Title

Deconstructing the Smoking-Preeclampsia Paradox through a Counterfactual Framework

Authors

Miguel Angel Luque-Fernandez[1,2*], Helga Zoega[3], Unnur Valdimarsdottir[1,3], Michelle A. Williams[2],

Running title

Smoking-Preeclampsia Paradox

Authors Affiliations

1. Department of Non-Communicable Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK
2. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
3. Center of Public Health Sciences, Faculty of Medicine, University of Iceland, Reykjavik, Iceland

***Corresponding author**

Miguel Angel Luque-Fernandez
Department of Non-Communicable Disease Epidemiology, Faculty of Epidemiology and Population Health
London School of Hygiene and Tropical Medicine
Keppel Street, London, WC1E 7HT
+44 02079588162
miguel-angel.luque@lshtm.ac.uk

Word count: Abstract: 304; Text: 3,664; **References**: 75; **Tables**: 3; **Figures**: 5

ABSTRACT

Background:

Although smoking during pregnancy may lead to many adverse outcomes, numerous studies have reported a paradoxical inverse association between maternal cigarette smoking during pregnancy and preeclampsia. Using a counterfactual framework we aimed to explore the structure of this paradox as being a consequence of selection bias.

Methods:

Using a case-control study nested in the Icelandic Birth Registry (1,309 women), we show how this selection bias can be explored and corrected for. Cases were defined as any case of pregnancy induced hypertension or preeclampsia occurring after 20 weeks' gestation and controls as normotensive mothers who gave birth in the same year. First, we used directed acyclic graphs to illustrate the common bias structure. Second, we used classical logistic regression and mediation analytic methods for dichotomous outcomes to explore the structure of the bias. Lastly, we performed both deterministic and probabilistic sensitivity analysis to estimate the amount of bias due to an uncontrolled confounder and corrected for it.

Results:

The biased effect of smoking was estimated to reduce the odds of preeclampsia by 28% (OR=0.72, 95%CI: 0.52, 0.99) and after stratification by gestational age at delivery (<37 vs. ≥37 gestation weeks) by 75% (OR=0.25, 95%CI: 0.10, 0.68). In a mediation analysis, the natural indirect effect showed and OR >1, revealing the structure of the paradox. The bias-adjusted estimation of the smoking effect on preeclampsia showed an OR of 1.22 (95%CI: 0.41, 6.53).

Discussion:

The smoking-preeclampsia paradox appears to be an example of (1) selection bias most likely caused by studying cases prevalent at birth rather than all incident cases from conception in a pregnancy cohort, (2) omitting important confounders associated with both smoking and preeclampsia (preventing the outcome to develop) and (3) controlling implicitly or explicitly for a collider (gestation weeks at delivery). Future studies need to consider these aspects when studying and interpreting the association between smoking and pregnancy outcomes.

Keywords: Preeclampsia; Smoking; Selection Bias; Epidemiology Methods; Perinatal Mortality

BACKGROUND

Preeclampsia complicates between 2 and 8% of all pregnancies and is a leading cause of maternal and infant morbidity and mortality worldwide.[1,2] Although smoking during pregnancy may lead to many adverse maternal and perinatal outcomes, numerous studies have reported a paradoxical inverse association between cigarette smoking during pregnancy and preeclampsia in different populations, namely the "smoking-preeclampsia paradox".[3-7]

In the epidemiological literature different explanations have been proposed to describe the paradoxical protective effect of established risk factors (e.g., smoking) on an outcome.[8-10] Depending on the way that the association between a risk factor and an outcome is induced, this paradoxical protective effect has been referred to as collider bias, gestational age paradox, incidence-prevalence bias, competing risk and Neyman bias[11-13] yet these are all special cases of a general type of bias named "selection bias".[14] Selection biases leading to paradoxical effects have been described as biases arising from inappropriate selection of study subjects from the source population.[15]

In the particular case of the smoking-preeclampsia paradox, many studies documenting the inverse association used case-control designs with prevalent cases and omitted important factors associated with both smoking and preeclampsia (e.g., fetoplacental pathologies associated with preterm delivery)[3,4,7,16-20]. Such unmeasured confounders would make gestational age at delivery a collider. Thus, controlling for the collider (gestational age at delivery) will cause the association between smoking and preeclampsia to be biased.[9,21]

We hypothesized that the smoking-preeclampsia paradox, is an example of selection bias caused by (1) studying cases prevalent at birth rather than all incident cases in a conception or pregnancy cohort, (2) omitting important confounders associated with both smoking and preeclampsia, and (3) controlling for gestational age at delivery.

In this article, we first used Direct Acyclic Grasps (DAGs) to illustrate the common bias structure. Next, we demonstrate how to correct for this selection bias using data from a case-control study nested within the Icelandic Medical Birth Register.

METHODS

To evaluate the effect of the smoking-preeclampsia selection bias, we used data from a population-based case-control study nested within all pregnancies in Iceland 1989-2004, resulting in birth at the Landspitali University Hospital.[22] That study aimed to investigate the combined effects of obesity and smoking on hypertensive disorders during pregnancy. Any case of hypertensive disorder during pregnancy was selected retrospectively from the electronic National Medical Birth Registry based on the International Classification of Disease, 10th Revision [ICD-10][23] codes, O10-16.

For the present study, we defined preeclampsia as any case of pregnancy induced hypertension (O13, O16) or preeclampsia (O14, O15, O11) occurring after 20 weeks' gestation. We excluded cases of preexisting hypertension (O10) and restricted the analysis to women with only one singleton pregnancy. The final dataset contained 376 preeclampsia prevalent cases matched on year of delivery to 933 normotensive mothers (N = 1,382) (Figure 1).

We included the following variables: maternal age at delivery, early pregnancy BMI, parity, gestational age at delivery, smoking during pregnancy, and preeclampsia status. Weight (kg) and height (m) were measured at the first prenatal visit, which occurred on average at pregnancy week 13 (median: 13.8, interquartile range: 12.0-15.3). Early pregnancy BMI was categorized according to international standards as kg/m^2: underweight (BMI <18.5), normal weight (BMI <18.5-24.9), overweight (BMI 25.0-29.9), obese (BMI ≥30.0). We merged underweight and normal weight given the reduced number of underweight women (n= 22) in analysis. We dichotomized maternal age into ≥35 years vs. <35, parity into multiparous (≥2 previous deliveries) and nulliparous (first delivery) women, and gestational age at delivery (preterm <37 weeks vs. term ≥37 weeks). Finally, maternal self-report of smoking during pregnancy was dichotomized (yes vs. no). Women who self-reported having quit smoking at the first prenatal visit were considered non-smokers (n= 37). Information on quantity or type of smoking was not available.

Using DAGs we described the smoking-preeclampsia paradox, then we used a mediation analysis to explore the structure of the bias, and finally we developed both deterministic and probabilistic sensitivity analysis to estimate the amount of bias and corrected for it.[21,24-27]

Notations

We employed notations of counterfactuals.[28] We denoted ***A*** our exposure of interest (e.g., maternal smoking status during pregnancy), ***M*** represented the mediator (e.g., preterm delivery, <37 gestation weeks) and ***Y*** represented the primary outcome of interest (e.g., preeclampsia). We used ***C*** to denote some set of baseline characteristics measured before or concurrent with the exposure ***A,*** and ***U*** some unmeasured confounders (e.g., miscarriage, fetal death, fetoplacental pathologies associated with preterm delivery, etc.). Finally, we used $\boldsymbol{Y_a}$ to denote the counterfactual outcome for each individual if the exposure had been set to level ***a*** and $\boldsymbol{M_a}$ denoted the intermediate if the exposure had been set to level ***a***.

