Observational Studies Volume (2016) 90-118 Submitted 4/16; Published 12/16

Assessing the dose-response relationship between maternal use of inhaled corticosteroids therapy and birth weight: a generalized propensity score approach

Mariia Samoilenko **mariia.samoilenko@uqam.ca**
Department of Mathematics, Université du Québec à Montréal
Research Center, Hôpital du Sacré-Coeur de Montréal
Montréal, Québec, Canada

Lucie Blais **lucie.blais@umontreal.ca**
Faculty of Pharmacy, Université de Montréal, Montréal
Research Center, Hôpital du Sacré-Coeur de Montréal, Montréal
Centre de recherche clinique Étienne-Le Bel, Centre hospitalier universitaire de Sherbrooke, Sherbrooke
Québec, Canada

Benoît Cossette **bcossette.chus@ssss.gouv.qc.ca**
Faculty of Pharmacy, Université de Montréal, Montréal
Pharmacy Department, Centre hospitalier universitaire de Sherbrooke, Sherbrooke
Québec, Canada

Amélie Forget **amelie.forget@umontreal.ca**
Faculty of Pharmacy, Université de Montréal
Research Center, Hôpital du Sacré-Coeur de Montréal
Montréal, Québec, Canada

Geneviève Lefebvre **lefebvre.gen@uqam.ca**
Department of Mathematics, Université du Québec à Montréal
Faculty of Pharmacy, Université de Montréal
Montréal, Québec, Canada

Abstract

Purpose: Inhaled corticosteroids (ICS) are the first-line controller therapy for asthma. The objective was to assess the impact of different ICS doses during pregnancy on birth weight (BW) using generalized propensity scores (GPS).
Methods: A cohort of 7374 pregnancies from 6197 asthmatic women giving birth in Quebec (Canada) in 1998-2008 was constructed. The average treatment effects (ATE) of ICS daily doses (0, >0-125, >125-250, >250 μg/day) during pregnancy on BW were estimated using multilevel GPS and a conventional multivariable approach. Additional analyses were done to evaluate the robustness of the results.
Results: Using GPS, we found no significant associations between ICS doses and BW (ATE for >0-125 vs 0 μg/day: 27.62 g, 95% confidence interval (CI): -8.68, 64.10; ATE for >125-250 vs 0 μg/day: 17.07 g, 95% CI: -55.85, 92.16; ATE for >250 vs 0 μg/day: -37.83 g, 95% CI: -117.74, 41.53). Similar results were obtained using the multivariable approach.
Conclusions: While, in our primary analyses, no significant differences were found between the BW of babies exposed to the higher ICS doses, as opposed to no use of ICS, our

sensitivity analyses, which adjusted for gestational age in the models, suggest the possibility of a small detrimental effect of the higher ICS doses on BW.

Keywords: Asthma, average treatment effect, generalized propensity scores, inhaled corticosteroids, mediator, pregnancy.

1. Introduction

The use of inhaled corticosteroids (ICS) by pregnant asthmatic women is generally regarded as safe with respect to birth weight (BW) (Gregersen and Ulrik, 2013; British Thoracic Society and Scottish Intercollegiate Guidelines Network, 2014). However, knowledge about the potential differential effects of ICS-dose categories on BW is scarce, mainly relying on three studies for which ICS doses taken during pregnancy were considered. Namazy et al. (2004) reported no evidence of a significant linear trend in mean BW across quartiles of ICS dose. Similarly, Bakhireva et al. (2005) found no evidence of a differential effect across quartiles of ICS dose on mean BW. In particular, they showed no reduced BW for the highest quartile of beclomethasone equivalent doses. More recently, Cossette et al. (2013) found no association between the lowest ICS doses and low BW (i.e., BW $<$2500 g), where the reference group was composed of babies born to asthmatic mothers unexposed to ICS. A non-significant increasing trend in the prevalence of low BW was nonetheless observed for ICS doses $>$125 μg/day (fluticasone-propionate equivalent), leaving doubts concerning the safety of higher ICS doses (Cossette et al., 2013). Because ICS are the first-line treatment for asthma during pregnancy (NAEPP expert panel report, 2005) and BW is one of the most important factors that affect infant health and mortality (Wilcox, 2001), it is worthwhile to further investigate this issue.

In the study of the relationship between ICS-dose categories and BW, it is a concern that pregnant women receiving low-dose and high-dose ICS may not be intrinsically comparable. For instance, high-dose users are likely to have more severe asthma than low-dose users, and severe asthma may lead to worse outcomes by itself (Cossette et al., 2013). While Cossette et al. (2013) adjusted for several covariates related to asthma severity and control in their study of the association between ICS doses and low BW, the modeling assumptions underlying the conventional multivariable approach they took to analyze the data might have led to residual confounding by indication. Indeed, the conventional approach supposes one's ability to correctly model the outcome given exposure and confounders, which can be difficult to achieve in practice (Brookhart et al., 2013). Alternatively, propensity score approaches, relying on the specification of the treatment allocation mechanism, could have been used.

One argument in favor of propensity score-based approaches is to better characterize the treated and untreated subjects and identify those with similar characteristics (Austin, 2011a). Another interesting feature of propensity scores is that the adequacy of the treatment model fitted to data can be determined by assessing the balance of the potential confounding covariates between the treatment groups after adjusting for the estimated propensity scores (Shah et al., 2005; Imai and Ratkovic, 2014). Generalized propensity scores (GPS) are an extension of standard propensity scores for multilevel or continuous treatments (Imbens, 2000; Hirano and Imbens, 2004). They possess the same attractive properties as

standard propensity scores for bias reduction, but their implementation is recognized to be more challenging (Spreeuwenberg et al., 2010).

Our study is a follow-up of the study by Cossette et al. (2013) and aims to evaluate the safety of three ICS-dose categories during pregnancy with respect to BW using a large cohort of asthmatic women who gave birth between 1998 and 2008 in Quebec, Canada. We implemented a GPS approach for categorical treatments (herein referred to as a multilevel GPS approach) to estimate the average treatment effects (ATE) associated with two low ICS-dose categories and one moderate-to-high ICS-dose category. Estimates obtained with GPS were also compared to those obtained from adjusted linear regression. Indeed, some authors have recently called for exploring and reporting the sensitivity of the results to changes in the statistical models (e.g., conventional versus propensity score approach for effect estimation) (Brookhart et al., 2013; Guo and Fraser, 2014). Our study thus follows this analytical strategy while providing additional guidelines as to how the GPS approach can be implemented in practice. To our knowledge, our study is the first that utilizes GPS for assessing the impact of asthma medication on perinatal outcomes. Moreover, our GPS analyses are performed on clustered data, thus departing from standard implementations of this technique.

2. Methods

2.1 Data source

Data on medication prescriptions filled in community pharmacies, outpatient medical visits, emergency-department visits, medical procedures, and hospitalizations were retrieved from two administrative databases in Quebec: the Régie de l'assurance maladie du Québec (RAMQ) and the Maintenance et exploitation des données pour l'étude de la clientèle hospitalière (MED-ECHO) databases (see Cossette et al. (2013) for additional information).

2.2 Study design

Our cohort was formed by applying the same inclusion and exclusion criteria as those used by Cossette et al. (2013) to define their cohort. The inclusion criteria were: singleton delivery between 1998 and 2008; women aged $\leq$ 45 years at beginning of pregnancy, with $\geq$ 1 diagnosis of asthma, $\geq$1 prescription for an asthma medication filled in the year before or during pregnancy, and covered by the RAMQ drug-insurance plan for at least 1 year before and throughout pregnancy. Exclusion criteria were: use of theophylline, cromoglycate, nedocromil, ketotifen, and long-acting beta2-agonist (LABA) use without ICS during pregnancy. To follow Cossette et al. (2013), we kept only the two most recent pregnancies if a woman had several pregnancies during follow-up. To comply with the definition of stillbirth (Breton et al., 2009), we also excluded two pregnancies resulting in dead newborn with BW $<$500 g. Hence, our cohort comprises n=7374 pregnancies from 6197 asthmatic women.

2.3 Perinatal outcome

The perinatal outcome of interest is BW measured on a continuous scale (grams).

2.4 Inhaled corticosteroid exposure during pregnancy

Average ICS daily dose (in fluticasone-propionate equivalent) was measured with an algorithm based on prescription renewals that was used in prior studies (Martel et al., 2005; Blais et al., 2007; Cossette et al., 2013). Treatment with ICS was defined in four categories: no use, >0-125 μg/day (likely to represent sporadic use or very low doses), >125-250 μg/day (low doses as recommended by the Global Initiative for Asthma (2016) guidelines), and >250 μg/day (moderate-to-high doses).

2.5 Potential confounding covariates

Our list of potential confounding covariates contains 27 variables (see first column of Table A1 in Appendix A): 26 variables that were identified as risk factors for the perinatal outcomes studied by Cossette et al. (2013) and baby's sex, which is as well-known predictor of BW (Kramer et al., 2001; Sheiner, 2007; Aibar et al., 2012). As opposed to Cossette et al. (2013), cyanotic congenital heart disease was not considered in our analyses because of its small prevalence within our cohort (25 pregnancies) and the empirical positivity problem for ICS-dose category >125-250 μg/day (0 pregnancies).

2.6 Generalized propensity scores

We refer the reader to Appendix B for a theoretical and methodological review of multilevel GPS; specific details concerning the implementation of GPS on our cohort are provided in Section 2.7.1.

2.7 Statistical analyses

Basic statistics were first obtained to report on the characteristics of the pregnancies and describe BW as a function of ICS-dose categories.

2.7.1 Primary analyses

We implemented a multilevel GPS approach for assessing the dose-response relationship between ICS and BW taking into account the full set of 27 potential confounding covariates.

We first fitted a multinomial logit model using all available potential confounding covariates to obtain the GPS associated to each ICS-dose category for each pregnancy in the cohort. This multinomial model was selected on the basis of goodness-of-fit measures (Akaike information criterion, Bayesian information criterion) over two other multinomial models (ordered probit and logit).

We verified the assumptions underlying the GPS approach prior to estimating the effect of ICS on BW. First, we checked the positivity by assessing the overlap in the GPS values for the four ICS-dose categories. In particular, we obtained, for each dose category, a box-plot of the GPS for the pregnancies exposed under this category and a box-plot of the GPS for all other pregnancies. As implied from Imbens (2000) and Hirano and Imbens (2004), the GPS associated to treatment level $T = j$ is a balancing score between covariates and the corresponding treatment indicator variable $D(j)$ where $D(j) = 1$ if $T = j$ and $D(j) = 0$ otherwise (see Appendix B for more details regarding the balance property of GPS; in our study, treatment levels $j = 1, 2, 3, 4$ correspond, respectively, to ICS-dose categories: 0, >0-

125, >125-250, >250 μg/day). Therefore, we also verified the balance between each covariate and each treatment indicator $D(j)$ $(j = 1, 2, 3, 4)$ within the strata of percentiles of GPS that were used for the construction of the GPS-based outcome regression models. More precisely, for each covariate, the average standardized differences (Austin, 2011a) between the groups defined by the values of $D(j)$ $(j = 1, 2, 3, 4)$ were calculated over all GPS strata. These results were compared to the initial balance diagnostic performed by calculating (for each covariate) the standardized differences between the pregnancies exposed and unexposed to $T = j$ $(j = 1, 2, 3, 4)$.

