Supplementary Information

Preconception diabetes and hypertension and severe maternal-neonatal morbidity among non-Hispanic Black women in the United States: a national retrospective cohort study of risk pathways

Supplementary Table S1. Temporally structured DAG: 41 variables across nine temporal tiers

Tier	Variables
1. Maternal sociodemographic	advanced maternal age (≥35), low education (≤ high school), unmarried, WIC, Hispanic ethnicity, foreign-born nativity
2. Paternal	father age risk, father low education, father race minority, father Hispanic
3. Preconception	obesity, smoking, PGDM, PGHTN, infertility, ART, prior preterm birth, prior infant death, short inter-pregnancy interval
4. Prenatal access	late/no prenatal care, ≤2 prenatal visits
5. Pregnancy complications	GDM, gestational hypertension, eclampsia, chorioamnionitis
6. Congenital anomalies	anencephaly, spina bifida, cyanotic congenital heart disease, diaphragmatic hernia, omphalocele, gastroschisis
7. Birth outcomes	multiple gestation, preterm birth, low birthweight, cesarean delivery, abnormal weight gain, low 5-min Apgar, male sex
8. Maternal morbidity	severe maternal morbidity (SMM) composite
9. Neonatal outcome	CRITICAL composite, infant death

The DAG contains 82 expert-specified directed edges (all forward in time). Medicaid (the payer for the delivery) is deliberately excluded from the causal DAG because it is recorded at delivery and therefore cannot serve as a preconception confounder or mediator; it is reported descriptively only. Edges were specified from obstetric/perinatal epidemiology literature; no data-driven model search was used for the primary analysis.

Supplementary Table S2. DAG-derived minimal back-door adjustment sets for each prespecified exposure

Exposure	Back-door adjustment set	Notes
Pre-gestational diabetes (PGDM)	advanced maternal age; obesity; pre-gestational hypertension; smoking	Same set for CRITICAL, SMM and the composite outcome
Pre-gestational hypertension (PGHTN)	advanced maternal age; obesity; pre-gestational diabetes; smoking	Same set for CRITICAL, SMM and the composite outcome
Gestational hypertension	obesity	Defined among women without PGHTN (PGHTN=0); restriction closes all back-door paths through PGHTN
Gestational diabetes (GDM)	obesity	Defined among women without PGDM (PGDM=0); restriction closes all back-door paths through PGDM
Assisted reproductive technology (ART)	advanced maternal age	Reported as an association: ART is a subset of any infertility treatment, so the infertility-treatment marker is not adjusted
Advanced maternal age ($\geq$35)	none	No ancestors in the DAG; no back-door paths exist

All sets were derived from the temporally structured DAG by the back-

door criterion and verified to d-separate the exposure from each outcome in the graph with all edges out of the exposure removed. Severe SMM (4-item proxy) occupies the same causal position in the DAG as the SMM composite and therefore uses the same adjustment set as SMM for each exposure. GDM and gestational hypertension effects are estimated in the stratum without the corresponding pre-gestational disease, in line with the mutually exclusive NCHS coding.

Supplementary Table S3. Extended-baseline sensitivity of composite-endpoint ACEs (Black women, 2023)

Exposure	Adjustment set	ACE (pp)	95% CI	n
PGDM	DAG minimal back-door set	16.25	15.17–17.33	475,676
PGDM	Minimal + WIC, low education, unmarried, nativity	16.15	15.03–17.37	449,976
PGHTN	DAG minimal back-door set	10.08	9.53–10.60	475,676
PGHTN	Minimal + WIC, low education, unmarried, nativity	10.21	9.64–10.84	449,976

Point estimates on the composite (SMM or CRITICAL) endpoint. The extended set adds the four sociodemographic baseline variables (WIC receipt, low education, unmarried status, foreign-born nativity) to the minimal DAG back-door sets; the effective sample shrinks because marital status is not reported by some states (11.1% of all records), with such records excluded from the marital-status denominator rather than coded as unmarried. Adding the full sociodemographic baseline changes the composite ACEs by less than 0.15 percentage points (PGDM −0.10; PGHTN +0.13), indicating that the minimal adjustment sets are not the source of the effect estimates.