For statistical analyses we used Stata v.13.1 (Statacorp, College Station, Texas, U.S.) and the user written macro "paramed".[29,30]

RESULTS

Structure of the paradox*: simulated scenarios*

Case-control studies nested within a birth cohort, with a density incidence sampling design and incident cases, require precise information regarding the onset time of both exposure and outcome. Accurate information regarding time is needed to prevent that the start of follow-up and the end of the study coincide (Figure 2).[31,32] Neglecting this principle is the source of the occurrence of incidence-prevalence biases in cases-control studies nested within a birth cohort.[13,31,32,33] The use of prevalent cases in case-control studies makes it impossible to account for the association between the onset time of preeclampsia and any other measured and unmeasured factors that avert the development of preeclampsia (e.g., fetoplacental pathologies associated with preterm delivery).

Using recent preeclampsia incidence estimates by gestation week published elsewhere,[34] we sought to depict in a simple figure the smoking preeclampsia paradox (Figure 2). Figure 2 shows a simulated scenario of the smoking-preeclampsia paradox. The curve represents preeclampsia incidence rates by gestation weeks and dashed, and solid lines represent time to delivery for cigarette smokers (n = 5) and non-smokers (n= 5). In this example, preterm deliveries among smokers avert the further development of preeclampsia. However, depending on the type of study design and the use of prevalent or incident cases, the point estimated of the effect of smoking on preeclampsia varies from 0.4 to 1.05.

As shown in figure 2A, using a classical case-control study, without accounting for time, shows a protective association between maternal smoking and preeclampsia. Figure 2B depicts the scenario of a case-control study with a density sampling design. Under this scenario, the use of prevalent cases to account for person-time at risk also shows a protective association. The onset time of the outcome is not ascertained and the time at risk coincides with the duration of the gestation. Therefore, this represents conditioning implicitly on gestation weeks at delivery. Finally, figure 2C shows the inversion of the effect between maternal smoking and preeclampsia when the study design accounts for the correct estimate of person-time at risk using incident cases in a cohort study design where the onset time of the outcome is ascertained.

Descriptive results with empirical data: *presenting the paradox*

Women with preeclampsia were slightly older, more likely to be overweight and obese and to be nulliparous as compared with normotensive controls. The frequency of smoking among cases was lower (15% vs. 20%), and the odds of preeclampsia were reduced among smokers as compared with non-smokers (odds ratio (OR) = 0.72, 95% confidence intervals (CI): 0.52, 0.99) (Table 1). Overall, smokers and non-smokers were similar with regards to maternal age, early pregnancy BMI, and parity (Table 2).

DAGs and analytic results: *unveiling the paradox*

Figure 3 depicts the structure of the classical epidemiologic approach to estimate unadjusted and adjusted total effect of smoking (***A***) on preeclampsia (***Y***) using logistic regression models. The effect of smoking on preeclampsia was estimated to be protective with a 28% reduction of odds among smokers compared with non-smoker women. Because we were attempting to estimate the effect of maternal smoking exposure on the risk of preeclampsia by using prevalent cases, we were conditioning implicitly on been born at “x” specific gestation weeks. This is an example of collider bias where conditioning on being born at “x” specific gestation weeks, ***U*** (placental pathology) becomes a confounder of the smoking-preeclampsia association and induces bias.

Figure 4, describes the collider stratification effect that occurs after stratification by gestational age at delivery. The protective effect of maternal smoking (***A***) on the risk of preeclampsia increased among preterm infants (<37 gestation weeks) with 75% of odds reduction (adjusted OR = 0.25, 95%CI: 0.10, 0.68). Stratification by term and preterm delivery status creates a new association between maternal

smoking (***A***) and unmeasured confounders (***U***) because maternal smoking and unmeasured confounders are both associated with preterm delivery and preeclampsia.

To disentangle the multiple pathways that may explain the association of an exposure (***A***) on an outcome (***Y***), we applied a mediation analysis to assess the extent to which the effect of an exposure is explained, or is not, by an intermediate variable or mediator (***M***).[35] Figure 5 depicts the scenario when smoking (***A)***, for each woman would have two possible counterfactual outcomes, Y_1 and Y_0, corresponding to what would have happened to the woman (e.g. with or without smoking). Likewise, we had two possible counterfactual intermediates M_1 (delivery before 37 gestation weeks) and M_0 (delivery after 37 gestation weeks). For each woman, we were able to observe only one of Y_1 or Y_0 corresponding to the exposure that was in fact received; and likewise for M_1 and M_0. The mediation analysis showed the structure of the bias, assuming the presence of unmeasured confounders (i.e. placental pathology) for the mediator-outcome relationship (i.e. preterm delivery and preeclampsia). According to the causal graph theory, conditioning on the mediator ***M*** (i.e. preterm delivery) induces a spurious association between the mediator-outcome confounder ***U*** (i.e. placental pathology) and the exposure ***A*** (maternal smoking), and induces bias. Therefore, the estimated total effect of (***A***) on (***Y***) is biased (Figure 5).[9]

In the mediation analysis, we defined the natural indirect effect (NIE) as the contrast between the counterfactual outcome when the exposure is fixed at ***A*** = 1 (smoking) comparing the effects if the distribution of the mediator were set to what it would have been with preterm versus term delivery ($Y_{1,M(1)} - Y_{1,M(0)}$). Intuitively, the natural indirect effect captures the effect of the exposure. Thus, if smoking positively affects the mediator (***M***) and the supposed mediator-outcome confounder (***U***) is positively associated with both the outcome and the mediator, the direct effect for a given level of ***M*** is likely to be biased downwards, corresponding to an apparent inverse direct effect of maternal smoking on preeclampsia.[9,35]

Figure 5 reveals how the direct effect is biased downwards showing an apparent protective effect. However, in contrast, to classical and stratified analyses (Figures 3 and 4), the NIE in our empirical example showed an OR >1 for the association between maternal smoking and preeclampsia mediated by preterm delivery, thereby revealing the structure of the bias.

Sensitivity analysis

Finally, to address this bias, we developed a sensitivity analysis to assess how such an unmeasured common cause (***U***) of the intermediate (***M***) and the outcome (***Y***) might affect the total effect estimated.[26,27,36]

Let **γ** denote the odds of preeclampsia (***Y***) comparing ***U*** = 1 and ***U*** = 0 conditional on smoking exposure (***A***), preterm delivery (***M***), and let π_{am} denote the prevalence of ***U*** among those with smoking status ***a*** =1 (smokers) or ***a*** = 0 (non-smokers) and prematurity status ***m*** (***m*** = 1 or ***m*** = 0). If ***U*** is associated with preeclampsia (***Y***) by the same factor that it is associated with preterm and term deliveries, then the ratio between the estimate not controlling for ***U*** and the estimate that would have been obtained after controlling for ***U*** is given by:[26]

$$B = \frac{1+(\gamma - 1)\pi_{1m}}{1+(\gamma - 1)\pi_{0m}} \quad (1)$$

The bias-adjusted OR (that would have been obtained when adjusting for ***U***) could be estimated by dividing the estimated odds ratio by the bias factor *B*.[52] For our sensitivity analysis we used fetal death as a surrogate of placental pathology. Fetal deaths among smokers are 1 to 4 times higher than among non-smokers[37,38] and odds ratios for the association between placental pathology with fetal death and preeclampsia ranges between 2 and 86.[34]

We fixed the prevalence of ***U*** (i.e., fetal death) for smokers with preeclampsia to π_{1m} = 0.08 and π_{0m} = 0.30 for non-smokers (RR = 3.7). We based our choice on previous evidence regarding the association between fetal deaths and maternal smoking (three times higher among smokers).[39,40] We fixed π_{0m} = 0.30, because if smoking were not the cause of fetal death, this renders some other explanation more likely, such as fetal malformations, maternal medical conditions, placental pathology.[9]

In the empirical example, we have shown that the odds of preeclampsia were reduced among smokers as compared with non-smokers (OR = 0.72 95%CI: 0.52, 0.99). However, if the association between ***Y*** (preeclampsia) and ***U*** (i.e., fetal death) were a 3.5-fold increase (***γ*** = 3.5),[34] we would have had a bias factor ***B*** in equation (1) of 0.69. Hence, the bias-adjusted OR for the association between preeclampsia and maternal smoking showed a positive association (Bias-adjusted OR = 0.72/0.69 = 1.05, 95%CI: 0.50, 2.00) (Table 3).

