Published in final edited form as:

Paediatr Perinat Epidemiol. 2017 January ; 31(1): 55–63. doi:10.1111/ppe.12322.

Reconstructing a pregnancy cohort to examine potential selection bias in studies on racial disparities in preterm delivery

Katherine J. Sapra^a, Ashok K. Chaurasia^b, Jennifer A. Hutcheon^c, and Katherine A. Ahrens^d

^aDepartment of Obstetrics and Gynecology, Weill Cornell Medical College, New York, NY ^bSchool of Public Health and Health Systems, University of Waterloo, Waterloo, Ontario, Canada

^cDepartment of Obstetrics and Gynaecology, University of British Columbia, Vancouver, British Columbia, Canada ^dDivision of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Rockville, Maryland

Abstract

Background—Epidemiologic studies examining preconception risk factors on perinatal outcomes are typically restricted to live births. By including only non-terminated pregnancies, estimates for the underlying pregnancy cohort may be subject to selection bias. We examined if potential selection bias due to induced termination by maternal race may result in different estimates of the non-Hispanic black — non-Hispanic white risk ratio (RR) for preterm delivery (PTD) among a reconstructed pregnancy cohort (“pseudo-pregnancy cohort”).

Methods—Using New York City registries of 1.6 million live births, spontaneous terminations, and induced terminations among non-Hispanic black and non-Hispanic white women (2000–2012), we multiply imputed PTD (<37 weeks) and early PTD (<32 weeks) outcomes for induced terminations based on maternal race, age, parity, marital status, nativity, and medical care payer to construct the pseudo-pregnancy cohort.

Results—Among non-Hispanic black and non-Hispanic white women 55% and 19% of pregnancies ended in induced termination and 13% and 8% resulted in PTD, respectively. Although several factors were associated with both PTD and induced termination, PTD RRs in the birth (RR 1.64, 95% confidence interval (CI) 1.62, 1.66) and pseudo-pregnancy (RR 1.63, 95% CI 1.56, 1.71) cohorts were similar. However, early PTD RR was somewhat larger in the birth (RR 2.80, 95% CI 2.71, 2.89) than pseudo-pregnancy (RR 2.47, 95% CI 2.23, 2.73) cohort.

Conclusions—Using birth certificate data—thereby excluding induced terminations—to estimate the PTD racial disparity did not produce biased estimates. Our data suggest observed PTD disparities likely are not artifacts of selection bias due to induced termination.

Keywords

preterm birth; racial disparities; loss to follow-up; selection bias; collider bias; induced termination

INTRODUCTION

Racial disparities in birth outcomes have been a concern in the US for decades.¹ Despite calls for action from various prominent institutions and organizations,^{2–4} the risks of preterm delivery and delivering a low birthweight infant remain 60% and 87% higher, respectively, among non-Hispanic black women compared with non-Hispanic white women.⁵ Most evaluations of racial disparities in perinatal outcomes, such as preterm delivery and low birthweight, are restricted to live births, where the outcome can be observed. This excludes the large proportion of pregnancies which would have had an observable perinatal outcome had the pregnancy proceeded. For example, in the US, the 11% and 17% estimated risks of preterm delivery are based upon only 69% and 49% of the total pregnancies among non-Hispanic white and non-Hispanic black women, respectively.^{6,7} Thus, it is possible that only examining live birth data may introduce selection bias when studying preconception risk factors, such as maternal race, on preterm delivery.⁸

As access to contraception and abortion are inconsistent across the US,⁹ the proportion and characteristics of women obtaining induced terminations (and thus excluded from the live birth cohort) can vary over time and geography.^{10,11} For example, women obtaining terminations in geographic areas with restrictive access may be more advantaged (e.g., non-Hispanic white, older, better-educated) and therefore at lower risk of preterm delivery than their counterparts who cannot obtain termination.^{12,13} Consequently, the racial disparity in preterm delivery estimated among live births may not reflect the true disparity in the underlying pregnancy cohort. Therefore, the objective of our analysis was to evaluate the extent to which selection bias due to induced termination might affect estimates of racial disparities in preterm delivery.