For each ICS-dose category separately, and using only pregnancies actually exposed to a given category, we then fitted a linear regression model for BW using percentile-based GPS strata as independent variables (GPS-based outcome regression model). Five and ten strata are the most commonly used numbers of strata in stratified propensity scores analyses (Wang et al., 2001; Kurth et al., 2006; Austin, 2011a,b; Brookhart et al., 2013; Ali, 2014). It has also been argued that the number of propensity scores strata should depend on sample size, and that with a large data set, it might be desirable to form more than five strata (Huppler Hullsiek and Louis, 2002). Therefore, we used ten GPS strata for the 0 and >0-125 μg/day ICS categories and five strata for the >125-250 and >250 μg/day categories due to smaller number of pregnancies within each of those two strata. For each baby in the cohort, the GPS-based outcome regression models were then used to predict their four conditional counterfactual BW given GPS values (Feng et al., 2012). The cohort average of these BW values was computed for each ICS-dose category (Feng et al., 2012). Finally, to compute the three ATE of interest, the BW average associated to each ICS-dose category >0 μg/day was subtracted to the average associated with no use of ICS (i.e., >0-125 versus 0 μg/day, >125-250 versus 0 μg/day, >250 versus 0 μg/day).

We also estimated the crude and regression-based adjusted (that is, conventional multivariable approach) ATE. The regression-based adjusted estimates were obtained by fitting a multiple linear regression model for BW as a function of ICS exposure and potential confounding covariates (as main effect terms)

$$E[Y|D(2), D(3), D(4), \boldsymbol{C}] = \alpha_0 + \alpha_2 D(2) + \alpha_3 D(3) + \alpha_4 D(4) + \boldsymbol{\alpha}_5' \boldsymbol{C} \tag{1}$$

where $D(j)$ $(j = 2, 3, 4)$ are, as defined above, the treatment indicator variables for the >0-125, >125-250 and >250 μg/day ICS-dose categories respectively, and $\boldsymbol{\alpha}_5' \boldsymbol{C}$ is a linear combination of the potential confounding covariates. The crude model corresponds to model (1) without the last term $\boldsymbol{\alpha}_5' \boldsymbol{C}$.

To account for the clustered nature of our data (that is, 1177 women with two pregnancies), we used cluster bootstrap (Field and Welsh, 2007) to calculate 95% confidence intervals (CI) with the percentile method (Chernick, 2007) for all ATE. The bootstrapped CI were obtained for each contrast and both approaches (GPS and conventional multivariable). Note that our bootstrap algorithm was applied at the GPS-based outcome regression modeling stage only (the standard error calculation does not take into account the uncertainty in estimating the GPS).

2.7.2 Additional analyses

Additional analyses were performed after obtaining the primary results. These pertain to three different aspects having the potential to affect the results, namely, 1) the choice of the

potential confounding covariates; 2) the violation of the positivity assumption with LABA use; 3) the adjustment for gestational age (GA), a potential mediator of the relationship between ICS and BW.

The results in Cossette et al. (2013) were obtained on the basis of a reduced subset of covariates selected by a change-in-estimate procedure. We performed a series of supplemental analyses to assess the robustness of our primary results to the choice of covariates within the initial set of covariates of size 27. These analyses were justified by the fact that confounder selection is a challenging issue, and that only a few covariates related to asthma were retained in the low BW regression model fitted by Cossette et al. (2013). The strategies for variable selection were a) use the covariates selected by Cossette et al. (2013) in the adjusted analysis of low BW, in addition to baby's sex (set 1); b) use baby's sex and all the covariates selected at least in one model for the perinatal outcomes studied in Cossette et al. (2013) (set 2); c) use a set of 18 potential confounding covariates obtained by applying the recently developed Bayesian Adjustment for Confounding algorithm (Wang et al., 2012; Lefebvre et al., 2014) (set 3) (see Table A1 in Appendix A).

In the primary analyses, adjustment for LABA, which is a marker of asthma severity, might be problematic. Indeed, to validly apply the GPS or the conventional multivariable approach, a strictly positive probability of no exposure to ICS ought to be possible for babies exposed to LABA during their mother's pregnancies (Westreich and Cole, 2010). However, medical guidelines do not recommend the use of LABA without ICS for the treatment of asthma (Lougheed et al., 2012; Namazy et al., 2014), and our cohort does not include pregnancies exposed to LABA that were unexposed to ICS. As such, it is ill-posed to define and predict the counterfactual BW outcomes associated with no use of ICS for babies who have been exposed to LABA. Secondary GPS analyses restricted to 6724 pregnancies unexposed to LABA (with the full set of covariates) were thus done, repeating each primary GPS analysis described above, while excluding 650 (8.81%) pregnancies exposed to LABA. Crude and adjusted regression (conventional) estimates were also obtained on this subcohort. These secondary analyses were not performed in the study by Cossette et al. (2013).

Assuming the existence of a causal relationship between ICS and GA and between GA and BW, each estimated ICS dose-category effect can be interpreted as a total effect that combines the direct and indirect effects of ICS on BW, and where the indirect effect is mediated through GA. As it is well-known that adjusting for an intermediate variable (GA) in causal models is problematic (VanderWeele et al., 2012; Richiardi et al., 2013; Morgan and Winship, 2015), we did not, as Cossette et al. (2013), included GA a priori in our models. Given the uncertainty about the mediating nature of GA in the present context however, we also performed sensitivity analyses and included GA in the corresponding models. Indeed, as GA is as a very strong predictor of BW (Kramer et al., 2001; Oken et al., 2003), an important motivation for considering these augmented models is the possible large gain in efficiency, especially for the highest ICS-dose category which corresponds to the smallest sample size among the exposure levels. These sensitivity analyses were performed both for the primary and the secondary analyses.

2.7.3 Estimation of a dose-response function

Our decision to consider multilevel GPS analyses was motivated by the desire to mimic the analytical strategy used in our reference paper (Cossette et al., 2013), which itself follows the clinical practice in the management of asthma. However, as it can be argued that the choice of a specific categorization of the exposure is to some extent arbitrary, it is of interest to investigate the relationship between ICS and BW when ICS-dose is considered as a continuous variable. As such, the general GPS approach described in Hirano and Imbens (2004) was implemented to better uncover possible heterogeneity of the exposure effect in regions corresponding to higher ICS-doses.

We selected a two-part model (Min and Agresti, 2002; Gelman and Hill, 2007) to compute the GPS associated to each pregnancy in our cohort. The first part of the exposure model was a logistic model for the dichotomous event of having zeroed or positive values of ICS dose, and, conditional on a positive value, the second part assumed a log-normal distribution. The discrete part of the model accounted for the fact that 43% of the pregnancies were unexposed to ICS, while the distribution of the positive doses of ICS was well-captured by the log-normal part. Of note, parameter estimation for this exposure model proceeded relatively simply because the likelihood function factors into two terms (Min and Agresti, 2002).

As Hirano and Imbens (2004), Kluve et al. (2012) and Bia et al. (2014), we used a blocking method to assess the balance of the covariates after GPS adjustment. The cohort was divided into four groups: the first group was defined by no exposure to ICS while the other three groups were formed by cutting the positive ICS-dose values at the 33th and 66th percentiles. For each pregnancy, we evaluated the GPS at the median of the exposure variable within each group. For each of these groups, and only using the GPS of the pregnancies associated to a given exposure group, we then created five blocks corresponding to the quintiles of the GPS evaluated at the median. Our check for balance then proceeded similarly to that described in Section 2.7.1.

The conditional expectation of the outcome (BW) given observed ICS exposure level t_i and the estimated GPS $\hat{r}(t_i, \boldsymbol{x}_i)$ was modeled according to the following quadratic equation:

$$E[Y_i|t_i, \hat{r}(t_i, \boldsymbol{x}_i)] = \alpha_0 + \alpha_1 t_i + \alpha_2 t_i^2 + \alpha_3 \hat{r}(t_i, \boldsymbol{x}_i) + \alpha_4 \hat{r}(t_i, \boldsymbol{x}_i)^2 + \alpha_5 t_i \hat{r}(t_i, \boldsymbol{x}_i). \tag{2}$$

The coefficients from Equation (2) were used to evaluate the dose-response function (DRF) at exposure level t by estimating the average potential outcome $E[Y(t)]$:

$$\hat{E}[Y(t)] = \frac{1}{n} \sum_{i=1}^{n} \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 \hat{r}(t, \boldsymbol{x}_i) + \alpha_4 \hat{r}(t, \boldsymbol{x}_i)^2 + \alpha_5 t \hat{r}(t, \boldsymbol{x}_i).$$

A DRF was obtained for both the full cohort and the subcohort which excluded the pregnancies exposed to LABA (using the full set of potentially confounding variables). We estimated contrasts defined as the difference in the DRF value at the median exposure level within each category of positive exposure to ICS (>0-125, >125-250, >250 μg/day) versus the null exposure value (0 μg/day). Sensitivity analyses for the choice of the potential confounding variables and the inclusion of GA in the GPS models were also performed.

2.8 Ethics approval

We obtained approval from the Commission d'accès à l'information du Québec prior to requesting and linking the information from the MED-ECHO and RAMQ databases. This study was approved by the ethics committee of the Hôpital du Sacré-Coeur de Montréal.

3. Results

The cohort includes 7374 pregnancies from 6197 asthmatic women. Table 1 presents the distribution of the pregnancies and the distribution of the 27 potential confounding covariates according to the average ICS daily dose categories. We remark that higher-ICS-dose users take more leukotriene-receptor antagonists, short-acting beta2-agonists (SABA), and oral and intranasal corticosteroids. They also have more severe asthma and have more emergency visits for asthma. The mean GA (in weeks) is very similar across ICS categories (standard deviation in parenthesis): 38.54 (2.54), 38.60 (1.99), 38.45 (2.12) and 38.41 (2.22) for ICS doses 0, >0-125, >125-250, <250 μg/day, respectively.

Table C1 in Appendix C provides the estimated regression coefficients (mean difference(s) for each potential confounding variable adjusted for ICS and other covariates) obtained from a linear regression model for BW. According to Table C1, Antiphospholipid syndrome, Eclampsia/pre-eclampsia, Placenta abruption, and Fetal-maternal hemorrhage are the most influential factors to explain BW (-322.33, -268.27, -364.03 and -271.99 g, respectively), possibly due to their negative association with GA (Ananth et al., 1999; Rubod et al., 2007; Goldenberg et al., 2008; Di Prima et al., 2011; Villar et al., 2012; Levy et al., 2015); in our data, these covariates are also the most important predictors of GA.

Figure 1 shows box-plots for the GPS values corresponding to each treatment category (primary analysis).

