on births, including Apgar scores, as well as maternal mortality by hour and day. For twins, we used the Apgar score of the first-born twin. The Sunnybrook Health Sciences Centre research ethics board approved the study.

We used Cox proportional hazard regression to generate hazard ratios (HR) and 95% confidence intervals (CI) for maternal death in relation to the newborn 5-minute Apgar score of 0–6 versus 7–10 (the referent). Maternal death was assessed at (1) 0–24 hours, (2) 0–42 days, (3) 43–365 days, and (4) 366 days to 5 years after the index delivery. HRs were adjusted (aHR) for maternal age at delivery. Model 4 was censored at the end of the study period of March 31, 2016.

Mothers of newborns with an Apgar score of 0–6 tended to smoke more and to have more health problems and obstetric complications than mothers of newborn with an Apgar score of score of 7–10 (eTable 1; http://links.lww.com/EDE/B352). More newborns with an Apgar of 0–6 were male, a twin, born earlier, and required resuscitation, in contrast to those with an Apgar score of 7–10.

The rate of maternal death within 24 hours of delivery was 8.4 versus 0.12 per 10,000 women, comparing those whose 5-minute newborn Apgar score was 0–6 versus 7–10—an aHR of 71 (95% CI = 25, 203; Figure). For maternal death from 0 to 42 days, the aHR was 19 (95% CI = 9.9, 38). For late maternal death from 43 to 365 days (aHR 6.9, 95% CI = 3.5, 14) and death 366 days to 5 years (aHR 1.6, 95% CI = 1.0, 2.7), the risk remained higher in women whose infant had a lower Apgar score, albeit less pronounced (Figure). At both under 24 hours and 0–42 days, no deaths occurred in mothers of infants with an Apgar score of 10.

Upon restricting to 386,313 term births otherwise unaffected by an anomaly or obstetrical complication, the rate of maternal death within 24 hours was 3.0 versus 0.08 per 10,000 women, comparing Apgar scores of 0–6 versus 7–10—an aHR of 42 (4.3, 400; eFigure 1; http://links.lww.com/EDE/B352). The aHRs for the other outcomes did not meaningfully differ from the null (eFigure 1; http://links.lww.com/EDE/B352).

A 5-minute Apgar score of 0–6 is associated with a higher risk of early and late maternal death, especially death within the first 24 hours of delivery. This observation reflects a previous finding of a higher risk of maternal ICU admission if her newborn's Apgar score is low.[3] It is of interest that a newborn Apgar score also reflects a higher risk of late maternal death, at least among more complicated pregnancies. While the causes of early maternal mortality are generally understood,[5] the same is not so for late maternal mortality.[6] The current findings highlight the potential use of the newborn 5-minute Apgar score as a reflection of not only newborn health but also the health of the mother. Among women with a lower newborn Apgar score, it remains to be determined whether increased surveillance is warranted, short term or long term.

ACKNOWLEDGMENTS

J.G.R. was involved with conception of the study, interpretation of the results, writing and editing of the article, as well as approval of the final version. A.L.P. was involved with conception of the study, data analysis, interpretation of results, editing of the article, and approval of final version. A.F.K. was involved in interpretation of results as well as writing and editing of the article.

Alyssa F. Kahane
Undergraduate Medical Education, Faculty of Medicine
University of Ottawa
Ottawa, Ontario

Alison L. Park
Keenan Research Centre, Li Ka Shing Knowledge Institute
St. Michael's Hospital
Toronto, Ontario
Institute for Clinical Evaluative Sciences
Toronto, Ontario

Joel G. Ray
Departments of Medicine and Obstetrics and Gynecology
Keenan Research Centre, Li Ka Shing Knowledge Institute
St. Michael's Hospital
Toronto, Ontario
Institute for Clinical Evaluative Sciences
University of Toronto
Toronto, Ontario
rayj@smh.ca.

