Examining the effects of pre-pregnancy weight and gestational weight gain on allergic disease development in offspring: a protocol for a population-based study using health administrative databases in Ontario, Canada

Sebastian A Srugo, Laura Gaudet, Daniel Corsi, Romina Fakhraei, Yanfang Guo, Deshayne B Fell

ABSTRACT

Introduction Over the last 20 years, excess maternal pre-pregnancy weight (overweight and obesity) and gestational weight gain have become the most common morbidities in pregnancy. These morbidities may pose a threat to fetal immunological development through associated metabolic dysfunction and inflammation and, as such, may partly explain the concurrent rise of paediatric allergic disease. We will examine the effect of maternal pre-pregnancy weight and gestational weight gain during pregnancy on the incidence of allergic diseases among offspring in Canada’s most populous province.

Methods and analysis We will conduct a retrospective, population-based cohort study of all singleton live births to residents of Ontario, Canada in 2012–2013 and 2013–2014. The study population will be defined using maternal-newborn records from the provincial birth registry, which captures information on maternal pre-pregnancy weight and gestational weight gain. The cohort will be linked with provincial health administrative databases, allowing for follow-up of neonates through early childhood until 2019 (5–7 years of age). Allergic disease development (asthma, rhinitis, atopic dermatitis and anaphylaxis) will be ascertained using diagnostic codes from healthcare encounters. Potential confounders have been identified a priori through a directed acyclic graph. Cox proportional hazards regression models will be employed to assess the associations between maternal pre-pregnancy weight, gestational weight gain and incident paediatric allergic disease. Several preplanned sensitivity analyses will be conducted, including a probabilistic bias analysis of outcome misclassification.

Ethics and dissemination Ethics approval was obtained from the Research Ethics Board of the Children’s Hospital of Eastern Ontario and the ICES Privacy Office. Findings will be disseminated in scientific conference presentations and peer-reviewed publications.

INTRODUCTION

In 1997, the WHO declared obesity a global epidemic. Since then, excess maternal pre-pregnancy weight (body mass index (BMI) ≥25 kg/m²) and gestational weight gain (GWG) have become the most common morbidities in pregnancy. Several studies have suggested these morbidities increase the risk of asthma among offspring, though their impact on other common allergic diseases is unclear. Further, most previous studies have been limited by small sample sizes and self-reported measures of the exposure or outcome.

What is already known on this topic?

► Excess maternal pre-pregnancy weight and gestational weight gain are common morbidities in pregnancy and may impact fetal immunological development.
► Several studies have suggested these morbidities increase the risk of asthma among offspring, though their impact on other common allergic diseases is unclear.

What this study hopes to add?

► This protocol details a large and extensive study to evaluate the impact of maternal pre-pregnancy weight and gestational weight gain on several paediatric allergic diseases.
► This study will employ a population-based approach and objective measures, increasing the generalisability and validity of the study findings.
► To ensure the robustness of results, several prespecified sensitivity analyses will be undertaken to measure the impact of selection biases, residual confounding and misclassification.
with metabolic dysfunction and inflammation, which pose a threat to the homeostasis of the fetal compartment and could result in impaired fetal immunological development and increased chronic disease susceptibility in later life. Therefore, the rise of excess weight and GWG during pregnancy may partially explain the concurrent rise in paediatric allergic disease, which is now one of the most common and earliest onset morbidities in childhood. One 2016 study conducted in Ontario found that 44% of toddlers suffered from food allergies, environmental allergies (wheezing, asthma, eczema or rhinitis), atopy or a combination thereof. Though one cross-sectional and several longitudinal studies have assessed the early-life programming of allergic diseases by maternal weight or GWG, they have been limited by small sample sizes (<1000) and self-reported measures of the exposure or outcome, which have reduced their generalisability and validity. Further, most have focused on asthma alone, none have examined anaphylaxis and none have calculated risk by the length of follow-up time, opting instead for aggregate rates (log-linear model), odds (logistic model) or prevalence ratios (log-binomial model). As well, none of these studies have been undertaken in Canada, which has a unique multicultural population. This proposed research will attempt to address these evidence gaps by carrying out a large population-based study in Canada’s most populous province.

