Comparison of Parametric and Nonparametric Estimators for the Association Between Incident Prepregnancy Obesity and Stillbirth in a Population-Based Cohort Study

Ya-Hui Yu, Lisa M. Bodnar, Maria M. Brooks, Katherine P. Himes, and Ashley I. Naimi

Table of Contents

WEB APPENDIX 1

Process of Data Linkage

We extensively cleaned the data to eliminate duplicate records and to identify and correctly match up twins and higher order gestations from the same pregnancy. When potential duplicate records matched on several mother characteristics (including baby's date of birth, mother's name, mother's date of birth, etc.) but had discrepant baby information (birthweight, sex, etc.), they were recoded as twins. We also created a unique maternal ID to identify repeat pregnancies from the same mother over time using a sequential, deterministic linkage strategy (1,2). The records were separated into batches by delivery year and then further into sub-groups by year of last live birth. These subgroups were then merged with birth records from the year indicated as the year of the last live birth. For instance, the 2011 batch of delivery records with 2010 listed as year of last live birth could potentially link with the 2010 delivery records to identify the sibling. Special consideration was given to the potential for two babies delivered to the same mother in the same calendar year, either due to multiple births, or one sibling delivered early in the calendar year and one at the end. Several attempts at linking with maternal identifiers took place for each batch, and additional maternal variables were used to validate the matches, with discrepancies manually checked and corrected. Any false positive matches identified through discrepant validation variables were recoded as unmatched. At the end of the process, 57% of the birth and fetal death records were linked with at least one sibling. Among those who could not be linked, 58% were the first birth for the woman or were missing information on year of previous last live birth and 38% had a last live birth from prior to 2003, and thus could not be linked.

WEB APPENDIX 2

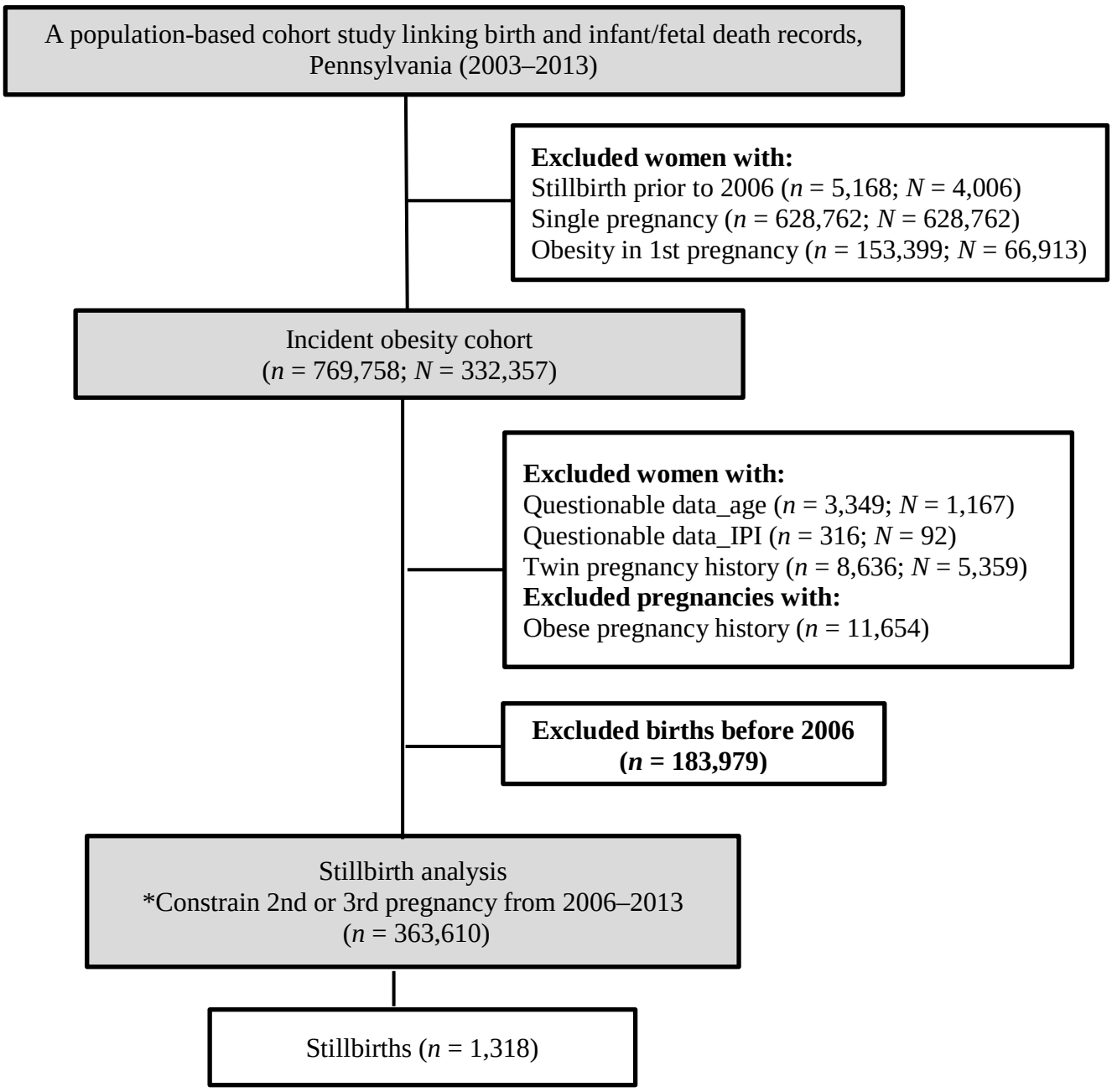
Example of Output From the SuperLearner Algorithm

