**Sensitivity analysis: Impact of potential selection bias from excluding missing data**

In our analysis, observations with missing data on the exposure (race), outcome (preterm birth), or mediator (trimester of prenatal care entry) were excluded: 2.2% of deliveries had missing data on race, 0.3% had missing data on gestational age (from which the outcome, preterm birth delivery was calculated), and 3.1% had missing data on date of prenatal care entry (which, along with gestational age, is how the mediator, trimester of prenatal care entry, was calculated). As a result, a total of 37,298 deliveries were excluded.

Deterministic imputation was used to assign values for the missing data based on three potential scenarios. For each scenario, we re-ran the mediation analysis on our primary outcome (total preterm birth) to assess the impact of each scenario on our findings:

1) The first scenario assumed a "worst" case scenario where all observations missing data on the exposure (race) were Black non-Hispanic, missing data on the mediator (timing of prenatal care entry) were those with inadequate care, and missing data on the outcome (preterm birth) were those who experienced the outcome,

2) The second scenario assumed an “intermediate” case scenario where all observations missing data on exposure (race) were Hispanic, and missing data on the mediator and outcome were treated in the same way as in scenario 1, and

3) The third scenario assumed a "best" case scenario where all observations missing data on the exposure (race) were White non-Hispanic, missing data on the mediator (adequacy of prenatal care) were those with adequate care, and missing data on the outcome (preterm birth) were those who did not experience the outcome.

Supplementary tables 1, 2, and 3 below shows the results of each of the scenarios. The sensitivity analysis shows that the various scenarios slightly exaggerated or attenuated the % mediated, but overall, excluding the missing data did not introduce significant bias to our results.

Supplementary table 1. “Worst” Case Scenario

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Black non-Hispanic**RR (95% CI) | **Hispanic**RR (95% CI) | **Other**RR (95% CI) |
| **Preterm Birth <37 Weeks** |
| Direct Effect | 1.0361 (1.0295, 1.0429) | 1.0739 (1.0601, 1.0879) | 1.1134 (1.0919, 1.1353) |
| Indirect Effect | 1.0084 (1.0077, 1.0090) | 1.0136 (1.0116, 1.0156) | 1.0153 (1.0106, 1.0201) |
| Total Effect | 1.0448 (1.0381, 1.0515) | 1.0885 (1.0747, 1.1026) | 1.1305 (1.1088, 1.1526) |
| % Mediated | 19.4 | 16.5 | 13.1 |

Supplementary table 2. “Intermediate” Case Scenario

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Black non-Hispanic**RR (95% CI) | **Hispanic**RR (95% CI) | **Other**RR (95% CI) |
| **Preterm Birth <37 Weeks** |
| Direct Effect | 1.0373 (1.0304, 1.0442) | 1.0735 (1.0599, 1.0873) | 1.1127 (1.0916, 1.1342) |
| Indirect Effect | 1.0059 (1.0052, 1.0065) | 1.0140 (1.0120, 1.0159) | 1.0165 (1.0119, 1.0211) |
| Total Effect | 1.0434 (1.0364, 1.0503) | 1.0885 (1.0748, 1.1023) | 1.1310 (1.1097, 1.1528) |
| % Mediated | 14.1 | 16.9 | 14.0 |

Supplementary table 3. “Best” Case Scenario

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Black non-Hispanic**RR (95% CI) | **Hispanic**RR (95% CI) | **Other**RR (95% CI) |
| **Preterm Birth <37 Weeks** |
| Direct Effect | 1.0365 (1.0298, 1.0432) | 1.0745 (1.0606, 1.0885) | 1.1141 (1.0926, 1.1360) |
| Indirect Effect | 1.0044 (1.0038, 1.0050) | 1.0065 (1.0046, 1.0083) | 1.0058 (1.0014, 1.0103) |
| Total Effect | 1.0410 (1.0343, 1.0478) | 1.0814 (1.0676, 1.0954) | 1.1205 (1.0990, 1.1425) |
| % Mediated | 11.1 | 8.5 | 5.4 |