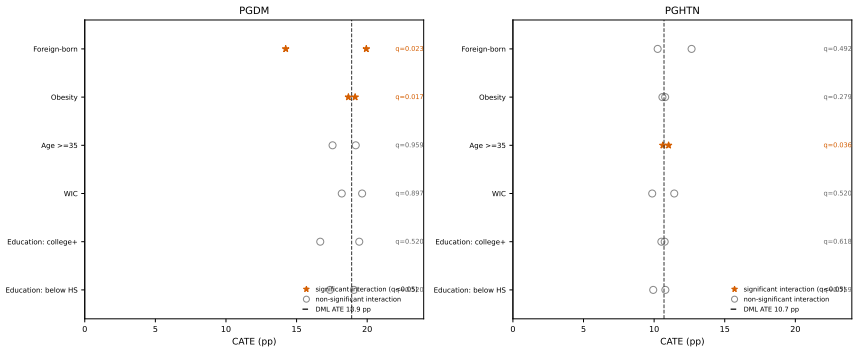

Figure 1: **Supplementary Figure S1.** DML CATE dot plot: effect heterogeneity within Black women. Double/debiased machine learning conditional average treatment effects (CATEs) for pre-gestational diabetes mellitus (PGDM) and pre-gestational hypertension (PGHTN) across six prespecified subgroups among non-Hispanic Black women (2023). For each subgroup, points show the CATE point estimate in each group (e.g., obese vs non-obese); asterisks mark subgroups with a statistically significant interaction and open circles mark the remainder. The dashed vertical line marks the overall DML ATE (18.9 pp for PGDM; 10.7 pp for PGHTN). Benjamini–Hochberg FDR-adjusted q-values are shown for each subgroup; significant modifiers ($q < 0.05$) are obesity and foreign-born nativity for PGDM, and age $\geq$35 for PGHTN. See Supplementary Table S10 for numerical values.

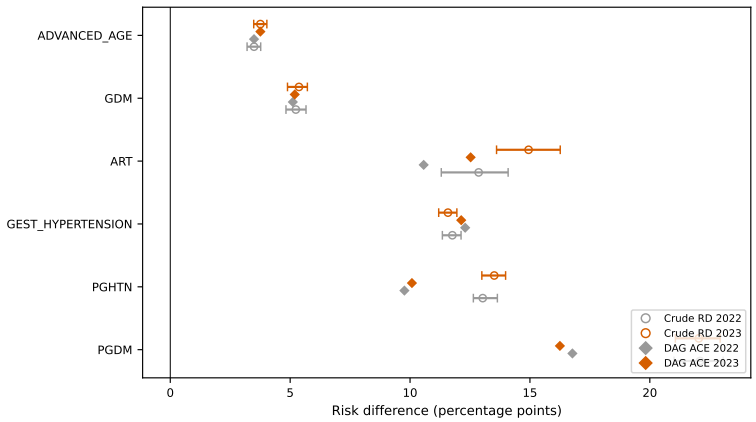

Figure 2: **Supplementary Figure S2.** Temporal replication: 2022 vs 2023 crude and adjusted effect estimates. Rank-order concordance of DAG-adjusted average causal effects (ACEs) across the six prespecified exposures between the 2022 and 2023 NCHS linked birth/infant-death cohorts. For each exposure, crude risk differences (open circles with horizontal 95% confidence intervals) and DAG-adjusted ACEs (filled diamonds) on the composite morbidity endpoint are plotted for 2022 (grey) and 2023 (orange). Spearman rank correlation of adjusted effects $\rho = 0.943$ ($p = 0.005$). ART and gestational hypertension swap second and third positions across years; all other exposures retain their rank.

Supplementary Table S4. Interventional mediation: full decomposition and disparity attribution