Table 2 provides the standardized differences of the potential confounding covariates across the groups defined by each treatment indicator's values. These results reveal the amount of evidence against initial balance between covariates and treatment indicators. We note that, except for ICS doses >0-125 μg/day, almost all asthma-related variables uniformly present standardized differences larger than the acceptable threshold to define balance (standardized difference of 0.20, see McCaffrey et al. (2013)). We thus reject distributional balance with respect to these covariates and the treatment indicator groups for each treatment. Table 3 reports the average standardized differences taken over the five or ten GPS strata (according to ICS-dose category). We observe from Table 3 an overall improvement of the balance for the problematic variables, but some of the standardized differences associated with variables SABA and Severity of asthma prior to pregnancy still remain larger than the maximum acceptable value.

The first part of Table 4 provides the results of the primary analyses: the unadjusted (crude) ATE, the ATE for both the multilevel GPS and the adjusted regression (conventional) approaches, and the results of the corresponding sensitivity analyses which consist in the inclusion of GA in the models. The second part of Table 4 presents the results for the secondary analyses in which we considered only the pregnancies unexposed to LABA. All primary analyses based on the GPS and conventional approaches reveal that babies exposed to ICS doses >0-125 and >125-250 μg/day did not significantly differ in BW as compared to babies unexposed to ICS (at level 0.05). The same conclusion is found for ICS-dose category

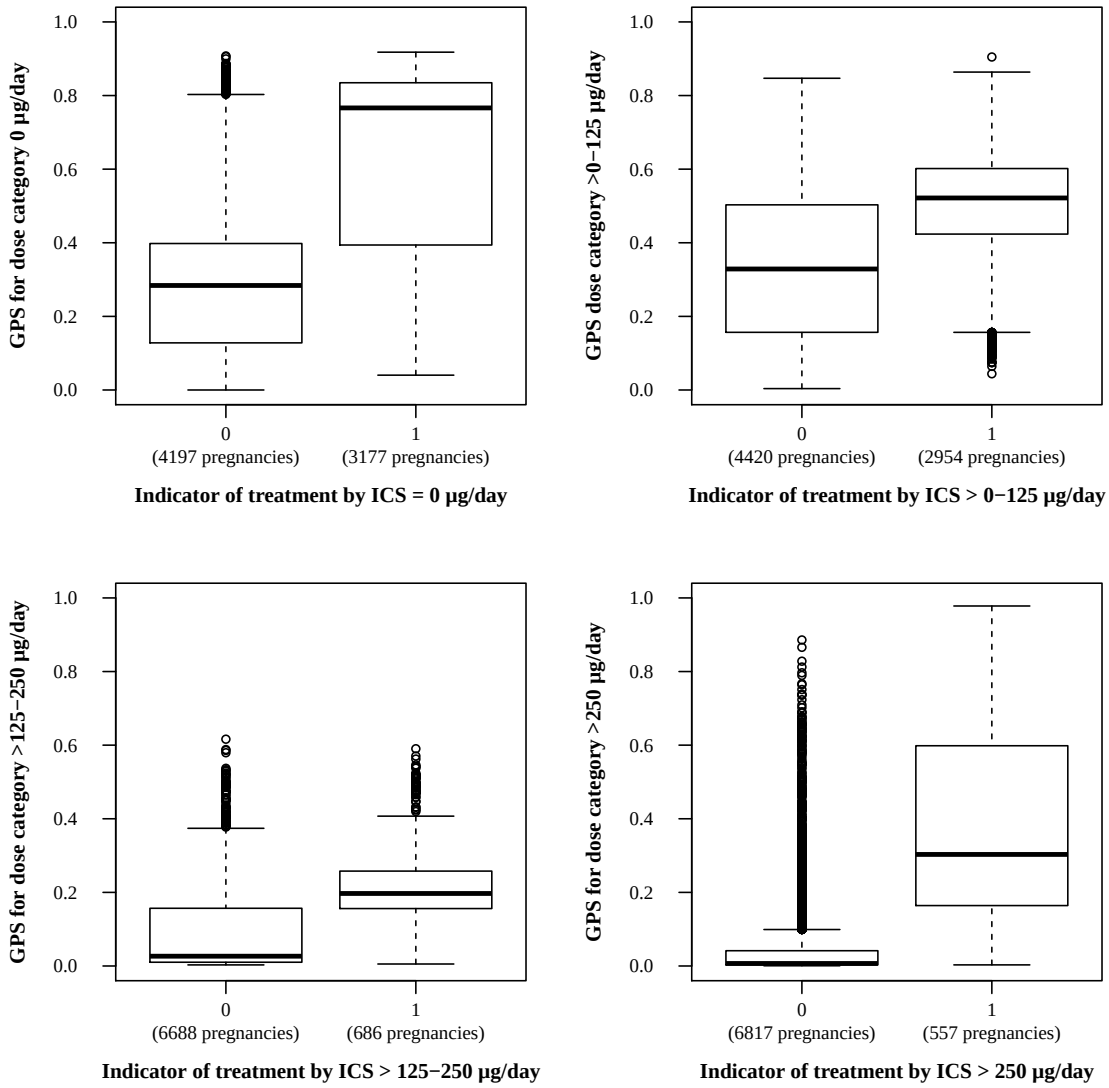

Figure 1: Overlap of the GPS between the treatment group indicators.

>250 μg/day in all primary GPS-based analyses. For ICS-dose category >250 μg/day, all primary results obtained from the conventional approach show a decrease in BW of approximately 50 g when the baby is exposed; the result is statistically significant (at level 0.05) when variable GA is included in the adjusted regression model. In general, the inclusion of GA in the models pertaining to the primary analyses does not substantially change the corresponding ATE point estimates.

The results for the sensitivity analyses associated to the choice of confounding variables appear in Appendix D. For both the GPS and the conventional approaches, the results based on the full set of 27 covariates and the reduced set 3 (refer to Table A1) were very similar.

In general, the results obtained by the adjusted conventional approach are consistent across the sets of covariates (full set, set 1, set 2, or set 3). The same tendency can be reported for the GPS approach for ICS-dose category >0-125 μg/day.

The secondary analyses, which were restricted to pregnancies unexposed to LABA, were initially justified on the basis of theoretical positivity problems. Such problems were also seen in practice through the extremely small estimated GPS associated to dose category 0 μg/day for pregnancies exposed to LABA (minimum GPS=1.89×10^{-11}). Eliminating these pregnancies improved overall positivity (results not shown). As clearly observed herein, GPS approaches make positivity issues explicit (Westreich et al., 2011), unlike regression approaches. All secondary GPS analyses show results pointing in the same direction as those obtained from the corresponding primary analyses. This is also true for the conventional approach, except for ICS-dose category >125-250 μg/day for which the ATE obtained for the primary and secondary analyses go in opposite directions.

For both the primary and secondary analyses, the confidence intervals associated with the multilevel GPS approach are wider than those obtained with the conventional multivariable approach. In particular, and unlike the GPS approach, the inclusion of GA in the adjusted regression models results in more narrow confidence intervals (22.72-25.58% narrower).

For the continuous GPS analyses, all standardized differences were smaller than 0.20 after GPS adjustment, except for no use of SABA (0.21), >3 doses/week of SABA (0.23), and use of LABA (0.24). The shape of the estimated DRF (see Fig. 2) suggests, starting from small positive ICS-dose values, a decreasing relationship between average daily ICS-dose during pregnancy and BW. Table 5 presents the ATE estimated using the approach described in Section 2.7.3. These results reveal a small decrease in BW for babies exposed to the median ICS-dose in >0-125 and >125-250 μg/day compared to those unexposed to ICS. For moderate-to-high doses (>250 μg/day), this decrease is approximately 70 and 50 g in the primary and secondary analyses, respectively. All contrasts are non-significant (at level 0.05). Table 5 also displays the robustness of the results to the inclusion of GA in the analyses. The results for the sensitivity analyses associated to the choice of the confounding variables appear in Appendix E.

4. Discussion

In the primary analyses, for ICS-dose category >250 μg/day, we found an ATE of -37.83 g (95% CI: -117.74, 41.53) for the multilevel GPS-based model and an ATE of -51.16 g (95% CI: -118.80, 14.17) for the adjusted regression model. The results from the continuous GPS approach agree with the above findings about the effect of the highest ICS-dose category on BW (ATE: -67.63, 95% CI: -138.80, 7.14). All these results are also in agreement with the results from Cossette et al. (2013) who found a non-significant trend for increased risk of low BW for the two highest ICS-dose categories (adjusted OR=1.20, 95% CI: 0.81, 1.78 for ICS >250-500 μg/day; adjusted OR=1.57, 95% CI: 0.86, 2.87 for ICS >500 μg/day). A decreasing relationship between ICS-dose and BW was similarly observed through the DRF obtained from the continuous GPS analyses. Most interestingly, the effect of exposure to ICS-dose category >250 μg/day on BW was significantly different from zero for the adjusted regression model including GA (ATE: -56.97, 95% CI: -109.30, -6.54). To our knowledge,

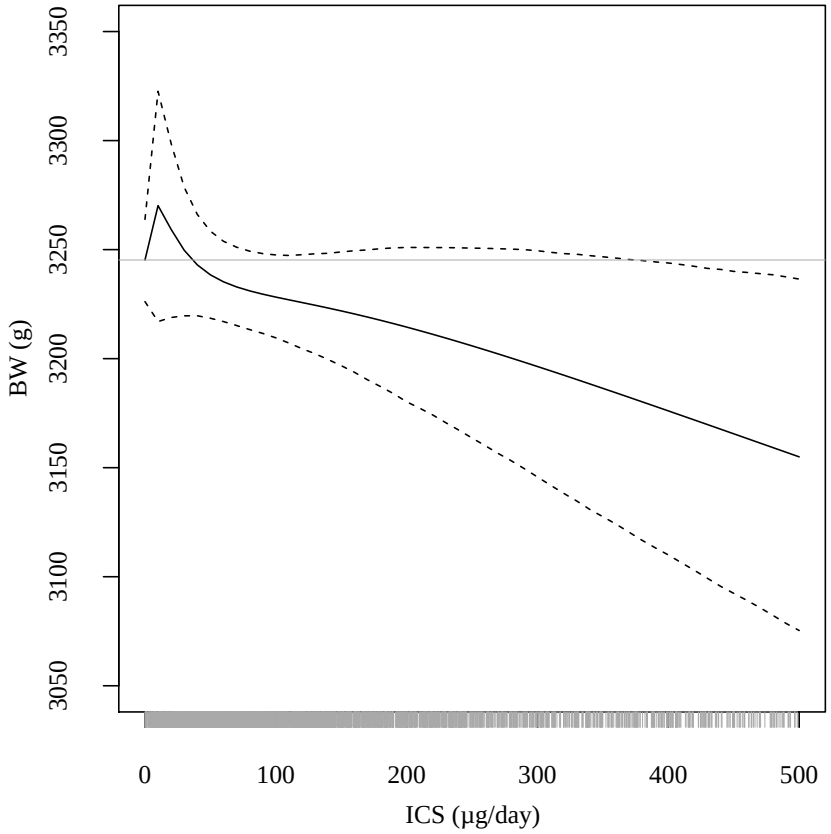

Figure 2: Estimated dose-response function with confidence bands.
Legend. Solid line: DRF; slashed line: DRF CI; pale solid line: predicted mean BW for ICS=0 μg/day.

this result is the first reported result which supports the detrimental effect of higher doses of ICS on BW-related variables.

