 0.50)	0.81 (0.49, 1.36)	-0.36 (-2.69, 1.98)
Unplanned cesarean section	1.03 (0.84, 1.27)	0.26 (-1.37, 1.89)	0.76 (0.40, 1.51)	-1.18 (-4.86, 2.50)
Emotional maturity				
Instrumental vaginal birth	0.93 (0.75, 1.16)	-0.47 (-1.95, 1.01)	0.91 (0.6, 1.38)	0.18 (-2.31, 2.68)
Unplanned cesarean section	1.21 (0.99, 1.48)	1.52 (-0.15, 3.19)	1.33 (0.76, 2.33)	2.67 (-1.65, 6.99)
Language and cognitive skills (school-based)				
Instrumental vaginal birth	0.65 (0.49, 0.86)	-2.06 (-3.27, -0.86)	0.90 (0.42, 1.90)	0.67 (-1.53, 2.87)
Unplanned cesarean section	0.81 (0.62, 1.05)	-1.14 (-2.46, 0.18)	1.25 (0.62, 2.53)	2.87 (-0.79, 6.53)
Communication skills and general knowledge				
Instrumental vaginal birth	0.75 (0.57, 0.97)	-1.53 (-2.80, -0.26)	1.14 (0.71, 1.85)	-0.69 (-1.47, 2.85)
Unplanned cesarean section	1.04 (0.82, 1.32)	0.26 (-1.20, 1.72)	1.60 (0.78, 3.27)	3.15 (-0.55, 6.86)