Initial estimation of Q
Procedure: SuperLearner
Model:
Y ~ SL.mean_All + SL.nnet_All + SL.earth_All + SL.ranger_All + SL.xgboost_All + SL.glmnet_0_All + SL.glmnet_0.25_All + SL.glmnet_0.5_All + SL.glmnet_0.75_All + SL.glmnet_1_All + SL.gam_3_All + SL.gam_4_All + SL.xgb.1_All + SL.xgb.2_All + SL.xgb.3_All + SL.xgb.4_All + SL.xgb.5_All + SL.xgb.6_All + SL.xgb.7_All + SL.xgb.8_All + SL.xgb.9_All + SL.xgb.10_All + SL.xgb.11_All + SL.xgb.12_All + SL.ranger_500_2_All + SL.ranger_1000_2_All + SL.ranger_500_5_All + SL.ranger_1000_5_All + SL.ranger_500_11_All + SL.ranger_1000_11_All + SL.glm.interaction_All

Coefficients:
SL.mean_All 0
SL.nnet_All 0
SL.earth_All 0
SL.ranger_All 0
SL.xgboost_All 0.2937091
SL.glmnet_0_All 0.1466514
SL.glmnet_0.25_All 0
SL.glmnet_0.5_All 0
SL.glmnet_0.75_All 0
SL.glmnet_1_All 0
SL.gam_3_All 0
SL.gam_4_All 0
SL.xgb.1_All 0
SL.xgb.2_All 0
SL.xgb.3_All 0.1847042
SL.xgb.4_All 0.125794
SL.xgb.5_All 0
SL.xgb.6_All 0
SL.xgb.7_All 0
SL.xgb.8_All 0
SL.xgb.9_All 0
SL.xgb.10_All 0
SL.xgb.11_All 0
SL.xgb.12_All 0
SL.ranger_500_2_All 0.1219778
SL.ranger_1000_2_All 0
SL.ranger_500_5_All 0
SL.ranger_1000_5_All 0
SL.ranger_500_11_All 0.1113301
SL.ranger_1000_11_All 0
SL.glm.interaction_All 0.0158334

