Birth Order, Birth Weight and Asthma: how to assess mediation and the presence of Unmeasured Confounding

Ordine di nascita, peso alla nascita e asma: come valutare la mediazione e la presenza di confondimenti non misurati

Rossella Murtas, Bianca DeStavola, Daniela Zugna, Rhian Daniel, Lorenzo Richiardi

Abstract If the goal is to measure the causal relation between two variables when a third is involved and plays the role of mediator, it is essential to explicitly define the relational assumptions among these and others relevant variables. However if any of this assumptions are not met, estimates of mediated effects may be affected by bias. One example where this occurs is the widely discussed situation known as the *Birth Weight paradox* which arises when birth weight is the mediator of interest and unmeasured confounders affect the mediator-outcome relation. In this paper we will focus on a setting where such paradox might arise where birth order act as the exposure for childhood asthma (by age 1.5). After estimating the direct and indirect effects, we give a plausible graphical explanation of the empirical results and explore the magnitude of the causal relations using sensitivity analysis.
Abstract *Quando l'obiettivo è studiare la relazione causale tra due variabili mediata dal basso peso alla nascita spesso ci si imbatte in risultati paradossali. In questo articolo forniamo una spiegazione grafica a tale paradosso nel caso in cui l'esposizione è l'ordine di nascita e l'outcome l'asma infantile.*

Key words: Mediation, Birth Weight Paradox, Unmeasured confounding

1 Aims, Materials and Methods

This work is motivated by the NINFEA study, an Italian web-based birth cohort of 6445 pregnant women started in 2005 in Turin and successively extended to the rest of the country from December 2007. The voluntary recruitment into the cohort begins during pregnancy with a first questionnaires web administered [2], with other follow up questionnaires planned at 6 and 18 months after delivery and when

Rossella Murtas University of Cagliari, Mathematical Department, e-mail: rmurtas@unica.it · Bianca DeStavola Centre for Statistical Methodology, London School of Hygiene and Tropical Medicine e-mail: bianca.destavola@lshtm.ac.uk

the children are 4 years old and 7 years old. The questionnaires contain questions on parent's and children's exposures and outcomes including environmental (smoking, traffic etc.), demographic (age, weight, height), medical (any sort of maternal disease like asthma, diabetes etc.) and occupational factors. The main goal of this project is to study the effect of prenatal exposures on recurrent wheezing or asthma by age 1.5 years and whether this is mediated by birth weight.

1.1 Background

Let's denote for simplicity X the exposure, Y the outcome, M the mediator of interest, C a set of baseline covariates and A a risk factor for Y. Mediation studies the relationship between two variables when a third is involved in the causal pathway. Let be $Y(x)$ the potential value that Y would take if X was being set to x, $M(x)$ the potential value that M would take if X was being set to x, $Y(x,m)$ the potential values that Y would take if X was set to the value x and M to m, $Y(x,M(\tilde{x}))$ the composite potential value that Y would take if X was set to x and M to $M(\tilde{x})$ (Rubin 1974 [3]). The term counterfactual means that these outcomes represent situations that may not actually occur (they may be counter to the fact).

If X, Y and M are dichotomic variables we can express effects in terms of risk ratios. Following Pearl 2001 [5]: the *Total Causal Effect* of X on Y is defined as $TCE = E\{Y(1)\}/E\{Y(0)\}$, the *Controlled Direct Effect* of X on Y when M is controlled at m is defined as $CDE(m) = E\{Y(1,m)\}/E\{Y(0,m)\}$, the *Pure Natural Direct Effect* of X on Y is $PNDE = E\{Y(1,M(0))\}/E\{Y(0,M(0))\}$ and the *Total Natural Indirect Effect* of X on Y is $TNIE = TCE/PNDE = E\{Y(1,M(1))\}/E\{Y(1,M(0))\}$. The CDE is able to quantify the sensitivity of Y to changes in X while the mediating factor (M) is controlled. The natural effects $PNDE$ and $TNIE$ are able to decompose the total causal effect of X on Y into two different effects: one mediated by M and one not mediated by M. All this definitions could be generalized to linear contexts simply defining them in terms of differences instead of ratios. Under certain assumptions we could identify the above quantities the observed data. For example for the Natural effects these are: no interference between units (subjects) of exposure and mediator on their relative outcomes, consistency (the potential variable $Y(x)$ must be equal to the observed outcome Y when the exposure is observed at that value x), no unmeasured confounding for the $X-M$, $X-Y$ and $M-Y$ relations (conditional on C) and no intermediate confounders. No unmeasured confounding for the exposure-outcome relation means that if the exposure is randomized, there is no loss to follow-up, no contamination and complete adherence, then $E(Y(x)) \equiv E(Y|X=x)$.