REFERENCES

1. Iliodromiti S, Mackay DF, Smith GC, Pell JP, Nelson SM. Apgar score and the risk of cause-specific infant mortality: a population-based cohort study. *Lancet*. 2014;384:1749–1755.
2. Persson M, Johansson S, Cnattingius S. Interpregnancy weight change and risks of severe birth-asphyxia-related outcomes in singleton infants born at term: A Nationwide Swedish Cohort Study. *PLoS Med*. 2016;13:e1002033.
3. Ray JG, Medcalf KE, Park AL. Association of newborn Apgar score with maternal admission to the intensive care unit. *JAMA Pediatr*. 2016;170:88–89.
4. Ray JG, Urquia ML, Berger H, Vermeulen MJ. Maternal and neonatal separation and mortality associated with concurrent admissions to intensive care units. *CMAJ*. 2012;184:E956–E962.
5. Nagaya K, Fetters MD, Ishikawa M, et al. Causes of maternal mortality in Japan. *JAMA*. 2000;283:2661–2667.
6. de Cosio FG, Jiwani SS, Sanhueza A, et al. Late maternal deaths and deaths from sequelae of obstetric causes in the Americas from 1999 to 2013: a trend analysis. *PLoS One* 2016;11:e0160642.

Graphical Diagnosis of Confounding Bias in Instrumental Variable Analysis

To the Editor:

Instrumental variable (IV) methods have become increasingly popular in estimating causal effects from observational health studies.[1–3] A key assumption for valid IV analysis is that the IV is independent of unmeasured confounders conditional on measured con-

The R functions for diagnosis of IV methods (iv.diagnosis and iv.diagnosis.plot) are available in the R package ivmodel available on CRAN. Figures in this article can be replicated from R files submitted with the article.

ISSN: 1044-3983/2018/2904-0e29
DOI: 10.1097/EDE.0000000000000822

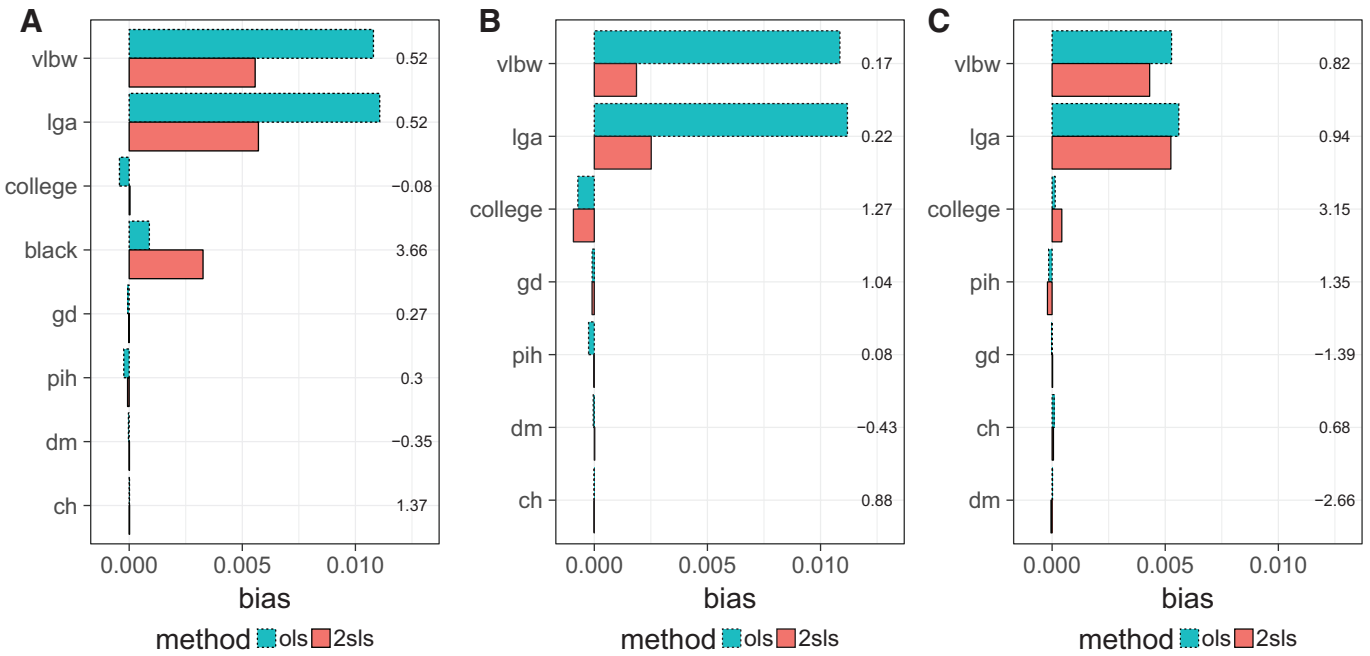

FIGURE. Diagnosis of IV (2SLS) bias compared with regression (OLS) bias from confounding. The bars show the absolute bias induced by a covariate that is computed by (1). The numbers on the right side of the bars are the bias ratios. Several covariates are recorded, including if the baby had vlbw, lga, mother had college degree (college), mother was African-American (black), mother had gd, dm, pih, or ch. A, Entire dataset. B, White mothers only. C, Black mothers only. ch indicates chronic hypertension; dm, diabetes mellitus; gd, gestational diabetes; lga, low gestational age; pih, pregnancy-induced hypertension; vlbw, very low birthweight.