Our study has some limitations that should be taken into account when interpreting the results, in addition to those already mentioned in Cossette et al. (2013). These pertain to the following three points: 1) balance and GPS modeling; 2) number of GPS strata and outcome regression modeling; 3) robustness of results to the choice of confounders.

Balance and GPS modeling. It is not excluded that other, more sophisticated, GPS models could have been more appropriate and led to greater balance between covariates and treatment indicators, thus reducing even further confounding by indication. The standardized differences associated with SABA and Severity of asthma prior to pregnancy were very large before GPS adjustment, but they still exceeded the loose threshold of 0.20 after GPS adjustment in the multilevel GPS analyses. These two variables were found marginally associated with BW, so our primary results need to be interpreted with caution. Of note, we experimented with a generalized boosted model for the GPS (McCaffrey et al., 2013), but

found that three additional variables related with asthma exceeded this threshold (results not reported). This comforted us in our decision to use a GPS model based on multinomial regression for our multilevel analyses. For the continuous GPS approach, SABA and LABA were found unbalanced after GPS adjustment, with standardized differences however close to the threshold.

Number of GPS strata and outcome regression modeling. In our GPS analyses, the validity of the predicted counterfactual outcomes, and thus of the estimated ATE, is intrinsically linked to the choice of an appropriate model for the outcome versus the GPS. In our multilevel GPS analyses, we made the decision to use deciles of GPS for the two smallest ICS-dose categories and quintiles of GPS for the two highest ICS-dose categories. In these analyses, a modeling strategy based on GPS strata was deemed preferable to minimize modeling assumptions and to increase the robustness of the predictions to more extreme values of GPS. In contrast, the predicted counterfactual outcome in our continuous GPS was modeled as a quadratic function of the exposure level and GPS. On the basis of resulting DRF, this model was selected over other modeling strategies, such as cubic regression, quadratic and cubic spline regressions, and penalized spline smoothing based on additive spline bases. Unlike the multilevel GPS approach, the continuous GPS approach does not rely on individual models applied to subsets of data corresponding to exposure categories of interest. However, our continuous GPS analyses rely on stronger parametric assumptions for the outcome model and could be more sensitive to outlying GPS values compared to our multilevel GPS analyses. While the multilevel and continuous GPS approaches did not agree in the direction of the ATE for the two lowest ICS-dose categories, they did agree in the small magnitude of the effects. More importantly, similar results were obtained for the highest ICS-dose category under both approaches. In general, it would be worthwhile to investigate the performances of multilevel versus continuous GPS for scientifically relevant exposure categories in various simulated scenarios. In addition to the median exposure strategy adopted in this paper, fully adapting continuous GPS analyses for a categorical exposure could be done by averaging the estimated dose-response function at the observed distribution of treatment within a given exposure category. This could however be computationally very expensive and not widely applicable, as herein.

Robustness of results to the choice of confounders. The supplemental analyses presented in Appendix D revealed that the conventional multivariate approach estimates were generally robust to different sets of confounding covariates. This was also seen for the multilevel GPS analysis for the >0-125 versus 0 μg/day category; results were however less stable across sets of confounding covariates for the GPS analysis for the >125-250 versus 0 category and the one for the >250 versus 0 μg/day category. Interestingly, the continuous GPS approach showed increased robustness with respect to the choice of confounding variables (refer to Appendix E). Because the outcome modeling is performed separately for each exposure category in the multilevel GPS approach (as opposed to the continuous GPS approach), it is reasonable to believe that the observed instability of the multilevel GPS estimates is a consequence of the small sample sizes for the corresponding subsets of observations. Recall that the two highest ICS-dose categories have prevalence of only 9.30% and 7.55%, respectively, in our cohort. Lastly, as in Cossette et al. (2013), we do not have information on the smoking status of the mothers. Using a method described in Schneeweiss et al. (2005),

Cossette et al. (2013) found their estimates robust to this unmeasured confounder; hence no additional analyses with respect to smoking were performed in our study.

The principal strength of our study is the care we took in performing analyses based on propensity scores that examined outcomes by dosage in attempt to better account for confounding by indication, which is a concern in this study. We feel, as others, that explicitly providing insights on potential positivity problem may be one of the main advantages of the propensity score approaches (Westreich et al., 2011). As such, while propensity score analyses are recommended, it is worthwhile to highlight the additional difficulty in implementing such type of analyses for an exposure with multiple levels, as opposed, for instance, to conventional analyses based on a multivariable model for the outcome. Herein, this difficulty was exacerbated because of the structure of the data. In our study, we have used cluster bootstrap to account for intracluster correlation, that is, possible similarity between a woman's consecutive pregnancies. However, only very limited information is currently available as to best apply GPS on clustered data, especially with a large number of small clusters (Li et al., 2013).

Whether one should include GA, a potential intermediate variable in the pathway between ICS and BW, in the models we considered is debatable. Interestingly, the inclusion of GA in the GPS models as opposed to the conventional ones had different impact on the results. Whereas no reduction in the variability of the estimates was seen for the GPS results, it had a large impact on the variability of the estimated effects obtained from the conventional models. These results suggest that the conventional approach may take further advantage of the inclusion of a strong predictor of the outcome in the modeling process than the GPS approach. Overall, we feel that the gain in efficiency outweighs the possible bias we might have introduced by including this variable in the models, especially since the point estimates were robust to the inclusion of GA in the models. As such, the conventional approach adjusted for GA appears to be the most appropriate in the present context. One obvious methodological extension would be the application of formal mediation analyses to refine our analysis of the effects of ICS doses on GA and BW.

Worldwide, ICS are the first-line treatment for asthma. While high doses ICS have been associated with increased malformations (Blais et al., 2007; Gregersen and Ulrik, 2013), our results suggest that such doses of ICS could also have a negative impact on the BW-axis of infants. To our knowledge, there does not exist a strict threshold for a clinical significance in BW variation. Although a reduction of 50 grams can arguably be qualified as small, the "oeuf-lait-orange" (egg-milk-orange) (OLO) prenatal nutrition program, that has been progressively implemented in the province of Québec (Canada) since the eighties, has recently been found to increase the BW of a participating baby of about 70 grams on average (Haeck and Lefebvre, 2016). Therein, the economic benefit of the OLO program (with cost of the order of $543 per baby in 2008) was established on the basis of decreased neonatal hospital costs and higher educational attainment and wages leading to accrued revenue for both the government and the person. Specifically, a 10% increase in BW increases high school completion by approximately 1 percentage point (Black et al., 2007), while the median earnings for those who completed high school is about $5,000 larger than for those who did not (Statistics Canada). The socio-economic significance of a decrease of about 50 grams in BW could similarly be argued.

When asthma is not controlled at low or medium doses of ICS, guidelines now recommend initiating combination therapy with LABA before escalating to high doses of ICS (Lougheed et al., 2012). Physicians thus have to decide between alternative regimens to treat asthma during pregnancy. Similar causal analyses could be done to assess the effect of different ICS/LABA combination therapies on BW and ultimately inform physicians on the best combination of ICS/LABA to keep asthma under control while limiting the impact of treatment on this important perinatal outcome.

Acknowledgments

We thank the Commission d'accès à l'information du Québec for authorizing the study and the Régie de l'assurance maladie du Québec and IRIS-Québec for assistance with the data.

This work was supported by the National Sciences and Engineering Research Council of Canada, the Canadian Institutes of Health Research and the Fonds de recherche Québec-Santé. Geneviève Lefebvre is a Chercheur-Boursier of the Fonds de recherche Québec-Santé. Lucie Blais co-chairs the Endowment Chair AstraZeneca in Respiratory Health.

Table 1: Distribution (frequency) of the potential confounding covariates according to the average ICS daily dose categories during pregnancy (% shown in parenthesis).

	Average ICS[a] doses during pregnancy (μg/day)				Total
	0 (None)	> 0-125	> 125-250	> 250	
Number of pregnancies	3177 (43.1)	2954 (40.1)	686 (9.3)	557 (7.6)	7374 (100)
Mother's & baby's characteristics					
Maternal age (years)					
< 18	50 (1.6)	64 (2.2)	13 (1.9)	4 (0.7)	131 (1.8)
18-34	2755 (86.7)	2514 (85.1)	575 (83.8)	439 (78.8)	6283 (85.2)
> 34	372 (11.7)	376 (12.7)	98 (14.3)	114 (20.5)	960 (13.0)
Baby's sex (male)	1584 (49.9)	1543 (52.2)	348 (50.7)	272 (48.8)	3747 (50.8)
Receipt of social assistance	1636 (51.5)	1720 (58.2)	416 (60.6)	363 (65.2)	4135 (56.1)
Urban residency	2581(81.2)	2348(79.5)	561(81.8)	449(80.6)	5939 (80.5)
Maternal chronic conditions					
Antiphospholipid syndrome	17 (0.5)	17 (0.6)	4 (0.6)	3 (0.5)	41 (0.5)
Chronic hypertension	86 (2.7)	88 (3.0)	17 (2.5)	24 (4.3)	215 (2.9)
Diabetes mellitus	102 (3.2)	104 (3.5)	34 (5.0)	30 (5.4)	270 (3.7)
Cystic fibrosis of the pancreas	14 (0.4)	13 (0.4)	2 (0.3)	8 (1.4)	37 (0.5)
Uterine defects	407 (12.8)	371 (12.6)	100 (14.6)	88 (15.8)	966 (13.1)
Pregnancy-related variables					
Gestational diabetes	288 (9.1)	288 (9.8)	76 (11.1)	80 (14.4)	732 (9.9)
Eclampsia/pre-eclampsia	86 (2.7)	95 (3.2)	20 (2.9)	21 (3.8)	222 (3.0)
Anaemia	443 (13.9)	429 (14.5)	110 (16.0)	93 (16.7)	1075 (14.6)
Placental conditions	129 (4.1)	109 (3.7)	26 (3.8)	26 (4.7)	290 (39)
Placenta abruption	112 (3.5)	113 (3.8)	20 (2.9)	20 (3.6)	265 (3.6)
Vaginal bleeding	438 (13.8)	388 (13.1)	78 (11.4)	68 (12.2)	972 (13.2)
Maternal infections	497 (15.6)	403 (13.6)	117 (17.1)	78 (14.0)	1095 (14.9)
Fetal-maternal hemorrhage	5 (0.2)	3 (0.1)	4 (0.6)	4 (0.7)	16 (0.2)
Gestational hypertension	179 (5.6)	189 (6.4)	43 (6.3)	53 (9.5)	464 (62.9)
Use of beta-blockers	25 (0.8)	27 (0.9)	8 (1.2)	2 (0.4)	62 (0.8)
Asthma-related variables					
Leukotriene-receptor antagonists	12 (0.4)	24 (0.8)	17 (2.5)	62 (11.1)	115 (1.6)
SABA (doses/week)					
0	1775 (55.9)	434 (14.7)	41 (6.0)	30 (5.4)	2280 (30.9)
> 0-3	896 (28.2)	1498 (50.7)	93 (14.0)	41 (7.4)	2528 (34.3)
> 3	506 (15.9)	1022 (34.6)	552 (80.0)	486 (87.3)	2566 (34.8)
Oral corticosteroids	117 (3.7)	397 (13.4)	148 (21.6)	154 (27.7)	816 (11.1)
Intranasal corticosteroids	228 (7.2)	408 (13.8)	146 (21.3)	147 (26.4)	929 (1.3)
$\geq$ 1 ED visit for asthma	201 (6.3)	523 (17.7)	145 (21.1)	137 (24.6)	1006 (13.6)
LABA	0 (0)	212 (7.2)	178 (26.0)	260 (46.7)	650 (8.8)
$\geq$ 1 hospitalization for asthma	9 (0.3)	43 (1.5)	17 (2.5)	20 (3.6)	89 (1.2)
Severity of asthma prior to pregnancy					
mild	2887 (90.9)	2471 (83.6)	404 (58.9)	135 (24.2)	5897 (80.0)
moderate	254 (8.0)	354 (12.0)	192 (28.0)	221 (39.7)	1021 (13.8)
severe	36 (1.1)	129 (4.4)	90 (13.1)	201 (36.1)	456 (6.2)