In the simplex case of M and Y dichotomic parametric specifications of the models for the diagram of Fig. 1 are:

$$logit[P(Y=1|X=x,M=m,C=c,A=a)] = \theta_0 + \theta_x x + \theta_m m + \theta_{xm} x \cdot m + \theta_c c + \theta_a a$$

$$logit[P(M=1|X=x,C=c)] = \beta_0 + \beta_x x + \beta_c c \quad (1)$$

1.2 An interesting problem

In 2006 S. Hernndez-Diaz et al. [1], discussed that if M in the model 1 is birth weight, conditioning on it could lead to paradoxical results. In their example infants born to smokers have higher risks of death and to be low birth weight (LBW less than 2500g) than infants born to nonsmokers, but in the LBW stratum maternal smoking appears not to be harmful for infants mortality. They suggested the presence of unmeasured $M-Y$ confounding as the reason of this paradox. The first aim of this report is to assess whether there is evidence of such paradox in the NINFEA dataset when the outcome of interest is recurrent wheezing or asthma up to 18 months of age (diagnosed from the doctor) the exposure is Birth Order dichotomized to be zero for first child and one otherwise and the mediator is Birth Weight dichotomized to be one for Low Birth Weight Infants (LBW less than 2500g) zero otherwise. The second aim is to estimate the direct and indirect effect of the exposure on the outcome and finally provide an alternative explanation to the observed paradox.

2 Preliminary Analysis on the NINFEA dataset

Among the 3900 children participating to NINFEA (at May 2013) we selected 3541 children with complete records. Among this children the outcome has a prevalence of 5.73% while the prevalence of first born children is 75,54% and the prevalence of Low birth weight is 4.74%.

All our analyses will be made referring to the DAG in figure 1 where we included as baseline covariates C child's year of birth (centered at 2009), maternal year of birth (centered at 1976) and maternal age (centered at 33). Maternal Asthma was a significant risk for infant asthma $OR = 1.73$ CI 95% (1.11;2.71), suggesting a possible genetic contribution.

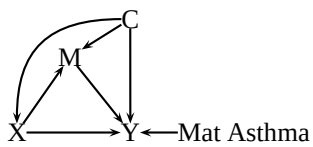

Fig. 1 Final Mediation DAG

Adjusting for C the exposure is strongly associated both with asthma and birth weight: X (Birth Order 2+) is harmful for asthma (OR=3.11 CI 95% (2.3;4.2)) and protective for low birth weight (OR=0.64 CI 95% (0.44;0.99)). On the other hand low birth weight seems to be a risk factor for asthma but it is non-significant

(OR=1.28 CI 95% (0.7;2.35)). Fitting a logistic regression model of Y with covariates X, M, XM and C the LBW ORs shown that X is a significantly harmful for normal birth weight infants (OR=3.3 CI 95% (2.4;4.5)) but has non-significant association in the low birth weight group (OR=1.01 CI 95% (0.21;4.9)). As pointed by VanderWeele et al. in [6] results such as these point towards an interaction between the exposure and the mediator which is likely to be a consequence of unmeasured $M-Y$ confounding. In fact given the general harmful association of birth order on asthma, conditioning on birth weight leads to a protective effect of X on Y in the low birth weight group and an harmful effect in the normal birth weight group.

3 Partitioning the causal effect

In order to estimate the controlled and the natural effects defined in Section 1.1 we used the approximate analytical formulas of Ananth and VanderWeele to the framework of logistic regression when the outcome is rare [7] that is indeed the setting considered in the DAG in Fig. 1 and assumed by the logistic models in equation 1.

From the models in 1 and the definitions introduced in section 1.1 we can specify the mediation quantities as risk ratio (the outcome is supposed to be rare so we can approximate the Odds Ratios as Risk Ratios) and estimate them using the results of Ananth end VanderWeele as follows (under the assumptions stated in 1.1):

$$CDE(0) = e^{\theta_x} \qquad CDE(1) = e^{\theta_x+\theta_{xm}} \qquad NDE = e^{\theta_x}\,\frac{1+e^{\theta_m+\theta_{xm}+\beta_0+\beta_c C}}{1+e^{\theta_m+\beta_0+\beta_c C}} \tag{2}$$

$$NIE = \frac{1+e^{\beta_0+\beta_c C}}{1+e^{\beta_0+\beta_x+\beta_c C}}\,\frac{1+e^{\theta_m+\theta_{xm}+\beta_0+\beta_x+\beta_c C}}{1+e^{\theta_m+\theta_{xm}+\beta_0+\beta_c C}} \tag{3}$$

The above statistics are estimated for the NINFEA data after fitting the model 1 by maximum likelihood estimation and inserting the relevant estimates in 2 and 3. Standard errors were obtained via bootstrap using the bias-corrected methods since in the bootstrap sample there was evidence of non-normality.

Th results show that there is evidence of causal effect of X on Y with $TCE = 3.16$ (CI 95% (2.3;4.2)) similar to the crude OR of 3.11 obtained in 3, but it is almost entirely attribute to the direct path $NDE = 3.14$ (CI 95% (2.3;4.1)) while $NIE = 1.03$ (CI 95% (0.98;1.02)). As pointed in the preliminary section 2, the exposure seems to act as a risk factor just in the normal weight group ($M = 0$) with $CDE(0) = 3.34$ CI 95% (2.5;4.4) while $CDE(1) = 1.03$ CI 95% (0.26;3.5).