Table 3 shows the result of bias-adjusted OR for different deterministic simulated scenarios. Furthermore, based on equation (1) we developed a probabilistic sensitivity analysis using Monte Carlo[41] simulations (10,000 replications) to bias adjust different OR estimates for values of π_{1m} ranging between 0.01 and 0.08, π_{0m} between 0.01 and 0.30, and γ (effect of **U** on **Y**) between 1 and 5.5. Based on these assumptions the probabilistic bias-adjusted OR was 1.22, (95%CI: 0.41, 6.53).

DISCUSSION

Using DAGs and mediation causal analyses we unveiled the structure and influence of selection bias (namely collider bias) on the association between maternal smoking during pregnancy and preeclampsia risk. Furthermore, we showed how potential unmeasured confounders may distort the analyses that condition on gestational age at birth. Finally, using a sensitivity analysis to address this bias, we showed how the inverse association of smoking on preeclampsia shifted from 28% risk reduction to a non-significant bias-adjusted effect of 22% risk increase of preeclampsia for smokers as compared with non-smokers.

The etiology of preeclampsia is not well understood, but several risk factors have been identified such as genetic factors,[42,43] nulliparity, multifetal gestations, maternal race, age, and pre-existing conditions such as pre-gestational hypertension, diabetes, kidney disease, obesity, and a prior history of preeclampsia.[44] However, even if smoking during pregnancy is recognized as the most important preventable risk factor for many adverse pregnancy and perinatal outcomes,[45-49] some authors have presented a paradoxical protective effect of smoking on preeclampsia.[3,4,7,16-20,50] The paradoxical risk reduction ranges from 10% to 50%[7] and an inverse dose-response relationship has also been reported.[50]

Overwhelming evidence supports the conclusion that cigarette smoking causes various adverse cardiovascular events including hypertension.[51-53] However, several underlying biological mechanism for the protective smoking effect on preeclampsia have been hypothesized (i.e., carbon monoxide-mediated inhibition of inflammation,[54,55] enhanced vasodilation,[56] suppression of platelet aggregation,[57] plasminogen activation,[54,58] apoptosis,[59] antiangiogenic factor[60,61] and, soluble fms-like tyrosine kinase-1[55,61]). These mechanisms remain poorly understood and are based on observational data with inconsistent evidence of questionable causal interpretation.[7,50]

Furthermore, establishing a clear biologic mechanism to explain the smoking-preeclampsia paradox has been complicated by the observation that active maternal smoking is also strongly associated with increased risk of perinatal conditions that end the pregnancy prematurely, such as fetal death and fetoplacental pathology associated with preterm or elary term delivery.[62-64] Thus, the premature end of the pregnancy avert further preeclampsia development.

To appreciate the statistical artifact that promotes this important perinatal epidemiological smoking-preeclampsia paradox, we invite readers to consider an example in cardiovascular epidemiology.[13] Suppose that a case-control study is carried out to investigate the relation between smoking and acute myocardial infarction (AMI), with cases interviewed one week after the coronary attack (prevalent cases). If AMI patients who are smokers die more frequently than AMI patients who are non-smokers, the analyses of surviving AMI cases will show a lower frequency of smoking, underestimating the true association of smoking with incident AMI.[12,13] The same holds for the smoking-preeclampsia paradox, where smoking is associated with miscarriage, fetal death and preterm delivery. It follows that pregnant women who are smokers and who experience higher risks of conditions related to early pregnancy termination cannot then have their pregnancies complicated by preeclampsia.[8]

The majority of the evidence supporting the protective effect of smoking on preeclampsia is based on prevalent cases assessed at birth.[3,4,7,16-20,50] An important consideration in designing a case-control study is whether or not it is possible to include incident or prevalent cases (i.e., prevalent cases of preeclampsia are ascertained at birth among the pregnancies that have survived from the time of its conception until birth).[65] The inability to fully account for outcomes of all conceptions because of the attrition of pregnancies, may lead to selection bias.[9,21,31,32]

Furthermore, in case-control studies with prevalent cases nested within a birth cohort, the selection of controls is dependent on the exposure (i.e., maternal smoking) when the exposure is associated with prematurity).[66-69] It is explained because the probability of selecting non-preterm controls is higher than the probability of selection preterm controls since preterm deliveries are less frequent than term deliveries. In addition, the duration of the pregnancy is considered the period at risk for developing preeclampsia. Thus, women who deliver at term have a higher probability of being selected as controls (over-represented because of longer duration of the gestation).[66-69] Hence, as our findings suggest, the exposure distribution among controls is not an estimate of the exposure distribution in the person-time

that gave rise to the cases over-representing non-preterm infants from both smokers and non-smokers women. On the other hand, when using prevalent cases the start of the follow-up and the end of the study coincides, time is not accounted correctly and may lead to an over-representation of exposed cases of long duration (i.e., smokers with preeclampsia delivering at term).[31-33,69] Thus, the sampling will be biased because the probability of selection at a given point in time depends on the time spent at risk.[66,69]

Given the observational nature of epidemiological research, paradoxical associations may arise when complex relations between risk factors and outcomes are modeled without considering the limitations of different study designs and the complex effect of time.[31,32,65-67] Causal inference and mediation analysis help to understand and disentangle epidemiological paradoxes such as the example illustrated in our present study.

Different methodological approaches have been suggested to unveil paradoxical effects in perinatal epidemiology.[8-10,12,70] The use of DAGs helps to clarify the structure of the biases, and distinguishes between biases resulting from (inappropriate) conditioning on common effects (collider bias) and lack of conditioning on common causes of exposure and outcome (confounding).[9,11,21,71] However, conditioning on an intermediate with sensitivity analysis has been described as the approach of greatest interest in perinatal epidemiology.[9] The advantage of this approach is that, after correction of the bias through sensitivity analysis, the effect of the exposure for individuals with the intermediate, correspond to the direct effect of the exposure on the outcome not through the intermediate.[9]

We used both, deterministic and probabilistic sensitivity analyses, to bias adjust the association between maternal smoking and preeclampsia. Deterministic sensitivity analysis provides an external adjustment of the observed measure of association upon the specification of a list of hypothetical values for the bias parameters without accounting for the uncertainty of the bias parameters.[41,73] Therefore, we used a probabilistic sensitivity analysis to overcome this limitation specifying prior probability distributions for the bias parameters that capture our uncertainty about the bias parameters.[41] On the other hand, the probabilistic approach requires estipulate deterministic prior distributions about the bias parameters[24,25,73] and this has been criticized because it reflects judgments of the investigator about sources of systematic error and therefore subjectivity (by varying the input prior distributions for

probabilistic sensitivity analyses).[73] However, to minimize subjectivity we used published evidence to support this uncertainty.[37,38]

Future observational case controls studies may have to consider the use incident cases and matching with controls by time (i.e. gestation weeks at delivery) in a density sampling design. In case this is not feasible, the smoking-preeclampsia paradox will have to be addressed considering the effect of potential unmeasured confounders between intermediates (gestation weeks) and outcome (preeclampsia), through a sensitivity analysis. Alternatively, in the case of longitudinal studies, investigators should consider conditioning on measured potential confounders, establishing the start of follow-up at a given specific gestation week, the closest to the conception of the gestation the better. Furthermore, to deal with the smoking-preeclampsia paradox, researchers will need to account for the exact onset time of preeclampsia in order to estimate precise person time at risk and finally, if available data, assessing smoking characteristics such us onset age, intensity and, duration.