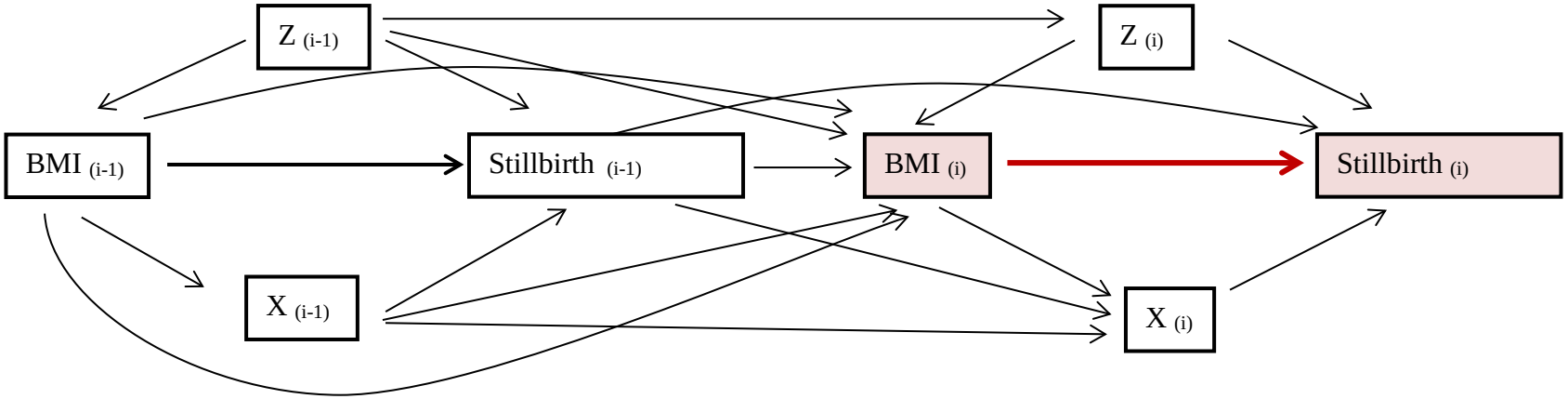
Web Table 1. Characteristics of the Past Pregnancy by Prepregnancy Obesity Status in a Population-Based Cohort Study of Stillbirths and Live Births, Pennsylvania, 2006–2013

	Nonobese No. (%) (*n* = 326,178)		Obese No. (%) (*n* = 37,432)	
Gestational weight gain *z* score (percentile)				
<20th	55,694	(17.1)	4,176	(11.2)
20th–80th	228,027	(69.9)	23,681	(63.3)
>80th	42,457	(13.0)	9,575	(25.6)
Gestational diabetes				
No	318,319	(97.6)	36,007	(96.2)
Yes	7,859	(2.4)	1,425	(3.8)
Gestational hypertension				
No	315,613	(96.8)	35,217	(94.1)
Yes	10,565	(3.2)	2,215	(5.9)
Smoked during pregnancy				
No	247,937	(84.3)	30,370	(81.1)
Yes	51,241	(15.7)	7,062	(18.9)
Birth facility's level of neonatal care				
Level 1	61,729	(18.9)	8,287	(22.1)
Level 2 or 2A	53,379	(16.4)	5,535	(14.8)
Level 3 or 3A/3B/3C	211,070	(64.7)	23,610	(63.1)
Use of Women, Infants, and Children program				
No	211,911	(65.0)	18,575	(49.6)
Yes	114,267	(35.0)	18,857	(50.4)
Death				
Live birth	322,881	(99.0)	36,893	(98.6)
Infant death	2,224	(0.7)	368	(1.0)
Fetal death	1,073	(0.3)	171	(0.5)
Preterm birth				
No	276,990	(84.9)	31,598	(84.4)
Yes	49,188	(15.1)	5,834	(15.6)
Small size for gestational age				
No	297,932	(91.3)	34,256	(91.5)
Yes	28,246	(8.7)	3,176	(8.5)
Large size for gestational age				
No	291,942	(89.5)	32,323	(86.4)
Yes	34,236	(10.5)	5,109	(13.7)
Prepregnancy body mass index				
Underweight	21,942	(6.7)	479	(1.3)
Normal weight	232,006	(71.1)	9,166	(24.5)
Overweight	72,230	(22.1)	27,787	(74.2)
Breastfeeding				
No	103,311	(31.7)	14,747	(39.4)
Yes	222,867	(68.3)	22,685	(60.6)
Delivery route				
Vaginal	256,401	(78.6)	27,087	(72.4)
Planned cesarean	35,609	(10.9)	4,826	(12.9)
Unplanned cesarean	34,168	(10.5)	5,519	(14.7)