METHODS

Study population

Data were obtained from the New York City (NYC) Bureau of Vital Statistics for the years 2000–2012. NYC requires registration of all live births, spontaneous terminations of pregnancy (miscarriages and stillbirths), and induced (elective) terminations of pregnancies irrespective of gestational age at delivery or termination. Live birth data come from birth certificate records while data on spontaneous and induced terminations are collected on separate, shorter, standardised forms completed by hospital staff. Live birth and spontaneous termination data were released as individual level data. Due to privacy concerns, the NYC Bureau of Vital Statistics released only aggregate level data for induced terminations, collapsing individual level data across study years (2000–2012) and abbreviated variable categories to prevent identification of individuals.

Maternal and pregnancy characteristics in our analysis included marital status, nativity, parity, medical care payer, maternal age, maternal race and gestational age at the time the pregnancy ended. These variables were chosen because the data were collected for live births, induced terminations, and spontaneous terminations; there were sufficient numbers of women within strata to allow release of the data for induced terminations; and they were considered to be important for both risk of preterm delivery and induced termination.

Gestational age of live births and spontaneous terminations were recorded as completed weeks since the start of last menstrual period based on the clinical estimate for gestational age. For induced terminations, four collapsed categories of gestational age at time of termination were available for analysis: 0–6, 7–8, 9–12, and 13–24 weeks. The Office of Human Subjects Research at the National Institutes of Health exempted the analysis herein from Institutional Review Board review.

Statistical analysis

We first imputed missing covariate data across all pregnancies using other available maternal characteristics and the pregnancy outcome. We then imputed birth outcomes for all pregnancies ending in induced termination. Birth outcomes for induced terminations were assumed to be missing-at-random (i.e., birth outcomes could be imputed given the maternal characteristics available). The imputation was performed in two steps: (i) assigning a status of live birth or spontaneous termination, and (ii) assigning a gestational age at birth for the imputed live births. The imputation model for live birth included the following variables: nativity, parity, marital status, maternal age, and maternal race. The imputation model for gestational age at birth category included the variables above plus medical care payer type. Medical care payer type was not included in the imputation model for live birth because a category ‘not applicable’ was used for >80% of spontaneous terminations, presumably reflecting that most of the spontaneous terminations were completed without any medical cost.

Multiple Imputation by Chained Equations (MICE) was used to perform the imputations¹⁴ using the R package ‘mice.’¹⁵ The MICE algorithm iteratively imputes values for variables with missing cells by estimating their distribution conditional on other variables. We specified live birth versus spontaneous termination as a dichotomous variable using logistic regression and gestational age at birth as a continuous variable using predictive mean matching. All models used linear additive terms for covariates with all two-way interactions. A total of 100 imputed datasets were generated; each imputed dataset was based on 50 iterations.

Log-binomial regression was used to estimate the risk of preterm delivery in non-Hispanic black compared with non-Hispanic white women. Among live births, preterm delivery (<37 weeks) and early preterm delivery (<32 weeks) were considered separately as outcomes. For each outcome, we ran regression models on two sets of data: the live birth cohort and the pseudo-pregnancy cohort. For the latter, modeling was limited to observations where there was either an observed or imputed live birth (i.e., spontaneous terminations, either observed or imputed, were excluded). The predicted probabilities, RR’s, and risk differences for the pseudo-pregnancy cohort were calculated for each imputed data set and combined using Rubin’s combining rules.¹⁶ These estimates then were compared with those from the live birth cohort. This comparison allowed us to estimate the magnitude of the selection bias introduced when excluding induced terminations, as would be typical when examining only the live birth cohort. As the total effect of race on preterm delivery was of interest, models were not adjusted for other covariates, following the guidance of Kaufman and Cooper.¹⁷

To address the related question of whether an increased induced termination rate in non-Hispanic white women would affect our estimate of the racial disparity in preterm delivery, we conducted a sensitivity analysis in which we reweighted the proportion of pregnancies in non-Hispanic white women ending in induced termination to mirror the proportion observed in non-Hispanic black women (i.e., up-weighted induced terminations and down-weighted live births and spontaneous terminations in non-Hispanic white women so that the weighted distribution of these outcomes in non-Hispanic white women was the same as the observed distribution of these outcomes in non-Hispanic black women). We then used these weights in the log-binomial models of the pseudo-pregnancy cohort and compared the disparity estimates using the weighted data against the non-weighted data. This allowed us to evaluate the racial disparity in preterm delivery in a hypothetical situation where utilization of termination among non-Hispanic white women was as high as among non-Hispanic black women, potentially as a result of more unintended pregnancies among non-Hispanic white women. Data were analyzed in SAS version 9.4 (SAS Institute, Cary, North Carolina).