However, using data from vital statistics the assessment of some potential confounders is not possible given that some pregnancies will not have been selected into the population in study because they were left truncated.[74] This situation will require other methodological approaches in the setting of longitudinal data analyses, such as accounting for competing risk for preeclampsia (e.g. miscarriages and very early pregnancy loss) preventing to develop the outcome under study.[75]

In conclusion, using the counterfactual framework, DAGs and a probabilistic sensitivity analysis to evaluate bias due to an uncontrolled source of censoring preventing the outcome to develop we have unveiled the structure of the paradox in the setting of a case-control observational study. In particular, we have shown that the smoking-preeclampsia paradox is likely an example of selection bias most likely caused by studying cases prevalent at birth rather than all incident cases in a conception or pregnancy cohort, omitting important confounders associated with both smoking and preeclampsia (preventing the outcome to develop), and controlling implicitly or explicitly for a collider (gestation weeks at delivery). Future studies will have to account for this bias or, at least, consider it in the interpretation of findings and weigh it against incomplete evidence on potential biologic mechanisms for this association. We hope that this study will guide future efforts in this area and help estimate true effects of cigarette smoking on maternal and perinatal outcomes, including preeclampsia.

Competing Interests

The authors declare that they do not have any conflict of interest associated with this research and the content is solely the responsibility of the authors.

Funding

This research was supported by the Rose Traveling Fellowship from the departments of Epidemiology and Biostatistics of the Harvard T.H. Chan School of Public Health and by the START Reintegration fellowship (#130814-051) from the Icelandic Centre for Research (Rannis).

Authors' contributions

MALF, UV and HZ are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. All authors designed the protocol. HZ provided the data. MALF wrote the manuscript. MALF developed and completed the statistical analysis. MW, UV, and HZ reviewed, edited and accepted the last version of the manuscript.

ANNEX (Stata syntax)

1. Simulation of preeclampsia cumulative incidence distribution by gestation weeks

```
set seed 12345
gen pe=24*(1-rbeta(2,5))+15
hist pe
```

2. Classic logistic regression, stratification and mediation analysis

```
*CLASSICAL APPROACH: collider bias

logistic PE SMOKING
logistic PE SMOKING BMI AGE
logistic PE SMOKING PARITY BMI AGE

*STRATIFICATION: collider stratification bias

logistic PE SMOKING PARITY BMI AGE if PTD==1
logistic PE SMOKING PARITY BMI AGE if PTD==0

*MEDIATION: collider bias

paramed PE, avar(SMOKING) mvar(PTD) cvars(AGE BMI PARITY) a0(0) a1(1) m(1) yreg(logistic) mreg(logistic) boot seed(1234)

paramed PE, avar(SMOKING) mvar(PTD) cvars(AGE BMI PARITY) a0(0) a1(1) m(1) yreg(logistic) mreg(logistic) nointer boot seed(1234)
```

3. Sensitivity analysis

```
 cc PE SMOKING

episensi 57 319 186 747, st(cc) reps(100)  nodots dpexp(uni(.01  .08)) dpunexp(uni(0.01 0.30)) drrcd(log-n(ln(20)+ln(2.2)/(2)  ln(20)-ln(2.2)/2*1.96)) sed(123)

episensi 57 319 186 747, st(cc) reps(1000)  nodots dpexp(uni(.01  .08)) dpunexp(uni(0.01 0.30)) drrcd(log-n(ln(20)+ln(2.2)/(2)  ln(20)-ln(2.2)/2*1.96)) sed(123)

episensi 57 319 186 747, st(cc) reps(10000)  nodots dpexp(uni(.01  .08)) dpunexp(uni(0.01 0.30)) drrcd(log-n(ln(20)+ln(2.2)/(2)  ln(20)-ln(2.2)/2*1.96)) sed(123)
```

FIGURES LEGENDS

Figure 1. Flowchart of the sampling selection process, n= 1,309

Figure 2. Simulation of preeclampsia incidence estimates by gestational age (weeks) at delivery and smoking status.

Footnote Figure 2:

The source of preeclampsia estimates by gestation weeks: Harmon QE, Huang L, Umbach DM, et al. Risk of fetal death with preeclampsia. Obstet Gynecol. 2015;125(3):628-635, In-house.

Figure 3. DAGs: collider bias (unadjusted and adjusted preeclampsia odds ratios by maternal smoking status).

Figure 4. DAGs: collider stratification scenario (preeclampsia adjusted odds ratios by maternal smoking status and stratified by gestation weeks at delivery).

Figure 5. DAGs: mediation analysis scenario (natural direct and indirect mediated effects).

TABLES

Table 1. Maternal BMI, age, parity and smoking status by preeclampsia status (cases), n= 1,309 (29% preeclampsia, n=376 women).

Variables	N (%)	Preeclampsia, n (%)	Odds Ratio (95%CI)	p-value*
Maternal age (years)				0.418
<35	1087 (83)	308 (28)	Ref.	
≥35	219 (17)	68 (31)	1.13 (0.81, 1.57)	
Parity				<0.001
Nulliparous	602 (46)	228 (38)	2.29 (1.77, 2.94)	
Multiparous	704 (54)	148 (21)	Ref.	
Early pregnancy BMI in kg/m^2				<0.001
Normal weight (<18-24.9)	763 (60)	174 (23)	Ref.	
Overweight (25-29.9)	323 (25)	106 (33)	1.65 (1.24, 2.20)	
Obese (≥30)	182 (15)	84 (46)	2.90 (2.07, 4.06)	
Pregnancy Smoking status				0.044
Smoker	933 (71)	57 (15)	0.72 (0.52, 0.99)	
Non-smoker	376 (29)	186 (20)	Ref.	

*Pearson Chi-square

Table 2. Smoking status by maternal BMI, age and parity, n= 1,309 (19% smokers, n= 243).

Variables	N (%)	Smokes, n (%)	Odds Ratio (95%CI)	p-value*
Maternal age (years)				0.399
<35	1087 (83)	197 (18)	Ref.	
≥35	219 (17)	45 (21)	1.13 (0.84, 1.53)	
Parity				0.726
Nulliparous	602 (46)	128 (18)	1.03 (0.88, 1.20)	
Multiparous	704 (54)	114 (19)	Ref.	
Early pregnancy BMI in kg/m^2				0.632
Normal weight (<18-24.9)	763 (60)	137 (18)	Ref.	
Overweight (25-29.9)	323 (25)	66 (20)	1.17 (0.84, 1.62)	
Obese (≥30)	182 (15)	34 (19)	1.05 (0.69, 1.60)	

*Pearson Chi-square

Table 3. Preeclampsia simulated bias adjusted odds ratios and 95%CI for maternal smoking.

π_{1m}	π_{0m}	γ(Effect of U on Y)	Bias (B)	Bias Adjusted OR	LCI (2.5%)	UCI (97.5%)
0.02	0.15	3.00	0.80	0.90	0.58	1.71
0.02	0.15	3.50	0.76	0.94	0.55	1.80
0.02	0.15	4.00	0.73	0.98	0.53	1.88
0.02	0.15	4.50	0.70	1.03	0.51	1.95
0.02	0.15	5.00	0.68	1.07	0.49	2.03
0.02	0.15	5.50	0.65	1.11	0.47	2.11
0.04	0.20	3.00	0.77	0.93	0.56	1.78
0.04	0.20	3.50	0.73	0.98	0.53	1.87
0.04	0.20	4.00	0.70	1.03	0.51	1.96
0.04	0.20	4.50	0.67	1.07	0.48	2.04
0.04	0.20	5.00	0.64	1.12	0.47	2.13
0.04	0.20	5.50	0.62	1.16	0.45	2.21
0.08	0.30	3.00	0.73	0.99	0.52	1.89
0.08	0.30	3.50	0.69	1.05	0.50	2.00
0.08	0.30	4.00	0.65	1.10	0.47	2.10
0.08	0.30	4.50	0.62	1.15	0.45	2.20
0.08	0.30	5.00	0.60	1.20	0.43	2.28
0.08	0.30	5.50	0.58	1.24	0.42	2.37

π_{1m}: Unobserved confounders for smoking mothers; π_{0m}: Unobserved confounders for non-smoking mothers;
γ: Effect of U on preeclampsia; OR: Odds Ratio; LCI: Lower confidence interval; UCI: Upper confidence interval