**Aims**

We will assess the impact of maternal pre-pregnancy BMI and GWG on the incidence of childhood allergic diseases (asthma, rhinitis, atopic dermatitis and anaphylaxis). Based on previous literature, we hypothesise that maternal overweight, obesity and excess GWG will all increase the risk of allergic diseases in offspring, with GWG incurring the greatest risk. This work may help inform public health initiatives for high-risk groups and increase early prevention, identification and treatment of early childhood morbidity.

**METHODS AND ANALYSIS**

**Design and setting**

We will conduct a population-based retrospective cohort study of all singleton live births to residents of Ontario, Canada between 1 April 2012 and 31 March 2014, with follow-up of children on 31 March 2019 (figure 1). The study will be conducted using multiple linked province-wide health databases within the secure network environment of ICES (https://www.ices.on.ca/). This protocol was developed following the Reporting of Studies Conducted Using Observational Routinely Collected Health Data guidelines.

**Data sources and population**

The birth cohort will be defined using maternal-newborn records from the provincial birth registry, the Better Outcomes Registry & Network (BORN) Information System (BIS). The BIS captures all hospital births ≥2500 g or ≥20 weeks of gestation in Ontario, as well as detailed clinical and demographic information about the birth and the mother. Deterministic and probabilistic methods have been used to link the BIS database to health administrative databases at ICES, enabling follow-up of offspring health outcomes into early childhood (5–7 years of age). All residents of Ontario are covered by publicly funded universal healthcare and are eligible to access the services captured in the health administrative databases. The health administrative databases at ICES to be used in this study include: the Canadian Institute for Health Information (CIHI) Discharge Abstract Database (DAD); the CIHI National Ambulatory Care Reporting System (NACRS); the Ontario Health Insurance Plan (OHIP) database; the Ontario Asthma Surveillance Information System (ASTHMA); the Ontario Drug Benefit (ODB) database; the Ontario portion of the federal Immigration, Refugees and Citizenship Canada Permanent Resident Database; and the Ontario Marginalisation Index. All data sources and reasons for use are displayed in figure 2. The study population will be restricted by sample exclusions (non-Ontario resident; stillbirth; twin or higher order multiple birth; mother’s age outside probable range (ie, <12 or >50 years); and infant died on date of birth) and administrative exclusions (duplicate records in BORN; invalid linkages and linkage warnings; invalid infant and mother identifiers; mothers without continuous OHIP eligibility throughout pregnancy; and infants who did not have records of healthcare eligibility within 90 days of birth). Figure 3 displays temporal aspects of the study design relating to eligibility, variable measurement and

---

**Figure 1** Length of study period for two two-year birth cohorts. Each birth cohort starts on April 1st and ends March 31st of the following calendar year.
follow-up; this figure was previously proposed by Schnee-weiss and colleagues.21

**Exposures**
The primary exposures of interest will be maternal pre-pregnancy BMI and GWG.

**Ascertainment**
Pre-pregnancy BMI will be calculated by dividing pre-pregnancy weight (kg) by height (m), squared. For approximately 70% of participants, data on maternal weight will be available from the first prenatal screening visit at around 12 weeks of gestation (available from the Prenatal Screening Ontario (PSO) encounter within the BIS database), where it is likely to have been objectively measured. From this, pre-pregnancy weight will be estimated by subtracting the typical first trimester weight gain (1.64 kg),22 as done previously.23 Ideally, pre-pregnancy measurements, which are rare in health administrative database studies, would be available; still, it has been recently noted that weight at first prenatal screening offers a valid estimate of pre-pregnancy weight after adjustment for first trimester weight gain.24 For subjects with no record in the PSO database (due to no prenatal screening), maternal pre-pregnancy weight will be captured from the Labour & Birth encounter within the BIS database, where the data may be based on either self-report or objective measurement. GWG will be calculated by subtracting maternal pre-pregnancy weight from weight recorded at delivery.
Figure 3
Study design and cohort structure. Adapted from Schneeweiss et al, 2019. Abbreviations: BMI, body mass index; GWG, gestational weight gain. Numbers in square brackets represent date ranges relative to the date of birth (day 0).