RESULTS

There were 1 593 256 pregnancies to non-Hispanic black and non-Hispanic white women in NYC from 2000–2012. Of these, 16 498 (1%) were missing data on maternal characteristics or gestational age. Among all pregnancies with complete information, 55% ended in live birth ($n=865\,554$), 5% in spontaneous terminations ($n=82\,711$), and 40% in induced terminations ($n=642\,419$). The study population was 59% non-Hispanic black and 41% non-Hispanic white. Compared with pregnancies in non-Hispanic black women, pregnancies in non-Hispanic white women were more likely to be among women who were older, married, US-born, nulliparous, and not on Medicaid (Table 1).

Among non-Hispanic black women, 55% of pregnancies ended in induced terminations compared with 19% of pregnancies to non-Hispanic white women. The percentage of pregnancies ending in induced termination differed across maternal characteristics and the magnitude of differences differed by maternal race (Table 2). For example, the proportion of pregnancies resulting in induced termination decreased with age for both non-Hispanic white and non-Hispanic black women; however, the decrease was from 60% to 13% from youngest to oldest non-Hispanic white women, but only 68% to 45% in non-Hispanic black women. Several characteristics associated with induced termination were also associated with preterm delivery, such as maternal age, race, and marital status (Table 3). Though the absolute differences in preterm delivery by these characteristics were relatively small (e.g. 2% higher for unmarried/unknown [10%] versus married women [8%]), the relative differences were much greater (e.g., 25% higher for unmarried/unknown versus married women [10%/8%]).

The risks of preterm delivery observed among non-Hispanic black and non-Hispanic white women whose pregnancies ended in live birth (13.0% and 8.0%, respectively) were lower than the imputed risks of preterm delivery in women who had induced terminations for both non-Hispanic black and non-Hispanic white women (16.4% and 14.7%, respectively), resulting in overall preterm delivery risks of 14.9% and 9.1% among non-Hispanic black and non-Hispanic white women, respectively, in the pseudo-pregnancy cohort (Table 3). Despite

the higher risk of preterm delivery in the pseudo-pregnancy cohort, the relative and absolute racial disparities in preterm delivery risk in the pseudo-pregnancy cohort were similar to those in the live birth cohort (Table 4).

For early preterm delivery, the risks were 3.1% and 1.1% among live births to non-Hispanic black and non-Hispanic white women, respectively, while among pregnancies ending in induced termination, imputed early preterm delivery risks were 6.2% and 6.8% in non-Hispanic black and non-Hispanic white women, resulting in overall risks of 5.0% and 2.0% in the pseudo-pregnancy cohort for non-Hispanic black and non-Hispanic white women. Consequently, the absolute racial disparity in early preterm delivery increased while the relative racial disparity decreased. Finally, when the distribution of pregnancy outcomes in non-Hispanic white women was reweighted to match those of non-Hispanic black women, the weighted risks of preterm and early preterm delivery in non-Hispanic white women increased to 11.5% and 4.0%, respectively, and the absolute and relative racial disparities in preterm and early preterm delivery subsequently decreased.

COMMENT

In this New York City pregnancy cohort, a large proportion (40%) of pregnancies were excluded from the live birth cohort due to induced termination. Induced terminations varied markedly by maternal race: 55% and 19% for non-Hispanic black and non-Hispanic white women, respectively. Several characteristics associated with increased risk of induced termination were also associated with increased risk of preterm delivery, including non-Hispanic black race, unmarried marital status, and being US-born. Despite these shared risk factors, accounting for the potential selection bias due to induced termination by multiply imputing missing preterm delivery outcomes had only a slight impact on the absolute and relative estimates of preterm delivery disparities; using birth data only to evaluate racial disparities in preterm delivery resulted in minimal selection bias. Results from our weighted analysis, however, suggested that if the percentage of pregnancies ending in induced termination changed from the status quo and became similar between non-Hispanic black and non-Hispanic white women, the racial disparity would be reduced as the risk of preterm delivery among non-Hispanic white women would increase. This finding highlights that changes in the distribution of characteristics among women who become and remain pregnant, potentially as a consequence of changing access to contraception or termination, might influence estimates of the racial disparity in preterm delivery.