Abbreviations: ED, emergency department; LABA, long-acting beta2-agonists; SABA, short-acting beta2-agonists.

[a]: fluticasone-propionate equivalent.

See Table A1 in Appendix A for more information on these potential confounders.

Table 2: Balance of potential confounding covariates across the groups defined by treatment indicators' values before GPS adjustment. *Standardized differences according to indicators of treatment by ICS.*[a]

	Standardized differences			
	0 (None)	> 0-125	> 125-250	> 250
Mother's and baby's characteristics				
Maternal age (years)				
< 18	0.03	0.05	0.01	0.1
18-34	0.08	0	0.04	0.18
> 34	0.07	0.01	0.04	0.22
Baby's sex	0.03	0.05	0	0.04
Receipt of social assistance	0.16	0.07	0.1	0.2
Urban residence	0.03	0.04	0.03	0
Maternal chronic condition				
Antiphospholipid syndrome	0	0	0	0
Chronic hypertension	0.02	0.01	0.03	0.08
Diabetes mellitus	0.04	0.01	0.07	0.09
Cystic fibrosis of the pancreas	0.02	0.01	0.04	0.11
Uterine defects	0.02	0.03	0.05	0.08
Pregnancy-related variables				
Gestational diabetes	0.03	0.02	0.01	0.05
Eclampsia/pre-eclampsia	0.03	0	0.04	0.06
Anaemia	0.01	0.02	0.01	0.04
Placental conditions	0.01	0.02	0.04	0
Placenta abruption	0.03	0	0.06	0.03
Vaginal bleeding	0.04	0.06	0.07	0.03
Maternal infections	0.02	0.04	0.07	0.08
Fetal-maternal hemorrhage	0.05	0.01	0	0.13
Pregnancy-induced hypertension	0.01	0.01	0.04	0.07
Use of beta-blockers	0.03	0.02	0.01	0.05
Asthma-related variables				
Leukotriene-receptor antagonists	0.18	0.1	0.07	0.45
SABA (doses/week)				
0	1.04	0.63	0.74	0.75
> 0-3	0.23	0.59	0.55	0.75
> 3	0.76	0.01	1.17	1.41
Oral corticosteroids	0.44	0.12	0.32	0.47
Intranasal corticosteroids	0.3	0.06	0.26	0.39
$\geq$ 1 ED visit for asthma	0.39	0.19	0.22	0.31
LABA	n/a	0.1	0.53	1.05
$\geq$ 1 hospitalization for asthma	0.16	0.04	0.11	0.17
Severity of asthma prior to pregnancy				
mild	0.51	0.16	0.53	1.52
moderate	0.31	0.09	0.4	0.67
severe	0.39	0.13	0.27	0.89

Abbreviations: ED, emergency department; LABA, long-acting beta2-agonists; SABA, short-acting beta2-agonists.

[a]: fluticasone-propionate equivalent, in μg/day.

Table 3: Balance of potential confounding covariates across the groups defined by treatment indicators' values after GPS adjustment.[a] *Average standardized differences (standard deviation) according to indicators of treatment by ICS.*[b]

	Average standardized differences			
	0 (None)	> 0-125	> 125-250	> 250
Mother's and baby's characteristics				
Maternal age (years)				
< 18	0.06 (0.05)	0.11 (0.05)	0.14 (0.06)	0.10 (0.19)
18-34	0.11 (0.09)	0.07 (0.05)	0.07 (0.05)	0.15 (0.17)
> 34	0.11 (0.09)	0.07 (0.07)	0.05 (0.05)	0.14 (0.17)
Baby's sex	0.08 (0.11)	0.07 (0.05)	0.06 (0.03)	0.15 (0.15)
Receipt of social assistance	0.15 (0.18)	0.11 (0.11)	0.08 (0.07)	0.05 (0.04)
Urban residence	0.10 (0.08)	0.08 (0.06)	0.10 (0.06)	0.08 (0.06)
Maternal chronic conditions				
Antiphospholipid syndrome	0.05 (0.04)	0.09 (0.05)	0.10 (0.05)	0.03 (0.05)
Chronic hypertension	0.08 (0.06)	0.05 (0.04)	0.15 (0.05)	0.11 (0.08)
Diabetes mellitus	0.08 (0.08)	0.05 (0.04)	0.10 (0.08)	0.11 (0.06)
Cystic fibrosis of the pancreas	0.08 (0.08)	0.08 (0.04)	0.02 (0.03)	0.08 (0.04)
Uterine defects	0.09 (0.08)	0.07 (0.08)	0.06 (0.05)	0.12 (0.09)
Pregnancy-related variables				
Gestational diabetes	0.09 (0.08)	0.07 (0.04)	0.06 (0.04)	0.08 (0.12)
Eclampsia/pre-eclampsia	0.09 (0.09)	0.07 (0.05)	0.11 (0.07)	0.04 (0.05)
Anaemia	0.12 (0.07)	0.09 (0.07)	0.06 (0.05)	0.05 (0.03)
Placental conditions	0.14 (0.13)	0.08 (0.05)	0.03 (0.02)	0.12 (0.09)
Placenta abruption	0.07 (0.08)	0.08 (0.07)	0.06 (0.03)	0.08 (0.06)
Vaginal bleeding	0.09 (0.10)	0.09 (0.06)	0.09 (0.03)	0.12 (0.08)
Maternal infections	0.09 (0.09)	0.10 (0.04)	0.07 (0.05)	0.09 (0.12)
Fetal-maternal hemorrhage	0.05 (0.03)	0.03 (0.04)	0.04 (0.05)	0.09 (0.06)
Pregnancy-induced hypertension	0.09 (0.08)	0.08 (0.07)	0.09 (0.06)	0.10 (0.10)
Use of beta-blockers	0.09 (0.11)	0.10 (0.05)	0.07 (0.06)	0.05 (0.06)
Asthma-related variables				
Leukotriene-receptor antagonists	0.07 (0.08)	0.07 (0.06)	0.09 (0.08)	0.12 (0.17)
SABA (doses/week)				
0	0.08 (0.15)	0.08 (0.06)	0.21 (0.21)	0.33 (0.34)
> 0-3	0.07 (0.12)	0.07 (0.06)	0.12 (0.08)	0.23 (0.20)
> 3	0.03 (0.04)	0.08 (0.07)	0.26 (0.30)	0.46 (0.57)
Oral corticosteroids	0.05 (0.07)	0.05 (0.05)	0.11 (0.04)	0.07 (0.05)
Intranasal corticosteroids	0.08 (0.07)	0.10 (0.05)	0.13 (0.09)	0.15 (0.12)
≥ 1 ED visit for asthma	0.05 (0.08)	0.10 (0.06)	0.07 (0.06)	0.07 (0.05)
LABA	0.10 (0.31)	0.06 (0.06)	0.15 (0.07)	0.12 (0.11)
≥ 1 hospitalization for asthma	0.04 (0.06)	0.08 (0.05)	0.04 (0.03)	0.06 (0.07)
Severity of asthma prior to pregnancy				
mild	0.12 (0.08)	0.08 (0.06)	0.16 (0.12)	0.15 (0.26)
moderate	0.10 (0.07)	0.07 (0.05)	0.13 (0.07)	0.31 (0.22)
severe	0.04 (0.10)	0.06 (0.04)	0.09 (0.09)	0.19 (0.16)

Abbreviations: ED, emergency department; LABA, long-acting beta2-agonists; SABA, short-acting beta2-agonists.

[a]: the average standardized difference is taken over 5 or 10 strata (according to ICS-dose category);
[b]: fluticasone-propionate equivalent, in μg/day.

Table 4: Average treatment effects (primary and secondary analyses) using multilevel GPS approach.

	Categories of ICS doses compared (μg/day)					
	> 0-125 vs 0 (None)		> 125-250 vs 0 (None)		> 250 vs 0 (None)	
	ATE	95% CI[a]	ATE	95% CI[a]	ATE	95% CI[a]
	Primary analyses (7374 pregnancies)					
Unadjusted (crude)	7.41	-21.76, 37.64	-6.6	-55.50, 43.37	-94.41	-149.27, -39.92
GPS analysis	27.62	-8.68, 64.10	17.07	-55.85, 92.16	-37.83	-117.74, 41.53
Adjusted regression	17.41	-15.15, 50.07	15.3	-41.56, 71.09	-51.16	-118.80, 14.17
	Sensitivity analyses: inclusion of GA in the primary analysis models					
GPS analysis	10.05	-25.56, 46.10	13.93	-57.51, 87.01	-52.38	-133.02, 29.14
Adjusted regression	3.42	-21.46, 28.16	14.68	-29.03, 57.17	-56.97	-109.30, -6.54
	Secondary analyses (6724 pregnancies unexposed to LABA)					
Unadjusted (crude)	12.48	-18.19, 42.70	-48.9	-103.01, 9.50	-91.34	-168.15, -14.50
GPS analysis	27.27	-8.79, 63.65	14.31	-69.42, 97.82	-84.54	-199.72, 33.02
Adjusted regression	26.17	-6.25, 60.15	-10.48	-71.61, 54.13	-32.53	-111.13, 45.29
	Sensitivity analyse: inclusion of GA in the secondary analysis models					
GPS analysis	13.8	-21.58, 50.12	17.36	-65.33, 100.08	-99.73	-215.05, 18.70
Adjusted regression	11.1	-13.35, 37.04	-2.4	-49.23, 44.34	-64.29	-123.84, -3.79

Abbreviations: ATE, average treatment effect (grams); CI, confidence interval; GA, gestational age; GPS, generalized propensity score; LABA, long-acting beta2-agonists.