For the purposes of illustrating the temporal relationship between maternal and infant study variables, the date of the LMP is shown as day -280 (i.e. 40 completed weeks of gestation). In the study, this will depend on the actual length of gestation.

Sample exclusions include: non-Ontario resident; stillbirth; twin or higher-order multiple birth; mother’s age outside probable range (i.e. <12 years or >50 years); and infant died on date of birth. Administrative exclusions include: duplicate records in BORN; invalid linkages and linkage warnings; invalid infant and mother identifiers; mothers without continuous OHIP eligibility throughout pregnancy; and infants who did not have records of health care eligibility within 90 days of birth.

Pre-existing maternal conditions are ascertained on delivery record but reflect previous history.

Only mothers who were continuously eligible to receive health care (OHIP) during the one-year period will be included.
Pre-pregnancy BMI will be classified according to WHO categories: underweight (<18.5 kg/m²), normal weight (18.5–24.9 kg/m²), overweight (25.0–29.9 kg/m²) or obese (≥30.0 kg/m²). The Institute of Medicine (IOM) recommendations for weight gain during pregnancy suggest total weight gain for singleton pregnancies should not exceed 12.5–18 kg, 11.5–16 kg, 7–11.5 kg or 5–9 kg based on pre-pregnancy BMI category, respectively. However, these recommendations do not account for gestational length. To do so, we will follow previously published methods to establish expected GWG based on IOM recommendations for weight gain by trimester and pre-pregnancy BMI, and calculate the ratio of actual to expected GWG. We will then employ the recommended ratio ranges calculated by Guo and colleagues to classify GWG into three categories: inadequate, adequate or excess weight gain. These ratio ranges were determined by dividing the lower and upper limits of the IOM GWG recommendations by the expected GWG within each pre-pregnancy BMI category. Pre-pregnancy BMI and GWG (calculated as the difference between observed and expected weight gain, based on the above methods) will also be examined as continuous variables.

Outcomes
Asthma, rhinitis, atopic dermatitis and anaphylaxis will be examined separately during follow-up based on International Classification of Diseases (ICD)-9 and ICD-10 diagnostic codes in health administrative data sets (DAD, NACRS and OHIP). ICD codes and algorithms can be found in online supplemental file A. We will employ a time-to-first-event analysis for all outcomes, in which the follow-up period will begin at birth and continue until either the outcome occurs, the infant dies or becomes ineligible to receive healthcare services (with a 90-day grace period) or until the end of the study period, whichever comes first. Follow-up time (person-days) will be calculated for each event separately.

Asthma
Using a validated algorithm, the ASTHMA database will identify infants with asthma diagnoses. The algorithm requires ≥1 hospitalisation and/or ≥2 ambulatory care visits for asthma within 2 years, where the earliest claim is deemed the date of diagnosis. The algorithm was first validated among children in 2006 (ages 1–8; sensitivity: 91%, specificity: 83%). and was recently revalidated in 2019 among younger children (ages 1–5; sensitivity: 81%, specificity: 90%). Since the accuracy of asthma diagnoses among young infants is debated, records fulfilling the algorithm for asthma diagnosis before the age of 6 months will be disregarded unless a further hospitalisation or ambulatory care visit for asthma is detected after 12 months of age.

Rhinitis and atopic dermatitis
Rhinitis and atopic dermatitis cases will be captured by an algorithm from Hill and colleagues, which requires ≥2 care visits occurring 26 months apart and where the date of diagnosis is the first visit. These algorithms displayed positive predictive values of 100% and 90% based on chart review, respectively (no sensitivity or specificity measures were published).