A few prior studies have considered the problem of induced termination in estimating racial disparities in birth outcomes.^{18,19} In a 1980 study, Joyce estimated the effect of induced termination on risks of adverse birth outcomes among non-Hispanic white and non-Hispanic black women; higher risks of adverse outcomes, including preterm delivery, were estimated in both non-Hispanic black and non-Hispanic white women whose pregnancies ended in induced terminations,¹⁸ similar to our findings. In a 1981 follow-up study, however, Joyce and Grossman concluded higher utilization of induced termination reduced the risk of observed low birthweight in infants born to non-Hispanic black, but not non-Hispanic white, women.¹⁹

We applied existing epidemiological theory on selection bias, particularly as it relates to collider bias.^{20,21,22} Collider bias arises when conditioning upon a common effect (collider) of the exposure and outcome.²⁰ Berkson's bias, which arises when restricting the analysis to a single stratum of the collider (e.g., observable gestational age), could apply to our scenario.²¹ In this instance, we posited that maternal race was the exposure, true gestational age at delivery had all pregnancies proceeded to their uninterrupted outcome was the outcome (dichotomized into preterm delivery vs. not preterm delivery), and observable gestational age (i.e., observed birth) was the collider. By conditioning upon observable gestational age, a spurious relationship between maternal race and underlying true gestational age at delivery could appear. However, we did not observe biased estimates. It may be because our hypothesized causal structure was incorrect or relationships were not very strong; if observed gestational age depended weakly or not at all on underlying true gestational age at delivery, bias would not be expected.²¹ Furthermore, it does not appear than an unmeasured common cause of the collider and outcome, such as access to healthcare, created collider bias in these data.²¹ Thus, it seems unlikely that strong collider bias was present in this analysis of racial disparities in preterm delivery among births, although some bias was found for the disparities in early preterm delivery.

The potential for selection bias in our study could be further classified as left truncation bias, as others in reproductive and perinatal epidemiology have previously suggested is the case with spontaneous terminations occurring prior to study enrollment.^{23,24,25} When using birth certificate data, which only registers deliveries occurring after a specified gestational age (e.g. 20 weeks), the underlying pregnancy cohort is left truncated; that is, only those pregnancies surviving long enough to reach the gestational age for registration are included. This could cause bias in the exposure-outcome estimate (i.e. racial disparity in preterm delivery) if the risk of left truncation (i.e. induced termination) is differential by exposure (i.e. maternal race). A previous simulation study suggested that left truncation bias due to increased risk of spontaneous termination among smokers may explain the observed inverse association between maternal smoking and preeclampsia.²⁴ In our paper, we were able to explore these potential biases using empirical data.

New York City has a higher rate of induced terminations per capita than the United States as a whole. In our population, 55% and 19% of pregnancies to non-Hispanic black and non-Hispanic white women, respectively, ended in induced termination compared with 36% and 12% among non-Hispanic black and non-Hispanic white women nationally.⁷ As there is no national database of registered induced terminations in the US, it would be challenging to conduct a similar analysis on a national level. Still, it may be useful to compare our study findings to state-level data from jurisdictions where induced termination is less easily accessed and data on pregnancy outcomes are available.