[a]: confidence intervals calculated using cluster bootstrap (Field and Welsh, 2007).

Table 5: Average treatment effects (primary and secondary analyses) using continuous GPS approach: difference in the dose-response function values estimated at the medians of compared ICS-dose categories.

Categories of ICS doses compared (μg/day)					
> 0-125 vs 0 (None)		> 125-250 vs 0 (None)		> 250 vs 0 (None)	
ATE[a]	95% CI[b]	ATE[a]	95% CI[b]	ATE[a]	95% CI[b]
Primary analyses (7374 pregnancies)					
-9.35	-22.93, 5.11	-27.07	-62.37, 10.68	-67.63	-138.80, 7.14
Sensitivity analyses: inclusion of GA in the primary analysis models					
-7.89	-21.56, 6.55	-29.58	-64.52, 7.78	-68.74	-139.44, 4.66
Secondary analyses (6724 pregnancies unexposed to LABA)					
-6.96	-23.88, 9.68	-21.28	-63.01, 20.93	-48.19	-139.12, 46.63
Sensitivity analyse: inclusion of GA in the secondary analysis models					
-7.41	-24.51, 9.25	-26.75	-67.89, 14.91	-50.54	-139.73, 42.24

Abbreviations: ATE, average treatment effect (grams); CI, confidence interval; GA, gestational age; LABA, long-acting beta2-agonists.

[a]: the median values are 0, 57.65, 176.20, and 392.43 μg/day, respectively;

[b]: confidence intervals calculated using cluster bootstrap (Field and Welsh, 2007).

Appendix A. Potential confounding covariates

We refer the reader to Table A.1.

Appendix B. Generalized propensity scores

B.1 Theory

We present a mathematical review of GPS; the counterfactual (potential outcome) framework (Rubin, 1974, 1978) is used for the description. Suppose we have a population of N units, along with a multi-valued treatment variable T which takes on values in the set $\mathcal{T} = \{1, 2, \ldots, m\}$. Let $Y_i(j)$ denote unit i's potential outcome under treatment level j $(i = 1, 2, \ldots, N)$, that is, the outcome that would be observed if unit i had received treatment $T = j$ (possibly contrary to the fact). Let $D_i(j)$ be unit i's indicator for receiving treatment $T = j$, i.e., $D_i(j) = 1$, if $T_i = j$ and $D_i(j) = 0$, otherwise. The observed outcome for unit i is $Y_i = Y_i(j)$ if $D_i(j) = 1$; in other words, the observed outcome coincides with the potential outcome for the treatment level actually received by the unit. The counterfactual representation described above is valid under the Stable Unit Treatment Value Assumption (SUTVA). For unit i, and its corresponding received treatment T_i, SUTVA asserts that the value of $Y_i(T_i)$ is stable (i.e., determined). SUTVA rules out hidden versions of treatments as well as interference between units (Rubin, 2010).

The average effect of treatment $T = k$ versus treatment $T = l$ $(k, l \in \mathcal{T})$ is defined as

$$\theta_{kl} = E[Y(k) - Y(l)] = E[Y(k)] - E[Y(l)], \tag{B.1}$$

where $E[Y(j)]$ refers to the mean of the potential outcome $Y_i(j)$ taken over all the units in the population.

Upon the observation of the potential outcomes $Y_i(k)$ and $Y_i(l)$ for every unit in a representative sample of size n drawn from the studied population, the effect θ_{kl} can be estimated by

$$\hat{\theta}_{kl} = \frac{1}{n}\sum_{i=1}^{n} Y_i(k) - \frac{1}{n}\sum_{i=1}^{n} Y_i(l). \tag{B.2}$$

Since either $Y_i(k)$ or $Y_i(l)$, or both, are typically unobserved for a given unit, we cannot use estimator (B.2) for the average treatment effect θ_{kl}. In the context of observational studies, Imbens (2000) proposed an approach for estimating of the mean potential outcomes $E[Y(j)]$ $(j \in \mathcal{T})$ based on GPS.

Imbens (2000) defined the generalized propensity score $r(j, \boldsymbol{X})$ as the conditional probability of receiving treatment $T = j$ given pre-treatment covariates $\boldsymbol{X}$, that is, $r(j, \boldsymbol{X}) = Pr(T = j|\boldsymbol{X})$. The following two assumptions are required for appropriately developing the GPS methodology:

1. **Positivity.** There is a positive probability of receiving all levels of treatment given every possible values $\boldsymbol{x}$ of the pre-treatment covariates $\boldsymbol{X}$ in the population:

$$Pr(D(j) = l|\boldsymbol{X} = \boldsymbol{x}) > 0 \qquad \forall j \in \mathcal{T}, \quad \forall \boldsymbol{x}, \quad l = 0, 1.$$

2. **Weak unconfoundness (given the pre-treatment covariates).** Assignment to treatment T is weakly unconfounded if, for all $j \in \mathcal{T}$, the exposure status for treatment $T = j$ and potential outcome $Y(j)$ are conditionally independent given the pre-treatment covariates:

$$D(j) \perp\!\!\!\perp Y(j)|\boldsymbol{X} \qquad \forall j \in \mathcal{T}.$$

Let $\beta(j,r)$ denote the mean potential outcome under treatment $T = j$ given that the GPS $r(j,\boldsymbol{X})$ equals r, that is $\beta(j,r) = E[Y(j)|r(j,\boldsymbol{X}) = r]$. Imbens (2000) showed that if treatment T is weakly unconfounded given the pre-treatment variables $\boldsymbol{X}$, then

$$E[Y(j)] = E[\beta(j, r(j,\boldsymbol{X}))]. \tag{B.3}$$

Moreover, Imbens (2000) proved that $\beta(j,r) = E[Y|T = j, r(T,\boldsymbol{X}) = r]$, and therefore $\beta(j,r)$ can be estimated in the subsample of units that actually received the jth level of treatment. This result is established by invoking the balancing property of GPS:

$$D(j) \perp\!\!\!\perp \boldsymbol{X}|r(j,\boldsymbol{X}) \qquad \forall j \in \mathcal{T},$$

i.e. the probability that $T = j$ does not depend on the value of $\boldsymbol{X}$ within strata with the same value of $r(j,\boldsymbol{X})$ (Hirano and Imbens, 2004).

B.2 Methodology

We present a regression-based GPS approach for the estimation of the ATE. Another approach is through inverse-probability-weighting (Feng et al., 2012). Matching is not frequently encountered with GPS (StataCorp LP, 2009; Feng et al., 2012; Rassen et al., 2013).

Step 1. *Estimation of the generalized propensity scores*
The first step is to estimate the propensity of unit i $(i = 1, 2, \ldots, n)$ to receive treatment level $j \in \mathcal{T}$; that is, we want to estimate the GPS $r(j,\boldsymbol{x}_i) = Pr(T = j|\boldsymbol{X}_i = \boldsymbol{x}_i)$ $\forall j, i$. To this end, one can use multinomial logit, multinomial probit or nested logit models if the treatment variable T is nominal (Tchernis et al., 2005). If T is ordinal, ordered logit or probit models can also be used (Lechner, 2001; Spreeuwenberg et al., 2010). Then, for each unit i, we obtain $\hat{r}(1,\boldsymbol{x}_i), \ldots, \hat{r}(m,\boldsymbol{x}_i)$, corresponding to the unit's estimated probabilities to receive all of the m levels of treatment (these probabilities sum to one for a given unit).

Step 2. *Estimation of* $\beta(j, r(j,\boldsymbol{X}))$
Within the subset of units that received treatment $T = j$ $(j = 1, \ldots, m)$, we regress the outcome Y on GPS $r(j,\boldsymbol{X})$ (or function thereof):

$$E\{Y_i|T = j, r(j,\boldsymbol{x}_i)\} = \alpha_j + \gamma_{j1} g_1(r(j,\boldsymbol{x}_i)) + \ldots + \gamma_{jk} g_k(r(j,\boldsymbol{x}_i)), \tag{B.4}$$

where $g_1, \ldots, g_k$ are user-defined regression basis functions. Examples of function of $r(j,\boldsymbol{X})$ include $g(r) = r$ or $g(r) = \log(r/(1-r))$. This function can also be defined by categories

of percentiles of $r(j, \boldsymbol{X})$, such as

$$\begin{aligned} g_1(r(j,\boldsymbol{x})) &= I(r(j,\boldsymbol{x}) \leq q_j(20)), \\ g_2(r(j,\boldsymbol{x})) &= I(q_j(20) < r(j,\boldsymbol{x}) \leq q_j(40)), \\ g_3(r(j,\boldsymbol{x})) &= I(q_j(40) < r(j,\boldsymbol{x}) \leq q_j(60)), \\ g_4(r(j,\boldsymbol{x})) &= I(q_j(60) < r(j,\boldsymbol{x}) \leq q_j(80)), \end{aligned}$$

where I is the indicator function and $q_j(20), \ldots, q_j(80)$ are the ordered quintiles of $r(j, \boldsymbol{x})$ for individuals with $T = j$. In practice, the true GPS are unknown and thus $r(j, \boldsymbol{x}_i)$ is replaced by its estimate $\hat{r}(j, \boldsymbol{x}_i)$ in model (B.4).

For each treatment level j, we then obtain estimated coefficients $\hat{\alpha}_j, \hat{\gamma}_{j1}, \ldots, \gamma_{jk}$ that are used to estimate $\beta(j, r(j, \boldsymbol{x}_i))$ for all the units in the sample:

$$\hat{\beta}(j, r(j, \boldsymbol{x}_i)) = \hat{\alpha}_j + \hat{\gamma}_{j1} g_1(\hat{r}(j, \boldsymbol{x}_i)) + \ldots + \hat{\gamma}_{jk} g_k(\hat{r}(j, \boldsymbol{x}_i)). \tag{B.5}$$

In this step, we emphasize that the linear function (B.5) permits the calculation of the predicted conditional potential outcomes associated with treatment level j for every unit, whether or not the unit actually received this level of treatment. In particular, Equation (B.5) reveals that all individuals having the same propensity to receive treatment level j are attributed the same value for the predicted potential outcome $Y(j)$ (that is, the same estimated $E[Y(j)|r(j, \boldsymbol{X}) = r]$ value).

Step 3. *Estimation of* $E[Y(j)]$
In consequence of (B.3), $E[Y(j)]$, $j \in \mathcal{T}$, is estimated as a simple average:

$$\hat{E}[Y(j)] = \hat{E}[\beta(j, r(j, \boldsymbol{X}))] = \frac{1}{n} \sum_{i=1}^{n} \hat{\beta}(j, r(j, \boldsymbol{x}_i)).$$

Step 4. *Estimation of the average effect of treatment* $T = k$ *versus treatment* $T = l$
We estimate the average treatment effect (B.1) as follows:

$$\hat{\theta}_{kl} = \hat{E}[Y(k)] - \hat{E}[Y(l)].$$

Step 5. *Calculation of the 95% confidence interval for* θ_{kl}
We calculate 95% confidence intervals for θ_{kl} by bootstrap (Chernick, 2007) (e.g., percentile or bootstrap-t intervals).