Anaphylaxis
We will employ criteria A and B from the algorithm by Walsh and colleagues to detect anaphylaxis in inpatient, emergency department and outpatient encounters using ICD codes for diagnoses and symptoms. In a validation study, this algorithm performed better than those of previous studies, with a sensitivity of 91% and positive predictive value of 67%. Some modifications were made to the algorithm for our study, including the removal of the three procedure/intervention codes which are not available in our outpatient data set and the inclusion of an ICD code for anaphylactic shock due to food reaction due to its high prevalence in childhood.

Covariates and confounders
The following covariates and potential confounders will be examined: maternal characteristics (maternal country of birth (for recent immigrants, otherwise Canada), maternal age at birth, maternal smoking, parity, maternal asthma and other pre-existing medical conditions, and use of drugs, alcohol or medications during pregnancy); neighbourhood characteristics using residential postal codes linked to census data (neighbourhood marginalisation index and rural residence); and birth or infant characteristics (mode of delivery, complications of pregnancy, season of birth, sex of infant, gestational age, birth weight and exclusive breast feeding at time of hospital discharge). Causal assumptions are presented in a directed acyclic graph (DAG; figure 4), adapted from two published DAGs and supplemented for our study through a review of the literature. Before statistical analysis, 10 multiple imputation data sets will be generated to deal with missing values. The results from the imputed data sets will be combined to produce final estimates, following best practices.

Statistical analyses
Sample characteristics will be presented as frequencies and proportions for categorical variables and means or medians for continuous variables (with measures of spread, such as SD or IQR), overall and stratified by exposure group to present potential confounding. Between-group differences will be assessed by standardised differences, which denote differences in units of SD. Standardised differences greater than 10% indicate an imbalance between the exposure groups. Crude incidence rates (per 100 000 person-days of follow-up) for each study outcome will be calculated for each exposure group and graphically presented using cumulative incidence curves.
For the primary analyses, Cox proportional hazards regression models will be employed to produce HRs with 95% CIs. Crude and adjusted analyses (using multivariable analysis methods) will assess differences in paediatric allergic disease incidence: (1) between infants born to underweight, overweight or obese mothers compared with those born to mothers with normal pre-pregnancy weight; and (2) between infants born to mothers who fell below or exceeded recommended GWG guidelines compared with those who met the GWG recommendation. Models will be adjusted for maternal age, maternal birth country, maternal pre-pregnancy health conditions, maternal medication and substance use during pregnancy, neighbourhood characteristics and parity—the minimal sufficient adjustment set identified in the DAG (Figure 4). Cox models will account for repeated pregnancies during the 2012–2014 period by adjusting the variance through generalised estimating equation methods.36 The following effect measure modifiers have been selected a priori: pre-pregnancy BMI (when GWG is the exposure), GWG (when pre-pregnancy BMI is the exposure) and infant sex. If estimates between strata are qualitatively different, we will introduce an interaction term or report strata-specific estimates based on published guidelines.37 The proportional hazards assumption of the Cox models will be assessed through Schoenfeld residual plots and Wald tests for interaction between exposure status and time.

Figure 4. Directed acyclic graph (DAG) of causal assumptions. Adapted from Dumas et al., 2016 and Harskamp-van Ginkel et al., 2015. Abbreviations: BMI, body mass index; GWG, gestational weight gain.
In secondary analyses, continuous measures of the exposures will be used and restricted cubic splines will be added to the Cox proportional hazards models to account for non-linear associations between exposures and outcomes. All analyses will be conducted using SAS V.9.4 (SAS Institute).

Preplanned sensitivity analyses
Assessment of potential selection biases and residual confounding
1. A complete case analysis will be conducted.
2. To understand the impact of exclusion criteria on selection bias, sociodemographic information of response, complete-case and imputed samples will be compared.
3. To account for potential residual confounding by maternal propensity to access healthcare services, we will assess typical maternal healthcare utilisation through outpatient visits in the OHIP database during a 1-year look-back period before pregnancy and adjust for this in our models. Only mothers who were continuously eligible for OHIP during the look-back period will be included.