Limitations of the study

Among the study's limitations was that the NYC's Bureau of Vital Statistics only provided aggregate level data on maternal characteristics collapsed across 2000–2012 for the induced terminations to prevent identification of individuals. Thus, details on maternal characteristics were limited. Also, by combining all years of data, we obtained only an average estimate;

thus, we may have missed clarifying details on how secular trends in utilization of termination and preterm delivery risk might have impacted the racial disparity in preterm delivery for the pseudo-pregnancy cohort. In addition, spontaneous terminations were likely underreported, accounting for only 5% of pregnancies in our pregnancy cohort, which is approximately 10% lower than expected.²⁶ Spontaneous terminations occurring at later gestational ages, particularly those after 6 weeks, were presumably more likely to be registered than earlier losses, as pregnancies ending in very early loss are less likely to receive medical care^{27,28} and, therefore, are less likely to be registered. While 5% and 6% of pregnancies ended in spontaneous termination among non-Hispanic black and non-Hispanic white women, respectively, some evidence suggests that non-Hispanic black women have an increased risk (2% excess risk) of spontaneous termination at 10 weeks of gestation.²⁹ Though spontaneous terminations are more likely to be registered at later gestational ages, the possibility exists that spontaneous terminations to non-Hispanic black women are underreported at a slightly greater rate than for non-Hispanic white women in the NYC registry. However, because the absolute increased risk of spontaneous termination for non-Hispanic black women is low, this potential underreporting likely has minimal impact on our results.

The validity of our imputations depends on the extent to which data were truly missing-at-random, an assumption challenging to verify in practice. Missing-at-random assumes that the observed missingness in a given dataset is dependent solely on the observed data, i.e., the missing information can in some sense be “recovered” by exploiting relationships seen in the observed data. Hence, the missing-at-random assumption becomes more plausible when more characteristics (and their relationships with other variables) are included in the imputation model.²⁹ Given confidentiality concerns, we had a relatively small set of variables available for analysis; however, these variables included maternal race, age, marital status, and medical care payer (as a proxy for socioeconomic status), which are key risk factors for preterm delivery³⁰ and are also associated with the missingness of the data (i.e., induced termination). Other important predictors of preterm delivery, such as prior preterm delivery, are not collected for induced terminations and so were not available for the multiple imputation model. The areas under the curve (AUC) for preterm and early preterm delivery were 0.60 and 0.67, respectively, which is relatively low, reflecting both the difficulty in predicting preterm delivery generally and the absence of certain important predictors of preterm delivery from the model.

Strengths of the study

Strengths of our study include the fact that, in New York City, all live births, induced terminations, and spontaneous terminations are required by law to be reported to the Department of Health and Mental Hygiene, and maternal characteristics and gestational age are recorded for each outcome. The large number of pregnancies occurring in NYC also provided greater power to detect small differences in racial disparities in preterm delivery in the live birth cohort vs. the pseudo-pregnancy cohort. Reporting of live births was nearly 100% complete,³¹ and the sensitivity of the induced termination reporting was 88% relative to induced termination surveillance.³² As a result of this relatively complete data on induced

terminations, our cohort was uniquely able to address the problem of ‘missing’ birth outcomes due to induced termination.

Conclusions

In this cohort of NYC pregnancies, we found that women whose pregnancies ended in induced termination had somewhat greater risks of preterm delivery than their counterparts whose pregnancies ended in live birth; however, the racial disparity in the pseudo-pregnancy cohort was not substantially different from the disparity observed in the birth cohort. Epidemiologists interested in disparities in birth outcomes may be encouraged by the relatively small magnitude of difference in preterm delivery estimates between the birth and pseudo-pregnancy cohorts, particular given limited availability of induced termination data in most instances. These results also provide reassurance that the racial disparities in preterm delivery observed in existing research using birth certificate data likely are not artifacts of selection bias due to induced termination, at least in areas where termination is readily accessed.

Acknowledgments

We thank the New York City Department of Health and Mental Hygiene, Bureau of Vital Statistics for providing the data. This work was completed while Drs. Sapra, Chaurasia, and Ahrens were with the Division of Intramural Population Health Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, Rockville, Maryland.