founders. While this assumption cannot be verified empirically, its plausibility can be assessed by checking if the IV is independent of the measured confounders. If the IV is not independent of a measured confounder, it raises concerns when the measured confounder may be a proxy for other unmeasured confounders.[4]

In practice, a common approach is to compute the balance of observed covariates across levels of the IV and compare it with the balance across levels of the treatment.[1] This method is criticized (rightly, in our opinion) by Jackson and Swanson[5] as the bias of an IV estimate can be amplified by how strongly the IV is predictive of the treatment. Let D be the treatment, Y the outcome, Z the binary instrument, X an observed covariate, and Y^d the potential outcome for treatment level d. By assuming the linear structural model $E[Y^d \mid X] = \alpha_0 + \alpha_1 d + \alpha_2 X$, where the only confounder is X, Brookhart and Schneeweiss[4] derived the following bias formulas for two-stage least squares (2SLS) IV and ordinary least squares regression (OLS) estimates that do not control for the measured covariate X:

$$\begin{aligned}&\text{bias}(\text{2SLS estimator})\\&\quad = \alpha_2 \cdot \frac{\text{E}[X \mid Z=1] - \text{E}[X \mid Z=0]}{\text{E}[D \mid Z=1] - \text{E}[D \mid Z=0]},\\&\text{bias}(\text{OLS estimator})\\&\quad = \alpha_2 \cdot (\text{E}[X \mid D=1] - \text{E}[X \mid D=0]).\end{aligned} \tag{1}$$

Based on (1), Jackson and Swanson[5] proposed to assess the potential bias of an IV analysis by: (1) plotting the bias components in (1) for each covariate; (2) displaying a table of bias ratios.

We propose to augment Jackson and Swanson's[5] diagnostics with a plot that considers absolute bias. Large bias ratios can be misleading because not all covariates are confounders. Because the bias ratio of the two terms in (1) does not depend on α_2, even if the bias ratio is large with respect to a covariate, the absolute bias may not be serious if α_2 is close to 0 ($\alpha_2 = 0$ if X is not a confounder). Nevertheless, bias ratios can be more robust than absolute biases when the linear structural model is misspecified. Our diagnostic barplots in the Figure show not only the bias ratios but also the absolute biases. The measured covariates are ordered from top to bottom by the bias amplifying factor $|\alpha_2|\sqrt{\text{Var}(X)}$ (whose square is the variance of Y explained by X), so that we might be most worried about high bias ratios in covariates nearer the top.

We demonstrate our proposal using the neonatal intensive care units (NICU) effectiveness study[3,6] in the Figure. In this study, D is an indicator of a premature baby being delivered in a high technology NICU and Y is in-hospital mortality. The IV $Z = 1$, if the excess travel time to the nearest high technology NICU is more than 10 minutes and $Z = 0$ otherwise. When using the entire dataset (Figure A), it is evident that the IV analysis is helpful to reduce confounding bias due to health condition of the babies (vlbw and lga). Also, although the bias ratio with respect to chronic hypertension (ch) is larger than 1, this variable and the other measured maternal comorbidities are essentially uncorrelated with the outcome and the absolute confounding bias

is negligible. This suggests that bias from the IV being correlated with unmeasured maternal comorbidities might not be concerning. Race (black) has bias ratio larger than 1 and is predictive of the outcome. After conditioning on race (Figure B, C), it appears that the IV analysis might be less biased than regression for white mothers but comparably biased for black mothers.