REFERENCES

1. Savitz DA, Danilack VA, Engel SM, Elston B, Lipkind HS. Descriptive epidemiology of chronic hypertension, gestational hypertension, and preeclampsia in New York State, 1995-2004. *Matern Child Health J.* 2014;18(4):829-838.
2. Jeyabalan A. Epidemiology of preeclampsia: impact of obesity. *Nutr Rev.* 2013;71 Suppl 1:S18-25.
3. Cnattingius S, Mills JL, Yuen J, Eriksson O, Salonen H. The paradoxical effect of smoking in preeclamptic pregnancies: smoking reduces the incidence but increases the rates of perinatal mortality, abruptio placentae, and intrauterine growth restriction. *Am J Obstet Gynecol.* 1997;177(1):156-161.
4. Marcoux S, Brisson J, Fabia J. The effect of cigarette smoking on the risk of preeclampsia and gestational hypertension. *Am J Epidemiol.* 1989;130(5):950-957.
5. Perni UC, Wikstrom AK, Cnattingius S, Villamor E. Interpregnancy change in smoking habits and risk of preeclampsia: a population-based study. *Am J Hypertens.* 2012;25(3):372-378.
6. Wikstrom AK, Stephansson O, Cnattingius S. Tobacco use during pregnancy and preeclampsia risk: effects of cigarette smoking and snuff. *Hypertension.* 2010;55(5):1254-1259.
7. England L, Zhang J. Smoking and risk of preeclampsia: a systematic review. *Front Biosci.* 2007;12:2471-2483.
8. Wilcox AJ, Weinberg CR, Basso O. On the pitfalls of adjusting for gestational age at birth. *Am J Epidemiol.* 2011;174(9):1062-1068.
9. VanderWeele TJ, Mumford SL, Schisterman EF. Conditioning on intermediates in perinatal epidemiology. *Epidemiology.* 2012;23(1):1-9.
10. Hernandez-Diaz S, Schisterman EF, Hernan MA. The birth weight "paradox" uncovered? *Am J Epidemiol.* 2006;164(11):1115-1120.
11. Cole SR, Platt RW, Schisterman EF, et al. Illustrating bias due to conditioning on a collider. *Int J Epidemiol.* 2010;39(2):417-420.
12. Hill G, Connelly J, Hebert R, Lindsay J, Millar W. Neyman's bias re-visited. J Clin Epidemiol. 2003;56(4):293-296.

13. Delgado-Rodriguez M, Llorca J. Bias. Journal of epidemiology and community health. 2004 Aug;58(8):635-41

14. Rothman KJ, Greenland S, Lash TL. *Modern epidemiology.* 3rd ed. Philadelphia ; London: Lippincott Williams & Wilkins; 2008.

15. Pearce N, Richiardi L. Commentary: Three worlds collide: Berkson's bias, selection bias and collider bias. *Int J Epidemiol.* 2014;43(2):521-524.

16. Martin CL, Hall MH, Campbell DM. The effect of smoking on pre-eclampsia in twin pregnancy. *BJOG.* 2000;107(6):745-749.

17. Misra DP, Kiely JL. The effect of smoking on the risk of gestational hypertension. *Early Hum Dev.* 1995;40(2):95-107.

18. Savitz DA, Zhang J. Pregnancy-induced hypertension in North Carolina, 1988 and 1989. *Am J Public Health.* 1992;82(5):675-679.

19. Engel SM, Janevic TM, Stein CR, Savitz DA. Maternal smoking, preeclampsia, and infant health outcomes in New York City, 1995-2003. *Am J Epidemiol.* 2009;169(1):33-40.

20. Engel SM, Scher E, Wallenstein S, et al. Maternal active and passive smoking and hypertensive disorders of pregnancy: risk with trimester-specific exposures. *Epidemiology.* 2013;24(3):379-386.

21. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology.* 2003;14(3):300-306.

22. T.A. Gudnadotir, Brian T. Bateman, S. Hernandez-Diaz, M.A. Luque-Fernandez, D.P. Geirs, U.Valdimarsdottir, H.Zoega. Body mass index, smoking and hypertensive disorders during pregnancy: a population based case-control study. Submitted for publication.

23. World Health Organization. International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) Version for 2010.; 2010.

24. Greenland S. Useful methods for sensitivity analysis of observational studies. *Biometrics.* 1999;55(3):990-991.

25. Greenland S. Basic methods for sensitivity analysis of biases. *International journal of epidemiology.* 1996;25(6):1107-1116.

26. VanderWeele TJ. Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology.* 2010;21(4):540-551.

27. Vanderweele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology.* 2011;22(1):42-52.

28. VanderWeele TJ, Hernan MA. From counterfactuals to sufficient component causes and vice versa. *Eur J Epidemiol.* 2006;21(12):855-858.

29. Emsley R, Liu H. PARAMED: Stata module to perform causal mediation analysis using parametric regression models. 2013.

30. Valeri L, Vanderweele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychological methods.* 2013;18(2):137-150.

31. Robins JM, Gail MH, Lubin JH. More on "Biased selection of controls for case-control analyses of cohort studies". *Biometrics.* Jun 1986;42(2):293-299.

32. Lubin JH, Gail MH. Biased selection of controls for case-control analyses of cohort studies. *Biometrics.* Mar 1984;40(1):63-75.

33. Neyman J. Statistics; servant of all sciences. *Science.* 1955;122(3166):401-406.

34. Harmon QE, Huang L, Umbach DM, et al. Risk of fetal death with preeclampsia. *Obstet Gynecol.* 2015;125(3):628-635.

35. Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. *Int J Epidemiol.* 2013;42(5):1511-1519.

36. VanderWeele TJ, Chiba Y. Sensitivity analysis for direct and indirect effects in the presence of exposure-induced mediator-outcome confounders. *Epidemiol Biostat Public Health.* 2014;11(2).

37. Wisborg K, Kesmodel U, Henriksen TB, Olsen SF, Secher NJ. Exposure to tobacco smoke in utero and the risk of stillbirth and death in the first year of life. *Am J Epidemiol.* 2001;154(4):322-327.

38. Kesmodel U, Wisborg K, Olsen SF, Henriksen TB, Secher NJ. Moderate alcohol intake during pregnancy and the risk of stillbirth and death in the first year of life. *Am J Epidemiol.* 2002;155(4):305-312.

39. Hyland A, Piazza KM, Hovey KM, et al. Associations of lifetime active and passive smoking with spontaneous abortion, stillbirth and tubal ectopic pregnancy: a cross-sectional analysis of historical data from the Women's Health Initiative. *Tob Control.* 2014.

40. Cnattingius S, Haglund B, Meirik O. Cigarette smoking as risk factor for late fetal and early neonatal death. *BMJ.* 1988;297(6643):258-261.

41. Orsini N, Bellocco R, Bottai M, Wolk A, Greenland S. A tool for deterministic and probabilistic sensitivity analysis of epidemiologic studies. *Stata Journal.* 2008;8(1):29-48.

42. Palei AC, Spradley FT, Warrington JP, George EM, Granger JP. Pathophysiology of hypertension in pre-eclampsia: a lesson in integrative physiology. *Acta Physiol (Oxf).* 2013;208(3):224-233.

43. Nejatizadeh A, Stobdan T, Malhotra N, Pasha MA. The genetic aspects of pre-eclampsia: achievements and limitations. *Biochem Genet.* 2008;46(7-8):451-479.

44. Ananth CV. Ischemic placental disease: a unifying concept for preeclampsia, intrauterine growth restriction, and placental abruption. *Semin Perinatol.* 2014;38(3):131-132.

45. Castles A, Adams EK, Melvin CL, Kelsch C, Boulton ML. Effects of smoking during pregnancy. Five meta-analyses. *American journal of preventive medicine.* 1999;16(3):208-215.

46. Cnattingius S. The epidemiology of smoking during pregnancy: smoking prevalence, maternal characteristics, and pregnancy outcomes. *Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco.* 2004;6 Suppl 2:S125-140.

47. Pueyo V, Guerri N, Oros D, et al. Effects of smoking during pregnancy on the optic nerve neurodevelopment. *Early human development.* 2011;87(5):331-334.