Assessment of potential exposure misclassification
4. To ensure that self-reported maternal weight does not bias our results, a subgroup analysis will be undertaken among the subset of participants with pre-pregnancy weight recorded in the prenatal encounter of the BIS database.

Assessment of potential outcome misclassification
5. Researchers from Denmark created validated algorithms that employ hospital diagnoses and dispensed prescription medications to identify cases of rhinitis (sensitivity: 84%, specificity: 82%) and atopic dermatitis (sensitivity: 74%, specificity: 73%) in childhood compared with parental reports of physicians diagnoses. In our study, data on prescription medications dispensed can only be ascertained for the subset of participants who were covered by the provincial drug funding system (ODB). Residents of Ontario qualify for ODB if they are either ≥65 years old or are enrolled in one of the following provincial programmes: home care and community services, employment assistance, assistance for persons with disabilities, or assistance with prescription drug costs for those in financial need. Starting on 1 January 2018, ODB was extended to any resident of Ontario covered by OHIP+: Children and Youth Pharmacare (≥24 years old). We will run a sensitivity analysis using these derived algorithms for rhinitis and atopic dermatitis in the subgroup of children whose families qualified for ODB (see online supplemental file A for algorithms and codes).
6. We believe that the algorithms employed may result in non-differential outcome misclassification, which usually produces bias towards the null or, in some circumstances, no bias. A probabilistic bias analysis will be conducted to evaluate the influence of this misclassification, following published methods.

Patient and public involvement
Patients and the public were not involved in the development of the protocol.

ETHICS AND DISSEMINATION
Approval was obtained from the Research Ethics Board of the Children’s Hospital of Eastern Ontario (protocol number 19/14PE) and the ICES Privacy Office (protocol number 2020 0901 237 000). Individual patient consent is not required for secondary analyses of routinely collected data, pursuant to privacy legislation in Ontario (Personal Health Information Protection Act, 2004). Further, under the same legislation, BORN and CIHI can collect, use and disclose personal health information without consent. Research finding will be disseminated to public health audiences through presentations at scientific conferences and publications in peer-reviewed journals.

Author affiliations
1School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada
2Department of Obstetrics and Gynecology, Queen’s University, Kingston, Ontario, Canada
3Department of Obstetrics and Gynecology, Kingston Health Sciences Centre, Kingston, Ontario, Canada
4OMNI Research Group, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada
5Children’s Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada
6Better Outcomes Registry & Network, Ottawa, Ontario, Canada
7ICES, Ottawa, Ontario, Canada

Twitter Sebastian A Srugo @sebsrugo

Contributors Conceptualisation: SAS, DBF, RF, LG. Funding acquisition: SAS, DBF, LG. Methodology: SAS, DBF, DC, YG, RF, LG. Supervision: DBF, LG. Writing, original draft: SAS. Writing, review and editing: SAS, DBF, RF, DC, YG, LG.

Funding This work was supported by a Canadian Institutes of Health Research (CIHR) Foundation Grant (grant number FDN-148438) to LG, a CIHR Frederick Banting and Charles Best Canada Graduate Scholarship–Master’s Award and an Ontario Graduate Scholarship to SAS, and by ICES, which is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care (MOHLTC). This study is based in part on data provided by Better Outcomes Registry & Network (‘BORN’), a part of the Children’s Hospital of Eastern Ontario.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Research Ethics Board of the Children’s Hospital of Eastern Ontario (protocol number 19/14PE) and the ICES Privacy Office (protocol number 2020 0901 237 000).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The data set from this study is held securely in coded form at ICES. While data sharing agreements prohibit ICES from making the data set publicly available, access may be granted to those who meet prespecified criteria for confidential access, available at www.ices.on.ca/DAS. The full data set creation plan and underlying analytic code are available from the authors upon request, understanding that the programs may rely upon coding templates or macros that are unique to ICES.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and
responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iD
Sebastian A Srugo http://orcid.org/0000-0001-9117-103X

REFERENCES