*n = no. of pregnancies; N = no. of women.

Web Figure 1. Flow chart showing selection of the analytical sample.

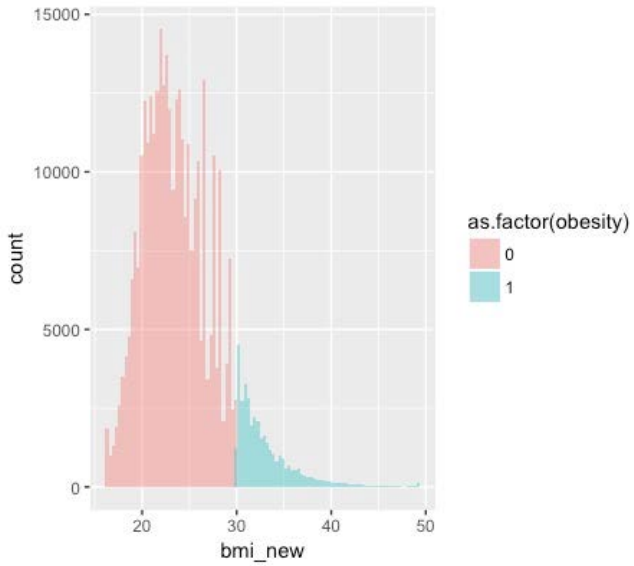


Z (Confounders of prepregnancy BMI and stillbirth):
(i) *Time-fixed variables*: maternal race, maternal height
(ii) *Time-vary variables*: age, maternal education, urban residence, percent Black residents, prepregnancy diabetes, prepregnancy hypertension, smoking status, insurance, inter-pregnancy interval, parity