4 An alternative explanation to the paradox

Section 3 supports absence of an indirect effect of X on Y via M. These results contradict the birth weight paradox shown in section 2: we would have expected that such paradox arises in the presence of indirect effect. It has been suggested in the literature that conditioning on an intermediate leads to paradoxical results in the case of unmeasured confounding U between M and Y: conditioning on the collider M will open a spurious direct path from X to Y able to delete the indirect effect of X on Y via M. According to the signs of the Odds Ratios estimated in section 2 it is clear that this unmeasured variable U should affect Y and M in the same direction for the negative indirect effect to be masked:

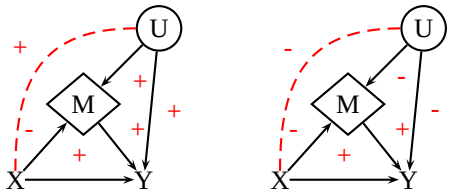

Fig.2 DAG with unmeasured confounding and signs

5 Conditioning on the risk of an intermediate

In 2012 VanderWeele et al. [6] proposed an alternative approach to dealing with such paradox. Instead of conditioning on the intermediate variable M in the model 1 they proposed to condition on the risk of the intermediate predicted by a set of baseline covariates, This approach will allow us to explore the magnitude of the association of X on Y among the group of infants who have an high risk to be low birth weight. The best candidates in terms of statistical and biological association are eclampsia, child's sex, maternal height and BMI (that we will call DET hereafter). This method will be able to avoid the paradox but it will not allow us to distinguish between direct and indirect effects. To apply this method we should first fit a logistic regression for M on DET and C and then predict the probabilities of each individuals based on their covariate values. After that we define two new variables called H_1 e H_2 (H=high risk) such that H_1 is one for children who have predicted probabilities of being LBW above the 90th percentile, and zero otherwise, and H_2 is equal to one for children who have predicted probabilities of being LBW above the 95th percentile, zero otherwise. In the first case 33.33% of LBW infants have a high risk to be LBW and 17.86% in the second case. To examine the X-Y association without adjusting for the mediator we could fit a logistic model for Y on X, H, $H \cdot X$, C and A where H represent either H_1 or H_2 and A the variable of maternal Asthma:

$$logit(P(Y = 1|x,h,c,a)) = \alpha_0 + \alpha_1 x + \alpha_2 h + \alpha_3 hx + \alpha_4 c + \alpha_5 a \qquad (4)$$

In the case of $H = H_1$ the OR of asthma comparing being birth order 2+ vs 1 among babies at low risk of LBW is 3.36 (CI 95% (2.44;4.3)) and 2.12 (CI 95% (0.89;5.00)) among those at high risk of LBW. On the other hand if $H = H_2$ the OR of asthma comparing being birth order 2+ vs 1 is 3.32 (CI 95% (2.44;4.54)) among the low risk group and 1.58 (CI 95% (0.46;5.40)) in the high risk group. Even if those results do not show anymore a protective effect of X on Y in the high risk group, it is important to underline that this approach is affected by the choice of the predictors in the model for M and possibly the risk cut-points.

6 Discussion

In both section 2 and 3 we have shown that using LBW as a mediator leads to paradoxical results. In fact the exposure resulted to act as a risk factor just in the Normal Birth Weight group. On the other hand, as pointed in section 4, there was no indirect effect of birth order on asthma mediated by LBW. The most likely explanation is the presence of unmeasured confounding between M and Y that distorts all previous analyses. Conditioning on M will open a bias path between X and U able to delete the indirect effect. Furthermore this unmeasured variable U should affect M and Y in the same direction for the negative indirect effect to be masked: it should be a risk factor (or a protection) for both simultaneously. Conditioning on the risk of being LBW, instead of conditioning on M itself, could be a possible alternative to explore the ”real” magnitude of the causal association but it is not able to differentiate between direct and indirect effects. It could be done repeating mediation analysis in the group of high risk infants. Further aims include analyzing in which situation the unmeasured confounding U is able to produce the birth weight paradox and in particular how strong the relationships between U and M and between U and Y should be to mask the indirect effect between X and Y via M.

References

1. S. Hernndez-Diaz, EF. Schisterman, MA. Hernn. *The birth-weight ”paradox” uncovered?* Am J Epidemiol. 2006; 164:1115-1120.
2. *NINFEA project* URL: https://www.progettoninfea.it/
3. Rubin, D. B. *Estimating causal effects of treatments in randomized and non-randomized studies* J. Educ. Psychol. 1974 66, 688701.
4. Holland P. *Statistics and causal inference* J. Amer. Statist. Assoc. 81, 945960 (1986).
5. J. Pearl. Direct and Indirect Effects, UCLA Cognitive Systems Laboratory, Technical Report (R-273), June 2001. In Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, San Francisco, CA: Morgan Kaufmann, 411-420, 2001.
6. T. VanderWeele et all. *Conditioning on intermediates in Perinatal Epidemiology* Epidemiology 2012; 23:1-9.
7. Cande V. Ananth, T. VanderWeele. *Placental Abruption and Perinatal Mortality with Preterm Delivery as a Mediator: Disentangling Direct and Indirect Effects* Am J Epidemiol. 2011 Jul 1;174(1):99-108.