48. Ram FS, McDonald EM. Response to 'Inhibitory effects of maternal smoking on the development of severe retinopathy of prematurity'. *Eye (Lond).* 2011;25(1):123-124; author reply 124.

49. Hogberg L, Cnattingius S, Lundholm C, D'Onofrio BM, Langstrom N, Iliadou AN. Effects of maternal smoking during pregnancy on offspring blood pressure in late adolescence. *Journal of hypertension.* 2012;30(4):693-699.

50. Conde-Agudelo A, Althabe F, Belizan JM, Kafury-Goeta AC. Cigarette smoking during pregnancy and risk of preeclampsia: a systematic review. *Am J Obstet Gynecol.* 1999;181(4):1026-1035.

51. Feng D, Liu T, Su DF, et al. The association between smoking quantity and hypertension mediated by inflammation in Chinese current smokers. *J Hypertens.* 2013;31(9):1798-1805.

52. D'Elia L, De Palma D, Rossi G, et al. Not smoking is associated with lower risk of hypertension: results of the Olivetti Heart Study. *Eur J Public Health.* 2014;24(2):226-230.

53. Mons U, Muezzinler A, Gellert C, et al. Impact of smoking and smoking cessation on cardiovascular events and mortality among older adults: meta-analysis of individual participant data from prospective cohort studies of the CHANCES consortium. *BMJ.* 2015;350:h1551.

54. Karumanchi SA, Levine RJ. How does smoking reduce the risk of preeclampsia? *Hypertension.* 2010;55(5):1100-1101.

55. Venditti CC, Casselman R, Young I, Karumanchi SA, Smith GN. Carbon monoxide prevents hypertension and proteinuria in an adenovirus sFlt-1 preeclampsia-like mouse model. *PLoS One.* 2014;9(9):e106502.

56. Zhang F, Kaide JI, Rodriguez-Mulero F, Abraham NG, Nasjletti A. Vasoregulatory function of the heme-heme oxygenase-carbon monoxide system. *Am J Hypertens.* 2001;14(6 Pt 2):62S-67S.

57. Fujita T, Toda K, Karimova A, et al. Paradoxical rescue from ischemic lung injury by inhaled carbon monoxide driven by derepression of fibrinolysis. *Nat Med.* 2001;7(5):598-604.

58. Brouard S, Otterbein LE, Anrather J, et al. Carbon monoxide generated by heme oxygenase 1 suppresses endothelial cell apoptosis. *J Exp Med.* 2000;192(7):1015-1026.

59. Liu XM, Chapman GB, Peyton KJ, Schafer AI, Durante W. Carbon monoxide inhibits apoptosis in vascular smooth muscle cells. *Cardiovasc Res.* 2002;55(2):396-405.

60. Llurba E, Sanchez O, Dominguez C, et al. Smoking during pregnancy: changes in mid-gestation angiogenic factors in women at risk of developing preeclampsia according to uterine artery Doppler findings. *Hypertens Pregnancy.* 2013;32(1):50-59.

61. Cudmore M, Ahmad S, Al-Ani B, et al. Negative regulation of soluble Flt-1 and soluble endoglin release by heme oxygenase-1. *Circulation.* 2007;115(13):1789-1797.

62. Hogberg L, Cnattingius S. The influence of maternal smoking habits on the risk of subsequent stillbirth: is there a causal relation? *BJOG.* 2007;114(6):699-704.

63. Raymond EG, Cnattingius S, Kiely JL. Effects of maternal age, parity, and smoking on the risk of stillbirth. *Br J Obstet Gynaecol.* 1994;101(4):301-306.

64. Gray R, Bonellie SR, Chalmers J, et al. Contribution of smoking during pregnancy to inequalities in stillbirth and infant death in Scotland 1994-2003: retrospective population based study using hospital maternity records. BMJ. 2009;339:b3754.

65. Geneletti S, Richardson S, Best N. Adjusting for selection bias in retrospective, case-control studies. Biostatistics. 2009 Jan;10(1):17-31

66. Richardson DB. An incidence density sampling program for nested case-control analyses. Occupational and environmental medicine. 2004 Dec;61(12):e59

67. Azzato EM, Greenberg D, Shah M, et al. Prevalent cases in observational studies of cancer survival: do they bias hazard ratio estimates? *Br J Cancer.* 2009;100(11):1806-1811.

68. Kyrklund-Blomberg NB, Cnattingius S. Preterm birth and maternal smoking: risks related to gestational age and onset of delivery. *Am J Obstet Gynecol.* 1998;179(4):1051-1055.

69. Breslow NE, Day NE. Statistical methods in cancer research. Volume I - The analysis of case-control studies. *IARC Sci Publ.* 1980(32):5-338.

70. Hernan MA, Schisterman EF, Hernandez-Diaz S. Invited commentary: composite outcomes as an attempt to escape from selection bias and related paradoxes. *Am J Epidemiol.* 2014;179(3):368-370.

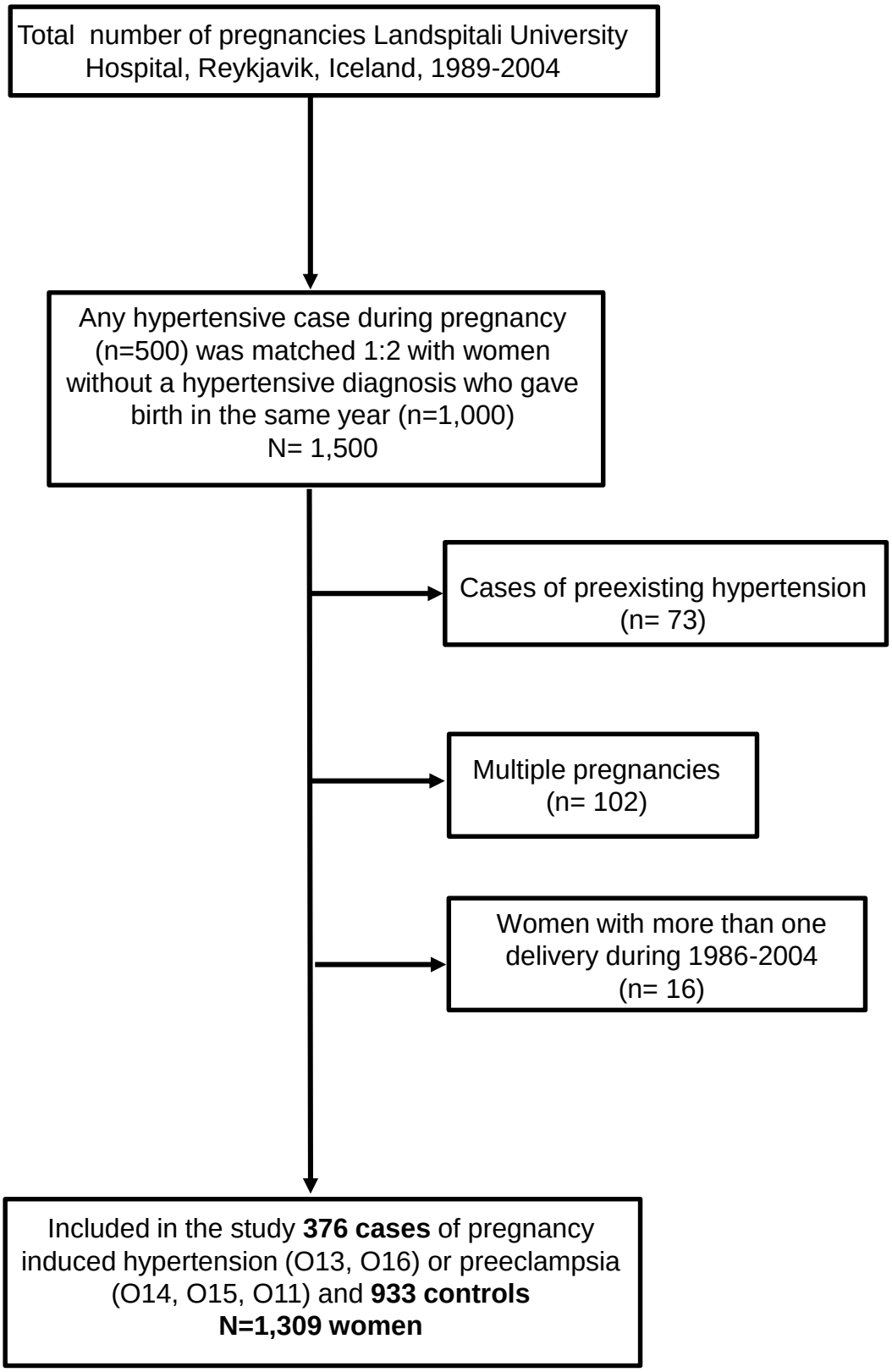
71. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology.* 2004;15(5):615-625.