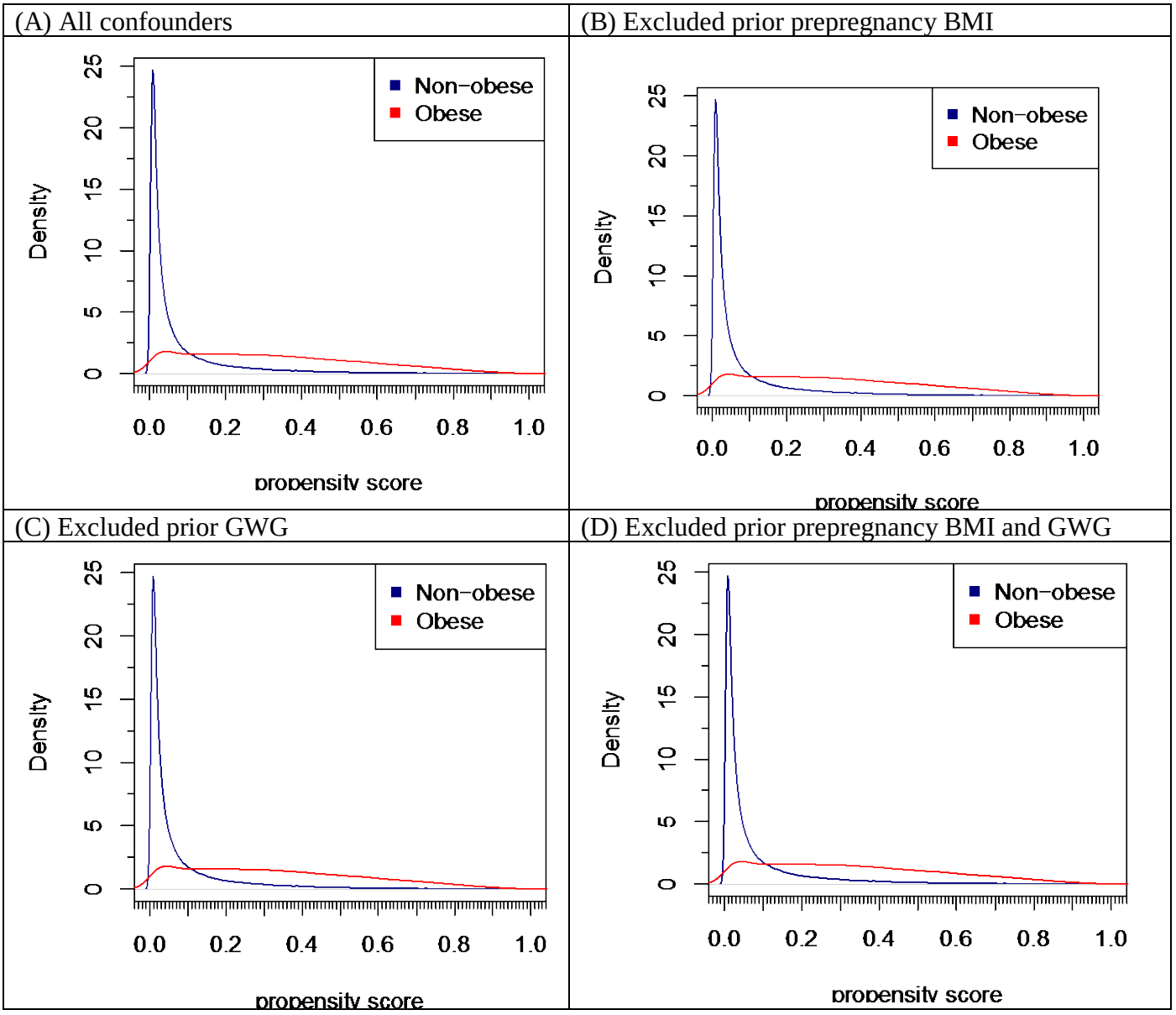
X (Mediators of prepregnancy BMI and stillbirth):
Gestational weight gain, gestational diabetes, gestational hypertension, smoking status during pregnancy, gestational age, apgar score, birth weight, birth facility level of neonatal care, neonatal care admission, breast feeding, delivery route, Women, Infants, and Children program usage,

Unmeasured variables: mental health and stress before pregnancy, diet quality, physical activity, access to government nutrition assistance in between pregnancies, social support, placental function, infection, inflammation.