REFERENCES

1. Alexander, G. Washington, DC: The National Academies Press; 2006. Prematurity at birth: determinants, consequences and geographic variation; p. 501-535.
2. American Public Health Association. Reducing Racial/Ethnic and Socioeconomic Disparities in Preterm and Low Birthweight Births (Policy No. 20062). 2006 [Accessed July 29, 2015] <http://www.apha.org/policies-and-advocacy/public-health-policy-statements/policy-database/2014/07/18/10/01/reducing-racial-ethnic-and-socioeconomic-disparities-in-preterm-and-low-birthweight-births>.
3. Butler, AS.; Behrman, RE. National Academies Press; 2007. Preterm Birth: Causes, Consequences, and Prevention.
4. March of Dimes. March of Dimes 2014 Premature Birth Report Card. 2014 [Accessed July 29, 2015] <http://www.marchofdimes.org/materials/premature-birth-report-card-united-states.pdf>.
5. Martin JA, Osterman M. Preterm Births—United States, 2006 and 2010. CDC Health Disparities and Inequalities Report—United States, 2013. 2013; 62(3):136.
6. Martin JA, Hamilton BE, Sutton PD, Ventura SJ, Mathews T, Kirmeyer S, et al. Births: Final data for 2007. National Vital Statistics Reports. 2010; 58:1–125.
7. Ventura SJ, Curtin SC, Abma JC, Henshaw SK. Estimated pregnancy rates and rates of pregnancy outcomes for the United States, 1990–2008. National vital statistics reports: from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System. 2012; 60(7):1–21.
8. Kramer MS, Zhang X, Platt RW. Analyzing risks of adverse pregnancy outcomes. American Journal of Epidemiology. 2014; 179:361–367. [PubMed: 24287468]
9. Guttmacher Institute. 2015 year-end state policy roundup. 2016 [April 13, 2016] Available from: <https://www.guttmacher.org/article/2016/01/2015-year-end-state-policy-roundup>.
10. Grossman D, Baum S, Fuentes L, et al. Change in abortion services after implementation of a restrictive law in Texas. Contraception. 2014; 90(5):496–501. [PubMed: 25128413]

11. Jerman J, Jones RK. Secondary measures of access to abortion services in the United States, 2011 and 2012: gestational age limits, cost, and harassment. *Women's Health Issues*. 2014; 24(4):e419–e424. [PubMed: 24981401]
12. Upadhyay UD, Weitz TA, Jones RK, Barar RE, Foster DG. Denial of abortion because of provider gestational age limits in the united states. *American Journal of Public Health*. 2014; 104:1687–1694. [PubMed: 23948000]
13. Cook PJ, Parnell AM, Moore MJ, Pagnini D. The effects of short-term variation in abortion funding on pregnancy outcomes. *Journal of Health Economics*. 1999; 18:241–257. [PubMed: 10346355]
14. Little, RJ.; Rubin, DB. *Statistical Analysis with Missing Data*. John Wiley & Sons; 2007.
15. Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*. 2011; 45(3)
16. Rubin, D. *Multiple Imputation for Nonresponse in Surveys*. New York: J. Wiley & Sons; 1987.
17. Kaufman JS, Cooper RS. Commentary: considerations for use of racial/ethnic classification in etiologic research. *American Journal of Epidemiology*. 2001; 154(4):291–298. [PubMed: 11495850]
18. Joyce T. The impact of induced abortion on black and white birth outcomes in the United States. *Demography*. 1987; 24(2):229–244. [PubMed: 3609406]
19. Joyce T, Grossman M. The dynamic relationship between low birthweight and induced abortion in New York City: An aggregate time-series analysis. *Journal of health economics*. 1990; 9(3):273–288. [PubMed: 10107847]
20. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004; 15:615–625. [PubMed: 15308962]
21. Westreich D. Berkson's bias, selection bias, and missing data. *Epidemiology*. 2012; 23:159–164. [PubMed: 22081062]
22. Pearce N, Richiardi L. Commentary: Three worlds collide: Berkson's bias, selection bias and collider bias. *International Journal of Epidemiology*. 2014; 43:521–524. [PubMed: 24585732]
23. Howards PP, Hertz-Picciotto I, Poole C. Conditions for bias from differential left truncation. *American Journal of Epidemiology*. 2007; 165(4):444–452. [PubMed: 17150983]
24. Lisonkova S, Joseph KS. Left truncation bias as a potential explanation for the protective effect of smoking on preeclampsia. *Epidemiology*. 2015; 26(3):436–440. [PubMed: 25695352]
25. Schisterman EF, Cole SR, Ye A, Platt RW. Accuracy loss due to selection bias in cohort studies with left truncation. *Paediatric and Perinatal Epidemiology*. 2013; 27(5):491–502. [PubMed: 23930785]
26. American College of Obstetricians and Gynecologists. *Early Pregnancy Loss*. 2013
27. Wang X, Chen C, Wang L, Chen D, Guang W, French J. Conception, early pregnancy loss, and time to clinical pregnancy: a population-based prospective study. *Fertility and Sterility*. 2003; 79(3):577–584. [PubMed: 12620443]
28. Wilcox AJ, Weinberg CR, O'Connor JF, et al. Incidence of early loss of pregnancy. *The New England Journal of Medicine*. 1988; 319(4):189–194. [PubMed: 3393170]
29. Mukherjee S, Velez Edwards DR, Baird DD, Savitz DA, Hartmann KE. Risk of miscarriage among black women and white women in U.S. prospective cohort study. *American Journal of Epidemiology*. 2013; 177(11):1271–1278. [PubMed: 23558353]
30. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychological Methods*. 2002; 7(2):147. [PubMed: 12090408]
31. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. *The Lancet*. 2008; 371:75–84.
32. Centers for Disease Control and Prevention. [Accessed July 29, 2015] User Guide to the 2012 Natality Public Use File. 2013. ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/DVS/natality/UserGuide2012.pdf
33. Toprani A, Madsen A, Das T, Gambatese M, Greene C, Begier E. Evaluating new york city's abortion reporting system: Insights for public health data collection systems. *Journal of Public Health Management and Practice*. 2014; 20:392–400. [PubMed: 24281129]