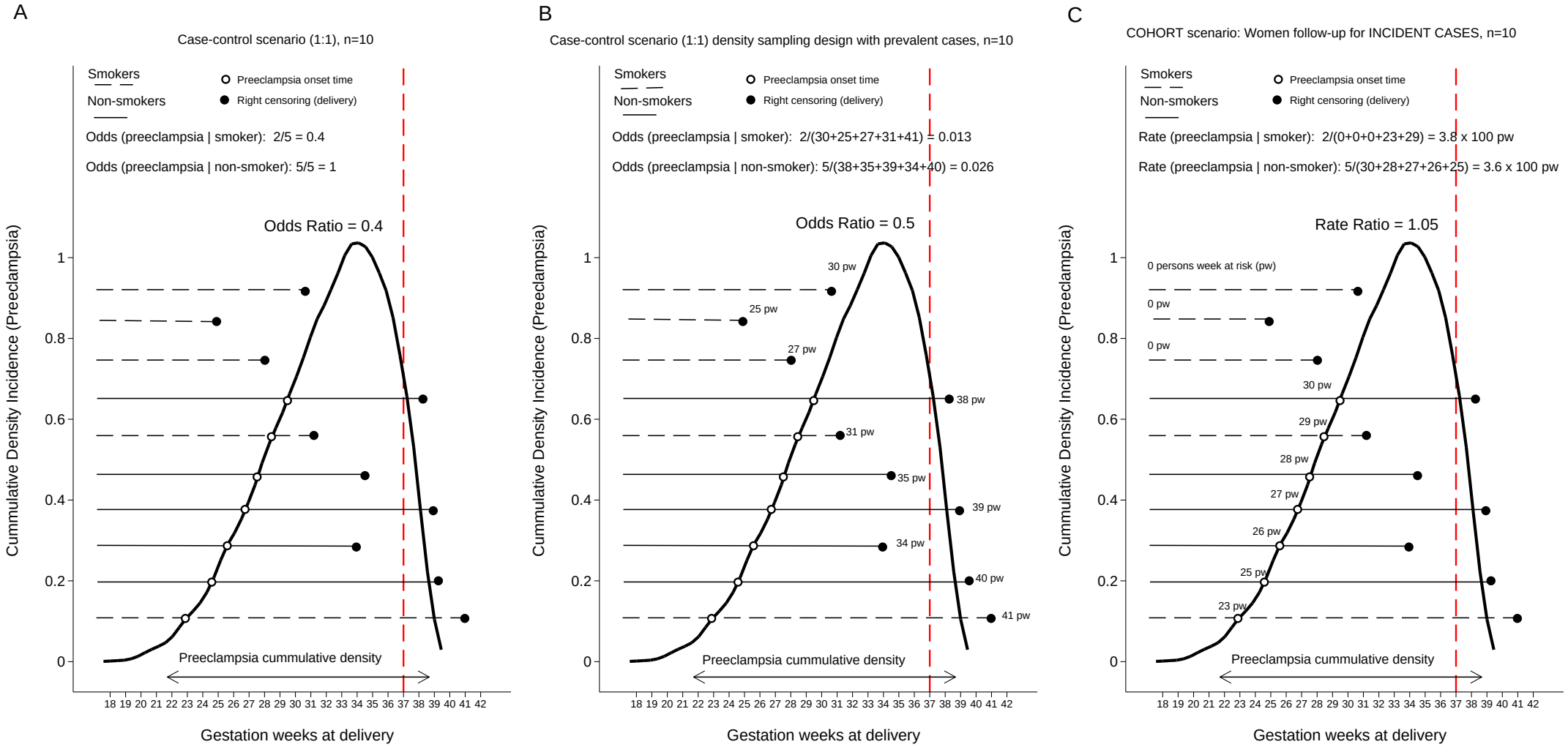
72. Greenland S. Sensitivity analysis, Monte Carlo risk analysis, and Bayesian uncertainty assessment. *Risk Anal.* 2001;21(4):579-583.

73. Arah OA, Chiba Y, Greenland S. Bias formulas for external adjustment and sensitivity analysis of unmeasured confounders. *Annals of epidemiology.* 2008;18(8):637-646.

74. Lisonkova S, Joseph KS. Left Truncation Bias as a Potential Explanation for the Protective Effect of Smoking on Preeclampsia. Epidemiology. 2015 Feb 18

75. Hernan MA, Schisterman EF, Hernandez-Diaz S. Invited commentary: composite outcomes as an attempt to escape from selection bias and related paradoxes. American journal of epidemiology. 2014 Feb 1;179(3):368-70