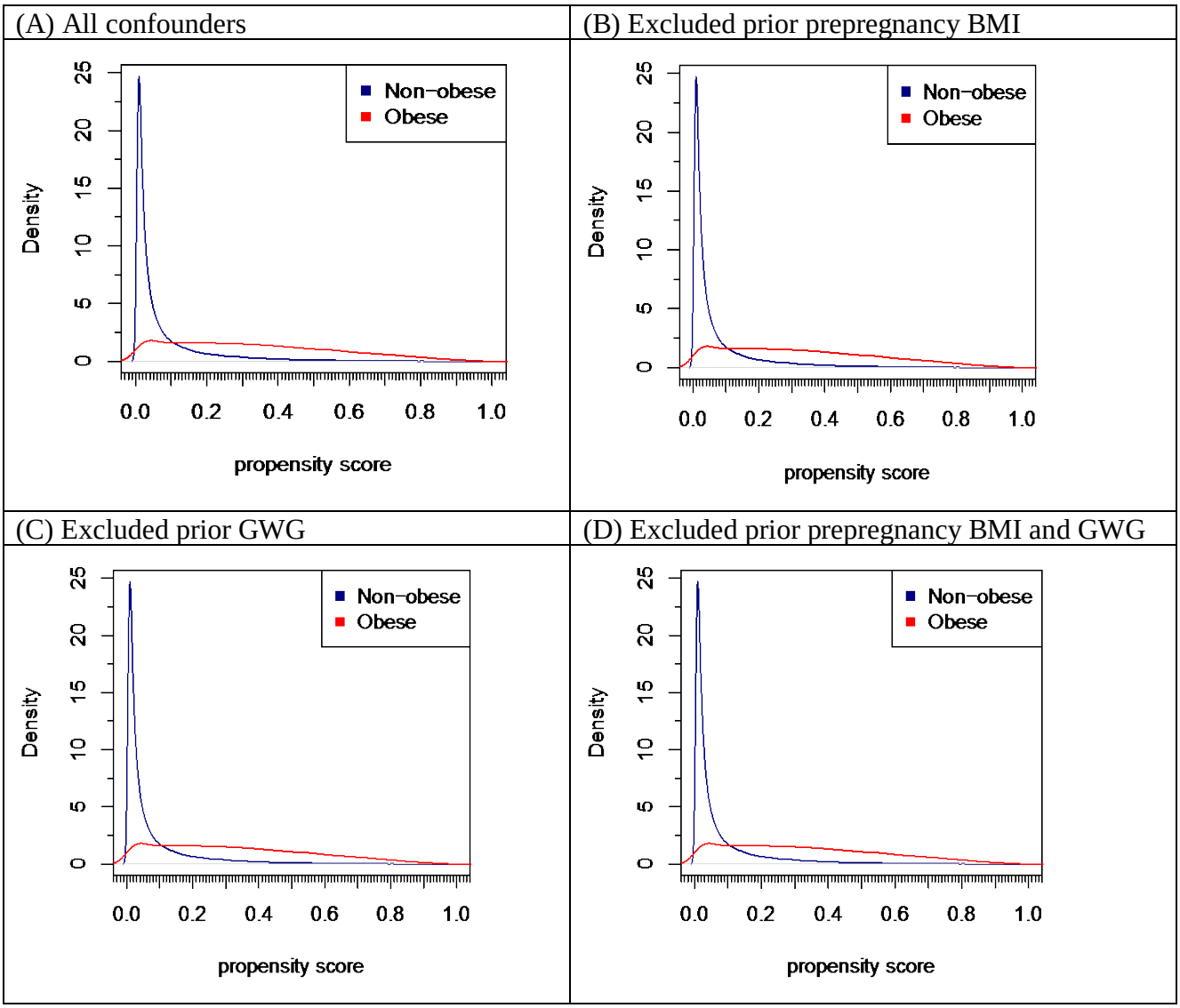
Web Figure 2. Causal diagram representing the relationship of prepregnancy body mass index (BMI) with stillbirth at the *i*th pregnancy.