Appendix C. Mean difference(s) for each potential confounding variable adjusted for ICS and other covariates

We refer the reader to Table C.1.

Appendix D. Primary analyses based on different strategies for covariate selection: multilevel GPS approach

We refer the reader to Table D.1.

Appendix E. Primary analyses based on different strategies for covariate selection: continuous GPS approach

We refer the reader to Table E.1.

References

Aibar, L., Puertas, A., Valverde, M., Carrillo, M. P., and Montoya, F. (2012). Fetal sex and perinatal outcomes. *J Perinat Med*, 40(3):271–6.

Ali, M. (2014). *Improving Propensity Score Methods in Pharmacoepidemiology*. Thesis, University of Utrecht, Utrecht, the Netherlands.

Ananth, C. V., Berkowitz, G. S., Savitz, D. A., and Lapinski, R. H. (1999). Placental abruption and adverse perinatal outcomes. *JAMA*, 282(17):1646–51.

Austin, P. C. (2011a). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res*, 46(3):399–424.

Austin, P. C. (2011b). A tutorial and case study in propensity score analysis: An application to estimating the effect of in–hospital smoking cessation counseling on mortality. *Multivariate Behav Res*, 46(1):119–151.

Bakhireva, L. N., Jones, K. L., Schatz, M., Johnson, D., and Chambers, C. D. (2005). Asthma medication use in pregnancy and fetal growth. *J Allergy Clin Immunol*, 116(3):503–9.

Bia, M., Flores, C., Flores-Lagunes, A., and Mattei, A. (2014). A Stata package for the application of semiparametric estimators of dose–response functions. *Stata Journal*, 14(3):580–604.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2007). From the cradle to the labor market? The effect of birth weight on adult outcomes. *The Quarterly Journal of Economics*, 122(1):409–439.

Blais, L., Beauchesne, M. F., Rey, E., Malo, J. L., and Forget, A. (2007). Use of inhaled corticosteroids during the first trimester of pregnancy and the risk of congenital malformations among women with asthma. *Thorax*, 62(4):320–8.

Breton, M. C., Beauchesne, M. F., Lemiere, C., Rey, E., Forget, A., and Blais, L. (2009). Risk of perinatal mortality associated with asthma during pregnancy. *Thorax*, 64(2):101–6.

British Thoracic Society and Scottish Intercollegiate Guidelines Network (2014). British guideline on the management of asthma. *Thorax*, 69 Suppl 1:1–192.

Brookhart, M. A., Wyss, R., Layton, J. B., and Sturmer, T. (2013). Propensity score methods for confounding control in nonexperimental research. *Circ Cardiovasc Qual Outcomes*, 6(5):604–11.

Chernick, M. R. (2007). *Bootstrap Methods: A Guide for Practitioners and Researchers*. John Wiley & Sons, Hoboken, NJ, 2nd edition.

Cossette, B., Forget, A., Beauchesne, M. F., Rey, E., Lemiere, C., Larivee, P., Battista, M. C., and Blais, L. (2013). Impact of maternal use of asthma-controller therapy on perinatal outcomes. *Thorax*, 68(8):724–30.

Di Prima, F. A., Valenti, O., Hyseni, E., Giorgio, E., Faraci, M., Renda, E., De Domenico, R., and Monte, S. (2011). Antiphospholipid syndrome during pregnancy: the state of the art. *J Prenat Med*, 5(2):41–53.

Feng, P., Zhou, X. H., Zou, Q. M., Fan, M. Y., and Li, X. S. (2012). Generalized propensity score for estimating the average treatment effect of multiple treatments. *Stat Med*, 31(7):681–97.

Field, C. A. and Welsh, A. H. (2007). Bootstrapping clustered data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(3):369–390.

Firoozi, F., Lemiere, C., Beauchesne, M. F., Forget, A., and Blais, L. (2007). Development and validation of database indexes of asthma severity and control. *Thorax*, 62(7):581–7.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Global Initiative for Asthma (2016). Global Strategy for Asthma Management and Prevention. *http://www.ginasthma.org/*.

Goldenberg, R. L., Culhane, J. F., Iams, J. D., and Romero, R. (2008). Epidemiology and causes of preterm birth. *Lancet*, 371(9606):75–84.

Gregersen, T. L. and Ulrik, C. S. (2013). Safety of bronchodilators and corticosteroids for asthma during pregnancy: what we know and what we need to do better. *J Asthma Allergy*, 6:117–25.

Guo, S. and Fraser, M. W. (2014). *Propensity Score Analysis: Statistical Methods and Applications*, volume 11. Sage Publications.

Haeck, C. and Lefebvre, P. (2016). A simple recipe: The effect of a prenatal nutrition program on child health at birth. *Labour Economics*, 41:77–89.

Hirano, K. and Imbens, G. W. (2004). The propensity score with continuous treatments. In Gelman, A. and Meng, X.-L., editors, *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, pages 73–84. John Wiley & Sons.

Huppler Hullsiek, K. and Louis, T. A. (2002). Propensity score modeling strategies for the causal analysis of observational data. *Biostatistics*, 3(2):179–193.

Imai, K. and Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):243–263.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3):706–710.

Kluve, J., Schneider, H., Uhlendorff, A., and Zhao, Z. (2012). Evaluating continuous training programmes by using the generalized propensity score. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175(2):587–617.

Kramer, M. S., Platt, R. W., Wen, S. W., Joseph, K. S., Allen, A., Abrahamowicz, M., Blondel, B., and Breart, G. (2001). A new and improved population-based Canadian reference for birth weight for gestational age. *Pediatrics*, 108(2):E35.

Kurth, T., Walker, A. M., Glynn, R. J., Chan, K. A., Gaziano, J. M., Berger, K., and Robins, J. M. (2006). Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol*, 163(3):262–70.

Lechner, M. (2001). Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. In Lechner, M. and Pfeiffer, F., editors, *Econometric Evaluation of Labour Market Policies*, pages 43–58. Physica, Heidelberg.

Lefebvre, G., Delaney, J. A., and McClelland, R. L. (2014). Extending the Bayesian Adjustment for Confounding algorithm to binary treatment covariates to estimate the effect of smoking on carotid intima-media thickness: the multi-ethnic study of atherosclerosis. *Stat Med*, 33(16):2797–813.

Levy, R. A., Dos Santos, F. C., de Jesus, G. R., and de Jesus, N. R. (2015). Antiphospholipid antibodies and antiphospholipid syndrome during pregnancy: Diagnostic concepts. *Front Immunol*, 6:205.

Li, F., Zaslavsky, A. M., and Landrum, M. B. (2013). Propensity score weighting with multilevel data. *Stat Med*, 32(19):3373–87.

Lougheed, M. D., Lemiere, C., Ducharme, F. M., Licskai, C., Dell, S. D., Rowe, B. H., Fitzgerald, M., Leigh, R., Watson, W., and Boulet, L. P. (2012). Canadian Thoracic Society 2012 guideline update: diagnosis and management of asthma in preschoolers, children and adults. *Can Respir J*, 19(2):127–64.

Martel, M. J., Rey, E., Beauchesne, M. F., Perreault, S., Lefebvre, G., Forget, A., and Blais, L. (2005). Use of inhaled corticosteroids during pregnancy and risk of pregnancy induced hypertension: nested case–control study. *BMJ*, 330(7485):230.

McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R., and Burgette, L. F. (2013). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Stat Med*, 32(19):3388–414.

Min, Y. and Agresti, A. (2002). Modeling nonnegative data with clumping at zero: a survey. *Journal of Iranian Statistical Society*, 1(1):7–33.

Morgan, S. L. and Winship, C. (2015). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press, 2nd edition.

NAEPP expert panel report (2005). Managing asthma during pregnancy: recommendations for pharmacologic treatment–2004 update. *J Allergy Clin Immunol*, 115(1):34–46.

Namazy, J. A., Chambers, C., and Schatz, M. (2014). Safety of therapeutic options for treating asthma in pregnancy. *Expert Opin Drug Saf*, 13(12):1613–21.

Namazy, J. A., Schatz, M., Long, L., Lipkowitz, M., Lillie, M. A., Voss, M., Deitz, R. J., and Petitti, D. (2004). Use of inhaled steroids by pregnant asthmatic women does not reduce intrauterine growth. *J Allergy Clin Immunol*, 113(3):427–32.

Oken, E., Kleinman, K. P., Rich-Edwards, J., and Gillman, M. W. (2003). A nearly continuous measure of birth weight for gestational age using a United States national reference. *BMC Pediatr*, 3:6.

Rassen, J. A., Shelat, A. A., Franklin, J. M., Glynn, R. J., Solomon, D. H., and Schneeweiss, S. (2013). Matching by propensity score in cohort studies with three treatment groups. *Epidemiology*, 24(3):401–9.

Richiardi, L., Bellocco, R., and Zugna, D. (2013). Mediation analysis in epidemiology: methods, interpretation and bias. *Int J Epidemiol*, 42(5):1511–9.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

Rubin, D. B. (1978). Bayesian inference for causal effects: the role of randomization. *The Annals of statistics*, pages 34–58.

Rubin, D. B. (2010). Reflections stimulated by the comments of Shadish (2010) and West and Thoemmes (2010). *Psychol Methods*, 15(1):38–46.

Rubod, C., Deruelle, P., Le Goueff, F., Tunez, V., Fournier, M., and Subtil, D. (2007). Long–term prognosis for infants after massive fetomaternal hemorrhage. *Obstet Gynecol*, 110(2 Pt 1):256–60.

Schneeweiss, S., Glynn, R. J., Tsai, E. H., Avorn, J., and Solomon, D. H. (2005). Adjusting for unmeasured confounders in pharmacoepidemiologic claims data using external information: the example of COX2 inhibitors and myocardial infarction. *Epidemiology*, 16(1):17–24.

Shah, B. R., Laupacis, A., Hux, J. E., and Austin, P. C. (2005). Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. *J Clin Epidemiol*, 58(6):550–9.

Sheiner, E. (2007). The relationship between fetal gender and pregnancy outcome. *Arch Gynecol Obstet*, 275(5):317–9.

Spreeuwenberg, M. D., Bartak, A., Croon, M. A., Hagenaars, J. A., Busschbach, J. J., Andrea, H., Twisk, J., and Stijnen, T. (2010). The multiple propensity score as control for bias in the comparison of more than two treatment arms: an introduction from a case study in mental health. *Med Care*, 48(2):166–74.

StataCorp LP (2009). *Stata Multiple-Imputation Reference Manual: Release 11*. A Stata Press Publication, College Station,Texas.