Qingyuan Zhao
Dylan S. Small
Department of Statistics
The Wharton School
University of Pennsylvania
Philadelphia, PA
qyzhao@wharton.upenn.edu

REFERENCES

1. Davies NM, Smith GD, Windmeijer F, Martin RM. Issues in the reporting and conduct of instrumental variable studies: a systematic review. *Epidemiology*. 2013;24:363–369.
2. Swanson SA, Hernán MA. Commentary: how to report instrumental variable analyses (suggestions welcome). *Epidemiology*. 2013;24:370–374.
3. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Stat Med.* 2014;33:2297–2340.
4. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. *Int J Biostat.* 2007;3.
5. Jackson JW, Swanson SA. Toward a clearer portrayal of confounding bias in instrumental variable applications. *Epidemiology*. 2015;26:498–504.
6. Lorch SA, Baiocchi M, Ahlberg CE, Small DS. The differential impact of delivery hospital on the outcomes of premature infants. *Pediatrics.* 2012;26:270–278.

Validation of Severe Maternal Morbidity on the US Certificate of Live Birth

To the Editor:

Severe maternal morbidity, unexpected outcomes of labor and delivery that have substantial short- or long-term health consequences, has been steadily increasing in recent years, affecting more than 50,000 women in the United States in 2013–2014.[1] It has been suggested that the rise is related to changes in the overall health of the childbearing population, including increases in maternal age,[2] prepregnancy obesity,[3] preexisting chronic medical conditions,[4] and cesarean delivery.[5] The 2003 revision of the US Certificate of Live Birth added specific checkbox items on severe maternal morbidity occurring within 24 hours of delivery to facilitate national surveillance of these outcomes, as well as research on maternal complications of pregnancy. The items included maternal transfusion, third- or fourth-degree perineal laceration, ruptured uterus, unplanned hysterectomy, and admission to the intensive care unit. Massachusetts implemented the 2003 revision in 2011. As part of a larger study evaluating subfertility and maternal–child health, we evaluated the accuracy of severe maternal morbidity reported on the birth certificate, using hospital discharge data as the reference.

We designed a population-based study of all live births in Massachusetts between 2011 and 2013 linked to hospital discharge data. We used International Classification of Diseases, 9th revision (ICD-9) codes in the hospital discharge data to identify maternal blood transfusion (procedure codes 99.x), unplanned hysterectomy (procedure codes 68.3–68.9), ruptured uterus (diagnosis codes 665.0 and 665.1) and, among vaginal births, third- or fourth-degree perineal laceration (diagnosis code 664.2 and 664.3); intensive care unit (ICU) admission was identified by Uniform Billing (UB-04) revenue codes 0200-0209. The rates for each severe maternal morbidity measure were calculated per 100,000 deliveries. Sensitivity (probability of being included on the birth certificate when it was reported on the hospital discharge) and positive predictive value (PPV, probability of being on the hospital discharge when it was reported on the birth certificate), specificity, and negative predictive value were calculated to measure the reporting of severe maternal morbidity measures by mode of delivery (vaginal, cesarean), except for third- or fourth-degree perineal laceration (vaginal births only). We obtained institutional review board approval from the Massachusetts Department of Public Health,

Table. Reporting of Maternal Morbidity on the Birth Certificate Versus Hospital Discharge Massachusetts, 2011–2013

Maternal Morbidity	Mode of Delivery	Rate/100,000 Deliveries		Numerator Present in Both Sources	Sensitivity	Positive Predictive Value
		Birth Certificate	Hospital Discharge			
Blood transfusion	All	171	1053	261	0.12	0.73
	Vaginal	92	637	97	0.11	0.75
	Cesarean	343	1962	162	0.13	0.72
Ruptured uterus	All	102	71	39	0.26	0.18
	Vaginal	45	6	2	0.22	0.03
	Cesarean	229	213	37	0.26	0.25
Unplanned hysterectomy	All	58	75	80	0.51	0.66
	Vaginal	22	16	11	0.48	0.35
	Cesarean	137	203	69	0.52	0.77
Admission to intensive care	All	132	194	81	0.20	0.29
	Vaginal	92	60	18	0.21	0.14
	Cesarean	219	485	63	0.20	0.44
Third- or fourth- degree perineal laceration	Vaginal	928	2746	1190	0.30	0.90

Supported by the National Institute of Child Health and Human Development, National Institutes of Health (grant R01 HD067270). The views expressed in this article are those of the authors and do not necessarily represent the official views of the National Institute of Child Health and Human Development or the National Institutes of Health.
Barbara Luke is a research consultant to the Society for Assisted Reproductive Technologies; all other authors report no conflicts of interest.

ISSN: 1044-3983/2018/2904-0e31
DOI: 10.1097/EDE.0000000000000828