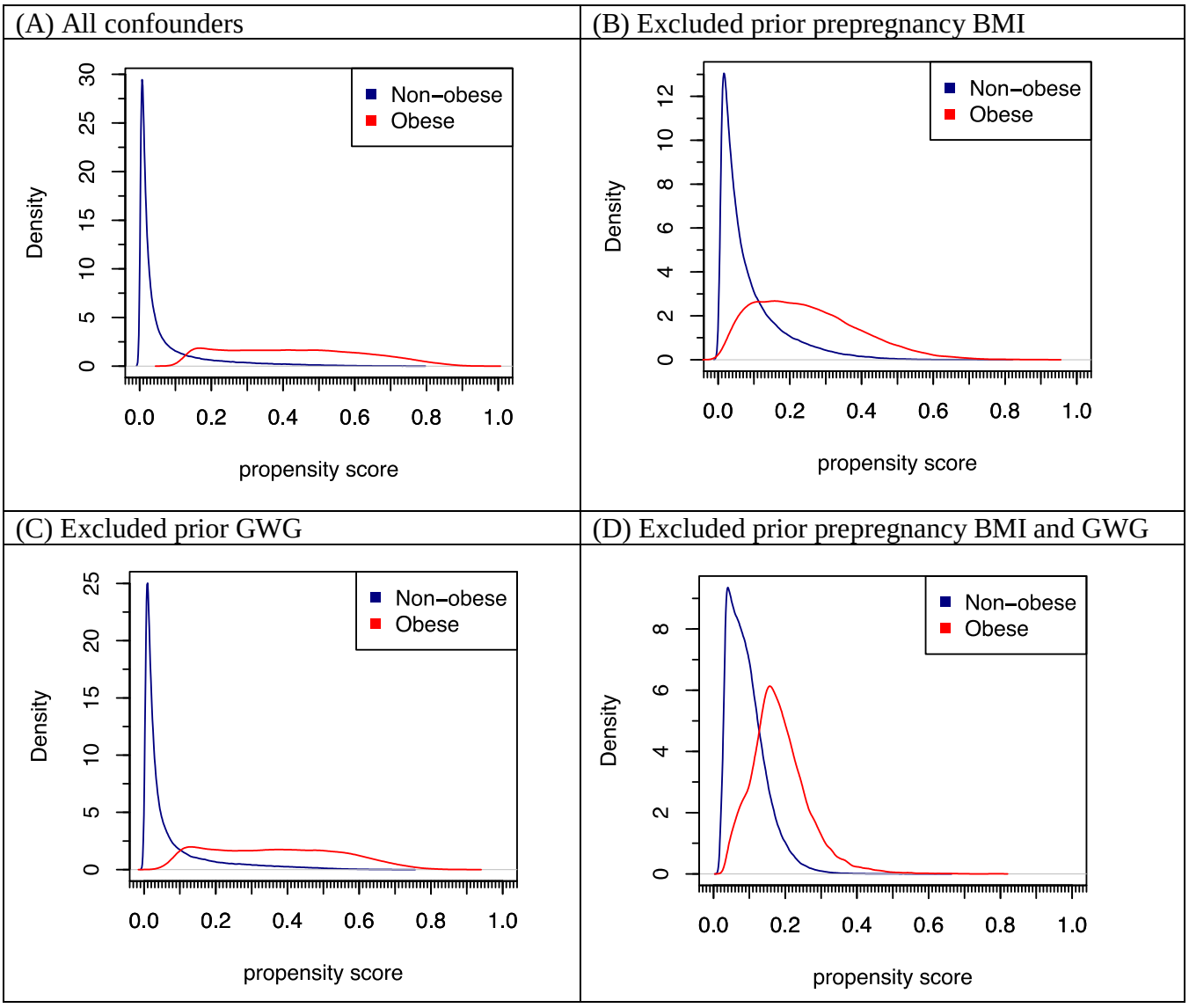
Web Figure 3. Distribution of body mass index (BMI) data by obesity status.



Web Figure 4. Distribution of propensity scores by obesity status, estimated by IPW.



Web Figure 5. Distribution of propensity scores by obesity status, estimated by parametric TMLE.



Web Figure 6. Distribution of propensity scores by obesity status, estimated by nonparametric TMLE.

References

1. Blakely T, Salmond C. Probabilistic record linkage and a method to calculate the positive predictive value. *Int J Epidemiol.* 2002;31(6):1246–1252.

2. Herman AA, McCarthy BJ, Bakewell JM, et al. Data linkage methods used in maternally-linked birth and infant death surveillance data sets from the United States (Georgia, Missouri, Utah and Washington State), Israel, Norway, Scotland and Western Australia. *Paediatr Perinat Epidemiol.* 1997;11(suppl 1):5–22.