Statistics Canada (Date Modified: 2010-10-06). Median 2005 earnings for full-year, full-time earners by education, both sexes, total - age group 25 to 64, for Canada, provinces and territories. http://www12.statcan.ca/census-recensement/2006/dp-pd/hlt/97-563/T803-eng.cfm?Lang=E&T=803&GH=4&SC=1&SO=99&O=A.

Tchernis, R., Horvitz-Lennon, M., and Normand, S. L. (2005). On the use of discrete choice models for causal inference. *Stat Med*, 24(14):2197–2212.

VanderWeele, T. J., Mumford, S. L., and Schisterman, E. F. (2012). Conditioning on intermediates in perinatal epidemiology. *Epidemiology*, 23(1):1–9.

Villar, J., Papageorghiou, A. T., Knight, H. E., Gravett, M. G., Iams, J., Waller, S. A., Kramer, M., Culhane, J. F., Barros, F. C., Conde-Agudelo, A., Bhutta, Z. A., and Goldenberg, R. L. (2012). The preterm birth syndrome: a prototype phenotypic classification. *Am J Obstet Gynecol*, 206(2):119–23.

Wang, C., Parmigiani, G., and Dominici, F. (2012). Bayesian effect estimation accounting for adjustment uncertainty. *Biometrics*, 68(3):661–71.

Wang, J., Donnan, P. T., Steinke, D., and MacDonald, T. M. (2001). The multiple propensity score for analysis of dose–response relationships in drug safety studies. *Pharmacoepidemiology and drug safety*, 10(2):105–111.

Westreich, D. and Cole, S. R. (2010). Invited commentary: positivity in practice. *Am J Epidemiol*, 171(6):674–7; discussion 678–81.

Westreich, D., Cole, S. R., Funk, M. J., Brookhart, M. A., and Sturmer, T. (2011). The role of the c-statistic in variable selection for propensity score models. *Pharmacoepidemiol Drug Saf*, 20(3):317–20.

Wilcox, A. J. (2001). On the importance–and the unimportance–of birthweight. *Int J Epidemiol*, 30(6):1233–41.

Table A.1: Inclusion of potential confounding covariates in full and reduced set analyses.

	Full set	Set 1[a]	Set 2[b]	Set 3[c]
Number of included variables	27	12	15	18
Mother's and baby's characteristics				
Maternal age at the beginning of pregnancy (< 18, > 18-34, > 34 years)	X	X	X	X
Baby's sex (male/female)	X	X	X	X
Receipt of social assistance [d] (yes/no)	X	X	X	X
Urban residency at delivery (yes/no)	X	X	X	X
Maternal chronic conditions [d]				
Antiphospholipid syndrome (yes/no)	X		X	X
Chronic hypertension (yes/no)	X			
Diabetes mellitus (yes/no)	X			
Cystic fibrosis of the pancreas (yes/no)	X			
Uterine defects (yes/no)	X			
Pregnancy-related variables				
Gestational diabetes (yes/no)	X	X	X	X
Eclampsia/pre-eclampsia (yes/no)	X	X	X	X
Anaemia during pregnancy (yes/no)	X		X	X
Placental conditions [e] (yes/no)	X	X	X	X
Placenta abruption (yes/no)	X	X	X	X
Vaginal bleeding (yes/no)	X	X	X	X
Maternal infections during pregnancy [f] (yes/no)	X			
Fetal-maternal hemorrhage (yes/no)	X	X	X	
Pregnancy-induced hypertension (yes/no)	X			
Use of beta-blockers during pregnancy (yes/no)	X			
Asthma-related variables				
Leukotriene-receptor antagonists during pregnancy (yes/no)	X			X
SABA during pregnancy (0, > 0-3, >3 doses/week)	X	X	X	X
Oral corticosteroids during pregnancy (yes/no)	X			X
Intranasal corticosteroids during pregnancy (yes/no)	X			X
≥ 1 ED visit for asthma during pregnancy (yes/no)	X			X
LABA during pregnancy (yes/no)	X	X	X	X
≥ hospitalization for asthma during pregnancy (yes/no)	X			
Severity of asthma in the year before conception [g] (mild, moderate, severe)	X		X	X

Abbreviations: BAC, Bayesian Adjustment for Confounding; BW, birth weight; ED, emergency department; LABA, long-acting beta2-agonists; SABA, short-acting beta2-agonists.

[a]: baby's sex and the covariates used by Cossette et al. (2013) in the adjusted analysis for low BW;
[b]: baby's sex and the covariates used at least in one model for the perinatal outcomes (low BW, preterm birth, small for gestational age) studied in Cossette et al. (2013);
[c]: covariates selected by applying the BAC algorithm as in Lefebvre et al. (2014);
[d]: in the year before or during pregnancy;
[e]: single umbilical artery, velamentous umbilical cord insertion, bilobate placenta, suboptimal implantation site, placenta previa, placental anomalies, and cord anomalies: yes, if at least one condition is present; no, otherwise;
[f]: urinary-genital infections, malaria, trypanosomiasis, cytomegalovirus, toxoplasmosis, rubella, herpes virus: yes, if at least one condition is present; no, otherwise;
[g]: the levels of asthma severity were measured according to an algorithm developed and validated in Firoozi et al. (2007).

Table C.1: Adjusted mean differences based on conventional multivariable approach.

	Adjusted mean differences[a] of BW (grams)	95% CI [b]
Mother's and baby's characteristics		
Maternal age (years)		
$<$ 18	-93.01	-212.39, 26.98
18-34	34.85	-11.94, 81.88
$>$ 34	Reference	
Baby's sex (female vs male)	-132.72	-157.75, -106.98
Receipt of social assistance	-127.47	-154.35, -100.87
Urban residency	52.28	16.29, 87.63
Maternal chronic conditions		
Antiphospholipid syndrome	-322.33	-520.44, -132.36
Chronic hypertension	-26.33	-112.01, 55.71
Diabetes mellitus	100.58	19.02, 180.12
Cystic fibrosis of the pancreas	-76.68	-242.83, 83.61
Uterine defects	-24.71	-67.34, 19.02
Pregnancy-related variables		
Gestational diabetes	92.04	47.60, 137.18
Eclampsia/pre-eclampsia	-268.27	-376.66, -159.27
Anaemia	70.26	30.61, 110.07
Placental conditions	-149.95	-225.58, -77.17
Placenta abruption	-364.03	-469.12, -258.95
Vaginal bleeding	-73.41	-122.32, -25.12
Maternal infections	-7.71	-45.02, 30.55
Fetal-maternal hemorrhage	-271.99	-779.59, 182.86
Pregnancy-induced hypertension	29.63	-31.69, 89.08
Use of beta-blockers	-99.71	-285.02, 79.43
Asthma-related variables		
Leukotriene-receptor antagonists	1.39	-111.60, 112.15
SABA (doses/week)		
0	33.57	-9.19, 74.91
$>$ 0-3	33.89	-2.72, 70.30
$>$ 3	Reference	
Oral corticosteroids	-13.42	-60.42, 31.85
Intranasal corticosteroids	21.51	-20.28, 60.94
$\geq$ 1 ED visit for asthma	37.99	-2.96, 79.43
LABA	17.51	-35.72, 71.67
$\geq$ 1 hospitalization for asthma	-65.90	-183.16, 51.55
Severity of asthma prior to pregnancy		
mild	53.21	-10.36, 116.99
moderate	59.63	-5.76, 123.80
severe	Reference	

Abbreviations: BW, birth weight; CI, confidence interval; ED, emergency department; LABA, long-acting beta2-agonists; SABA, short-acting beta2-agonists.

[a]: adjusted regression model (conventional multivariable approach) based on ICS exposure and the full set of covariates;

[b]: confidence intervals calculated using cluster bootstrap (Field and Welsh, 2007).

Table D.1: Average treatment effects (primary analyses, 7374 pregnancies) estimated with different strategies for variables selection using multilevel GPS approach.

	Categories of ICS doses compared (μg/day)					
	> 0-125 vs 0 (None)		> 125-250 vs 0 (None)		> 250 vs 0 (None)	
	ATE	95% CI[a]	ATE	95% CI[a]	ATE	95% CI[a]
GPS analysis (set 1[b])	22.49	-12.43, 57.86	-16.55	-97.70, 61.40	1.13	-67.63, 73.68
GPS analysis (set 2[c])	26.44	-8.80, 62.23	-0.60	-78.71, 74.20	-32.00	-119.44, 59.31
GPS analysis (set 3[d])	26.41	-9.60, 63.00	6.25	-64.87, 76.70	-36.51	-123.31, 47.00
Adjusted regression (set 1[b])	20.12	-11.45, 52.46	19.51	-37.26, 75.81	-54.64	-118.14, 12.01
Adjusted regression (set 2[c])	20.53	-10.71, 52.83	19.13	-38.11, 75.11	-45.44	-111.07, 23.66
Adjusted regression (set 3[d])	17.87	-14.51, 50.87	14.46	-43.32, 70.70	-51.89	-120.32, 13.04

Abbreviations: ATE, average treatment effect (grams); BAC, Bayesian Adjustment for Confounding; BW, birth weight; CI, confidence interval; GPS, generalized propensity score.
[a]: confidence intervals calculated using cluster bootstrap (Field and Welsh, 2007);
[b]: baby's sex and the covariates used by Cossette et al. (2013) in the adjusted analysis for low BW;
[c]: baby's sex and the covariates used at least in one model for the perinatal outcomes studied in Cossette et al. (2013);
[d]: covariates selected by applying the BAC algorithm as in Lefebvre et al. (2014).

Table E.1: Average treatment effects (primary analyses, 7374 pregnancies) estimated with different strategies for variables selection using continuous GPS approach: difference in the dose-response function values estimated at the medians of compared ICS-dose categories.

	Categories of ICS doses compared (μg/day)					
	> 0-125 vs 0 (None)		> 125-250 vs 0 (None)		> 250 vs 0 (None)	
	ATE[a]	95% CI[b]	ATE[a]	95% CI[b]	ATE[a]	95% CI[b]
Set 1[c]	-8.75	-22.45, 6.26	-27.04	-61.57, 11.66	-68.02	-140.57, 9.84
Set 2[d]	-7.76	-21.65, 7.18	-22.93	-58.57, 15.68	-55.77	-128.85, 23.33
Set 3[e]	-9.78	-23.38, 4.69	-28.70	-63.99, 8.93	-72.16	-144.82, 2.07

Abbreviations: ATE, average treatment effect (grams); BAC, Bayesian Adjustment for Confounding; BW, birth weight; CI, confidence interval; GPS, generalized propensity score.
[a]: the median values are, respectively. 0, 57.65, 176.20 and 392.43 μg/day;
[b]: confidence intervals calculated using cluster bootstrap (Field and Welsh, 2007);
[c]: baby's sex and the covariates used by Cossette et al. (2013) in the adjusted analysis for low BW;
[d]: baby's sex and the covariates used at least in one model for the perinatal outcomes studied in Cossette et al. (2013);
[e]: covariates selected by applying the BAC algorithm as in Lefebvre et al. (2014).