Table 1

Characteristics of Pregnancies Registered to Non-Hispanic White and Non-Hispanic Black Women in New York City, 2000–2012

	Non-Hispanic Black (59%) n=936,141	Non-Hispanic White (41%) n=657,115
	n (%)	n (%)
Maternal age (years)		
<20	129,265 (14)	22,281 (3)
20–24	248,651 (27)	103,758 (16)
25–29	229,382 (25)	148,808 (23)
30–34	178,878 (19)	198,650 (30)
35	149,690 (16)	183,384 (28)
Missing	275 (0)	234 (0)
Marital Status		
Married	177,389 (19)	479,477 (73)
Unmarried/Unknown	751,523 (80)	171,349 (26)
Missing ^a	7,229 (1)	6,289 (1)
Nativity		
US-born	560,344 (60)	446,981 (68)
Foreign-born/Unknown	375,797 (40)	210,134 (32)
Parity		
Nulliparous	362,893 (39)	327,516 (50)
Parous	566,212 (60)	323,473 (49)
Missing ^a	7,036 (1)	6,126 (1)
Payer		
Medicaid	469,899 (50)	151,364 (23)
Other payer	431,290 (46)	472,166 (72)
Not applicable ^b	34,952 (4)	33,585 (5)

^aMissing category was only applicable for spontaneous terminations in 2011–2012 when marital status and parity were not collected

^bThis category was only applicable for spontaneous terminations (2000–2012)

Table 2
Maternal Characteristics and Pregnancy Outcome by Maternal Race, New York City 2000–2012