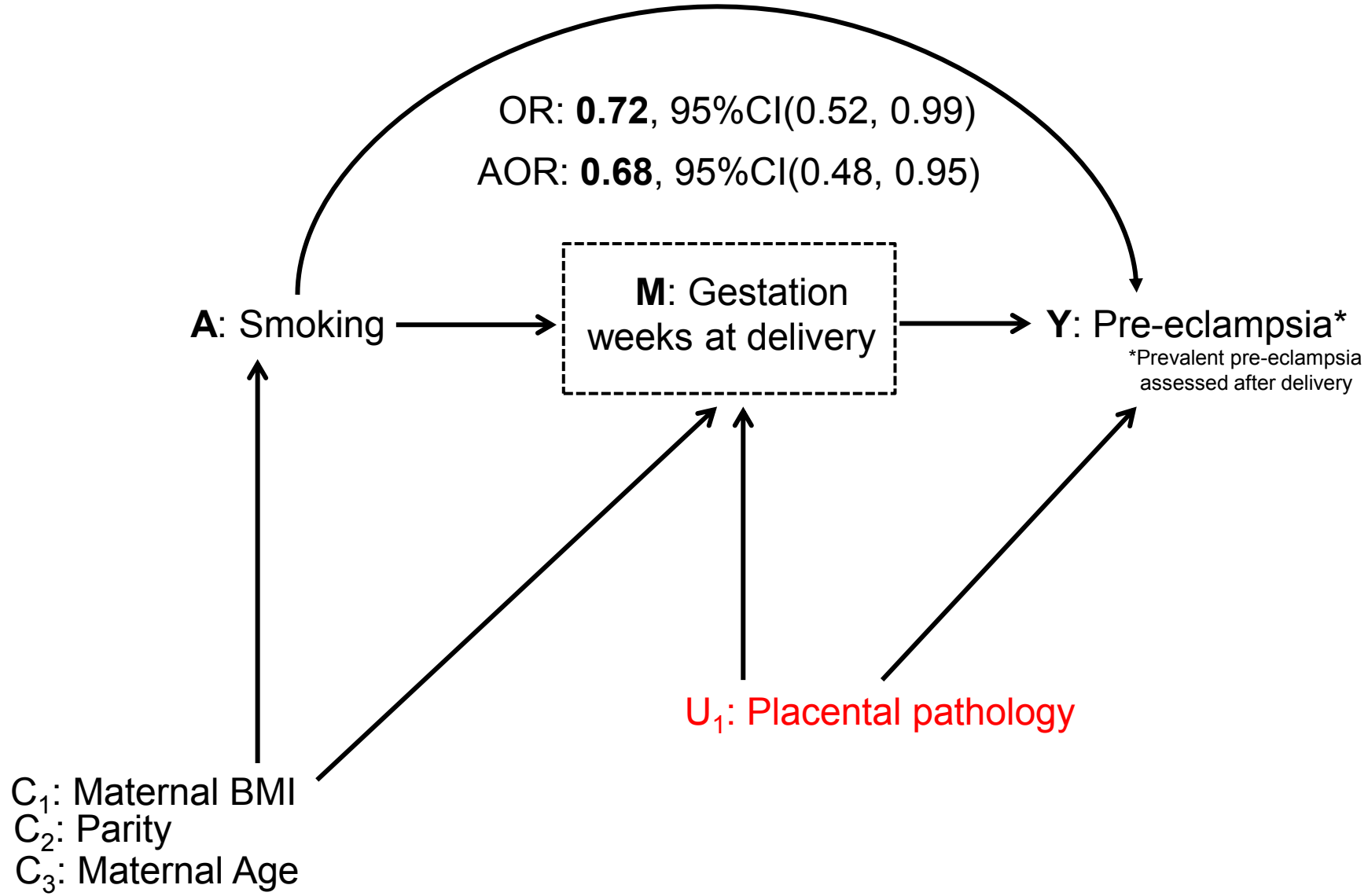

Total number of pregnancies Landspitali University Hospital, Reykjavik, Iceland, 1989-2004
Any hypertensive case during pregnancy (n=500) was matched 1:2 with women without a hypertensive diagnosis who gave birth in the same year (n=1,000)
N= 1,500
Cases of preexisting hypertension (n= 73)
Multiple pregnancies (n= 102)
Women with more than one delivery during 1986-2004 (n= 16)
Included in the study **376 cases** of pregnancy induced hypertension (O13, O16) or preeclampsia (O14, O15, O11) and **933 controls**
N=1,309 women


A
Case-control scenario (1:1), n=10
Smokers
Non-smokers
Preeclampsia onset time
Right censoring (delivery)
Odds (preeclampsia | smoker): 2/5 = 0.4
Odds (preeclampsia | non-smoker): 5/5 = 1
Odds Ratio = 0.4
Cummulative Density Incidence (Preeclampsia)
Preeclampsia cummulative density
Gestation weeks at delivery
B
Case-control scenario (1:1) density sampling design with prevalent cases, n=10
Odds (preeclampsia | smoker): 2/(30+25+27+31+41) = 0.013
Odds (preeclampsia | non-smoker): 5/(38+35+39+34+40) = 0.026
Odds Ratio = 0.5
30 pw
25 pw
27 pw
38 pw
31 pw
35 pw
39 pw
34 pw
40 pw
41 pw
C
COHORT scenario: Women follow-up for INCIDENT CASES, n=10
Rate (preeclampsia | smoker): 2/(0+0+0+23+29) = 3.8 x 100 pw
Rate (preeclampsia | non-smoker): 5/(30+28+27+26+25) = 3.6 x 100 pw
Rate Ratio = 1.05
0 persons week at risk (pw)
0 pw
0 pw
30 pw
29 pw
28 pw
27 pw
26 pw
25 pw
23 pw

Classical approach (case control study with prevalent cases):

$E(Y|A) = \log(OR) = \beta_0 + \beta_1 Smoking$

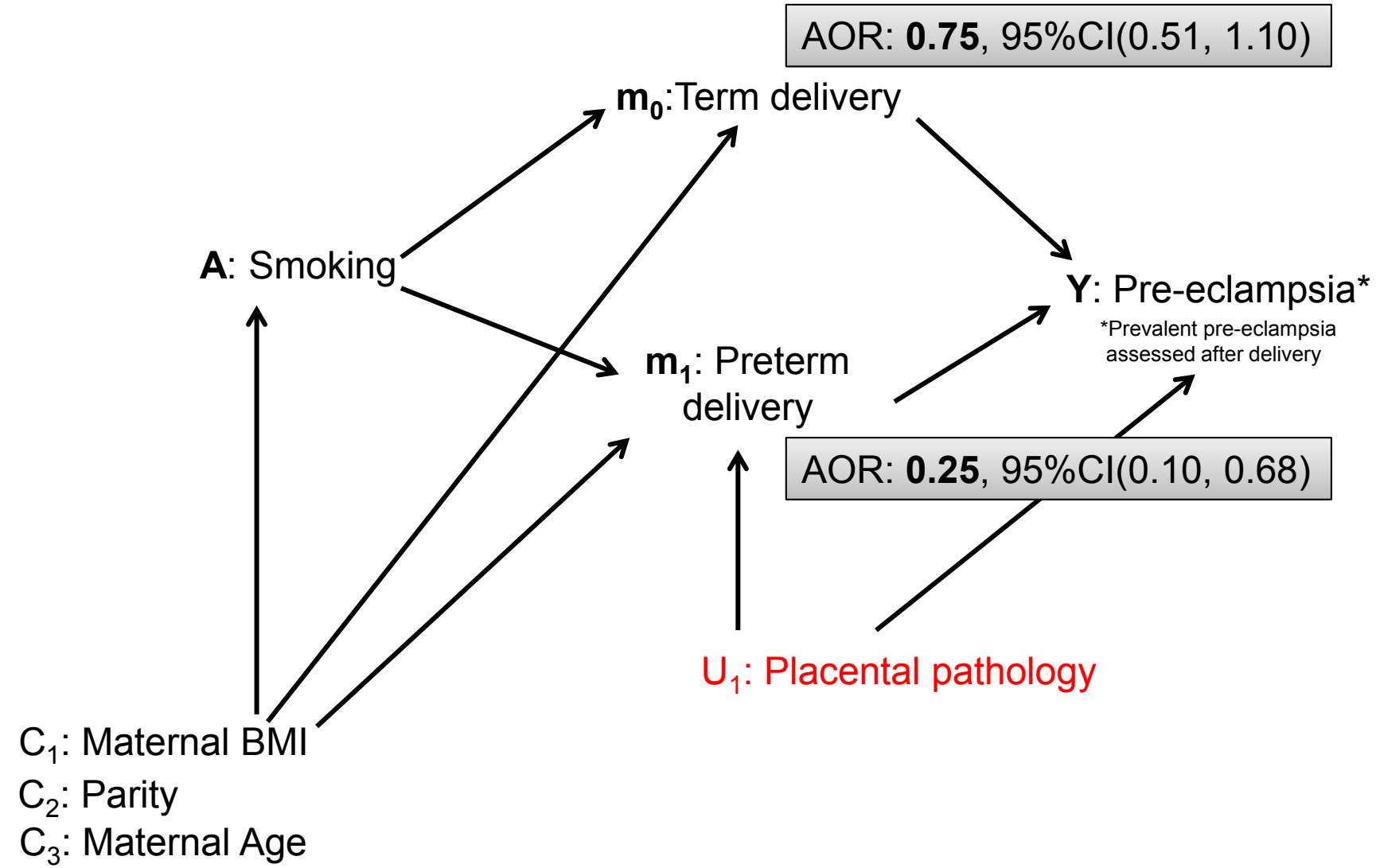
$E(Y|A,C) = \log(OR) = \beta_0 + \beta_1 Smoking + \beta_3 BMI + \beta_3 AGE + \beta_4 Parity$



Classical approach (case control study): Stratification and collider effect

$E(Y|A,C\ m_0) = \log(OR) = \beta_0 + \beta_1 Smoking + \beta_3 BMI + \beta_3 AGE + \beta_4 Parity$

$E(Y|A,C\ m_1) = \log(OR) = \beta_0 + \beta_1 Smoking + \beta_3 BMI + \beta_3 AGE + \beta_4 Parity$



Mediation analysis: Marginal Total Effect: **0.68**, 95%CI(0.45, 0.97)

Assumptions:

(1) is $Y_{am} \coprod A|C$

(2) is $Y_{am} \coprod M|C,A$

(3) is $M_a \coprod A|C$

(4) is $Y_{am} \coprod M_{a^*}|C$