	Non-Hispanic Black (n=936,141)				Non-Hispanic White (n=657,115)			
	Live birth (40%) n=374,607	Spontaneous termination (5%) n=44,108	Induced termination (55%) n=517,426	Live birth (75%) n=492,663	Spontaneous termination (6%) n=39,459	Induced termination (19%) n=124,993		
	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)		
Maternal age (years)								
<20	37,247 (29)	3,606 (3)	88,412 (68)	8,175 (37)	761 (3)	13,345 (60)		
20–24	88,312 (36)	8,272 (3)	152,067 (61)	67,061 (65)	3,867 (4)	32,830 (32)		
25–29	94,939 (41)	9,536 (4)	124,907 (54)	110,227 (74)	6,798 (5)	31,783 (21)		
30–34	83,964 (47)	9,587 (5)	85,327 (48)	164,259 (83)	11,094 (6)	23,297 (12)		
35	70,145 (47)	12,832 (9)	66,713 (45)	142,941 (78)	16,705 (9)	23,738 (13)		
Missing	0 (0)	275 (100)	0 (0)	0 (0)	234 (100)	0 (0)		
Marital Status								
Married	118,987 (67)	3,576 (2)	54,826 (31)	437,208 (91)	13,637 (3)	28,632 (6)		
Unmarried/Unknown	255,620 (34)	33,303 (4)	462,600 (62)	55,455 (32)	19,533 (11)	96,361 (56)		
Missing ^a	0 (0)	7,229 (100)	0 (0)	0 (0)	6,289 (100)	0 (0)		
Nativity								
US-born	204,015 (36)	23,498 (4)	332,831 (59)	344,533 (77)	24,987 (6)	77,461 (17)		
Foreign-born/Unknown	170,592 (45)	20,610 (5)	184,595 (49)	148,130 (70)	14,472 (7)	47,532 (23)		
Parity								
Nulliparous	160,416 (44)	14,809 (4)	187,668 (52)	231,567 (71)	16,181 (5)	79,768 (24)		
Parous	213,987 (38)	22,467 (4)	329,758 (58)	260,993 (81)	17,255 (5)	45,225 (14)		
Missing ^a	204 (3)	6,832 (97)	0 (0)	103 (2)	6,023 (98)	0 (0)		
Payer								
Medicaid	240,510 (51)	3,048 (1)	226,341 (48)	127,406 (84)	616 (0)	23,342 (15)		
Other payer	134,097 (31)	6,108 (1)	291,085 (67)	365,257 (77)	5,258 (1)	101,651 (22)		
Not applicable ^b	NA	34,952 (100)	NA	NA	33,585 (100)	NA		

^a Missing category was only applicable for spontaneous terminations in 2011–2012 when marital status and parity were not collected

This category was only applicable for spontaneous terminations (2000–2012)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3
Percent of Births Occurring Preterm by Maternal Characteristics, New York City, 2000–2012

	Live birth cohort (n=867,270)		Induced termination cohort ^a (n=557,459)		Pseudo-pregnancy cohort ^b (n=1,424,729)	
	Non-Hispanic Black (n=374,607)	Non-Hispanic White (n=492,663)	Non-Hispanic Black (n=456,076)	Non-Hispanic White (n=101,383)	Non-Hispanic Black (n=830,683)	Non-Hispanic White (n=594,046)
	% Preterm	% Preterm	% Preterm	% Preterm	% Preterm	% Preterm
Overall	13	8	16	15	15	9
Maternal age (years)						
<20	12	9	14	15	14	12
20–24	11	6	15	15	14	9
25–29	12	7	16	14	14	8
30–34	13	8	18	14	15	9
35	16	10	22	15	19	10
Marital Status						
Married	12	8	13	9	12	8
Unmarried/ Unknown	14	10	17	17	16	14
Nativity						
US-born	14	8	16	15	15	9
Foreign-born/ Unknown	12	7	16	14	14	8
Parity						
Nulliparous	12	8	15	15	14	9
Parous	14	8	17	13	16	9
Payer						
Medicaid	13	6	14	10	13	7
Other payer	14	8	19	16	17	10

^aIncludes induced terminations that were imputed as having live births

^bIncludes all live births and induced terminations that were imputed as having live births

Table 4

Risk of Preterm Delivery and Early Preterm Delivery in Non-Hispanic Black Compared with Non-Hispanic White Women, New York City 2000–2012

	Risk Ratio (95% confidence interval)	Risk Difference (95% confidence interval) (Excess cases per 100 births)
Preterm delivery<37 weeks		
Birth cohort	1.64 (1.62, 1.66)	5.07 (4.94, 5.20)
Pseudo-pregnancy cohort ^a	1.63 (1.56, 1.71)	5.76 (5.15, 6.38)
Weighted analysis ^b	1.29 (1.21, 1.38)	3.38 (2.51, 4.25)
Preterm delivery<32 weeks		
Birth cohort	2.80 (2.71, 2.89)	1.98 (1.92, 2.04)
Pseudo-pregnancy cohort ^a	2.47 (2.23, 2.73)	2.96 (2.60, 3.33)
Weighted analysis ^b	1.26 (1.10, 1.45)	1.03 (0.46, 1.60)

^a After multiple imputation of outcomes for induced terminations

^b Assuming similar proportion of induced terminations among Non-Hispanic White women as observed among Non-Hispanic Black women