Pathway	Covariates adjusted for	Total effect (pp)	Direct effect (pp)	Indirect effect (pp)	Proportion (95% CI)
ART → multiple gestation → CRITICAL (NH Black)	advanced maternal age	11.61 (10.20–12.80)	6.77 (5.48–7.91)	4.84 (4.44–5.24)	41.7% (37.3–47.0)
NH White	as above	7.39 (7.01–7.78)	4.97 (4.63–5.31)	2.42 (2.32–2.54)	32.7% (31.2–34.2)
PGDM → preterm birth → CRITICAL (NH Black)	advanced maternal age; ART; cesarean delivery; eclampsia; gestational diabetes; gestational hypertension; infertility treatment; multiple gestation; obesity; pre-gestational hypertension; smoking	17.20 (15.99–18.16)	6.58 (5.56–7.19)	10.63 (10.00–11.30)	61.8% (59.2–65.3)
NH White	as above	18.58 (17.84–19.27)	7.24 (6.68–7.74)	11.34 (11.00–11.72)	61.1% (59.4–63.0)
PGHTN → preterm birth → CRITICAL (NH Black)	advanced maternal age; cesarean delivery; eclampsia; gestational diabetes; gestational hypertension; multiple gestation; obesity; pre-gestational diabetes; smoking	10.30 (9.87–10.84)	3.78 (3.42–4.18)	6.53 (6.26–6.80)	63.4% (61.0–65.5)
NH White	as above	9.58 (9.30–10.04)	4.77 (4.54–5.12)	4.81 (4.62–5.00)	50.2% (48.2–51.9)
Racialised group → prenatal care access → composite morbidity (pooled)	advanced maternal age; nativity; low education; unmarried	1.94 (1.63–2.30)	1.85 (1.55–2.21)	0.09 (0.07–0.10)	4.4% (3.4–5.7)
Racialised group → deprivation → metabolic disease (pooled)	advanced maternal age; nativity	3.54 (3.36–3.79)	3.14 (2.94–3.40)	0.40 (0.27–0.47)	11.4% (7.6–13.6)
Racialised group → deprivation → composite morbidity (pooled)	advanced maternal age; nativity	3.14 (2.86–3.46)	2.37 (2.11–2.73)	0.77 (0.64–0.87)	24.5% (19.9–28.0)

Covariates for the modifiable-preconception pathways are the union of the DAG-derived back-door sets for the exposure-outcome, exposure-mediator and mediator-outcome relations. For the metabolic pathways this set includes the mediator-outcome confounders gestational disease (gestational diabetes, gestational hypertension, eclampsia), cesarean delivery and multiple gestation, all of which occur after the exposure. Interventional ef-

fects were evaluated by Monte Carlo simulation: post-exposure covariates were drawn from their conditional distributions under the relevant counterfactual exposure level, fitted as logistic models $P(L_j \mid X, C_{\text{base}})$ (the joint distribution approximated as the product of the marginal models), and the outcome and mediator models were evaluated at the simulated values; pathways without post-exposure covariates (ART; the disparity decompositions) used the standard regression-based mediation formula. NCHS mutual-exclusivity coding ensures all simulated exposure–L combinations lie within the observed support. Effects are on the risk-difference scale in percentage points; the proportion mediated is the indirect-to-total risk-difference ratio, and 95% confidence intervals are percentile bootstrap (200 resamples for Black women and the disparity decompositions, 100 for White women in the modifiable-preconception pathways). The disparity-decomposition estimates are descriptive proportions of the racial disparity attributable to each mediator, not causal effects of racial category. Because the temporal position of cesarean delivery relative to preterm birth is uncertain, a sensitivity analysis that excluded cesarean delivery from the post-exposure set (Black women, 100 resamples) gave PGDM and PGHTN proportions of 51.5% and 54.2%, respectively; full sensitivity results are reported in the analysis-code repository.

Supplementary Table S5. Policy-scenario input parameters (non-Hispanic Black women, 2023)

Burden estimates were computed as $N \times$ prevalence $\times$ effectiveness $\times$ risk difference, with $N = 491{,}753$ Black women and effectiveness varied across a uniform 20/40/60/80% ladder for all four scenarios. Each scenario used a distinct risk-difference quantity: the DAG-adjusted ACE of a composite "PGDM or PGHTN" exposure (metabolic), the interventional indirect effect of late/no prenatal care (access; estimated on the pooled Black/White sample with adjustment for advanced maternal age, nativity, low education and unmarried status, matching the disparity decomposition in Supplementary Table S4), the crude preterm-birth association (preterm-birth-related burden reduction; preterm is a downstream mediator, so the RD is labelled as an association), and the outcome-specific Black–White excess (racial disparity reduction target; not a causal RD). Severe-SMM inputs use the four-item severe-SMM proxy.

All risk differences are in percentage points (pp). Risk differences are shown to two decimal places; back-calculation from Table 4 of the main text

Scenario	Risk-difference input	SMM	CRITICAL	Composite
Metabolic risk-difference reduction	DAG-adjusted ACE, composite PGDM/PGHTN exposure (prevalence 6.941%; union of PGDM/PGHTN back-door sets excluding the exposure components)	0.90	12.72	12.89
Prenatal care access	Interventional indirect effect of late/no prenatal care (Black vs White)	0.003	0.10	0.09
Preterm-birth-related burden reduction	Crude preterm-birth RD, association (preterm prevalence 14.650%)	1.84	49.09	48.68
Racial disparity reduction target	Black–White outcome excess (disparity target)	0.33	3.25	2.98

may differ marginally owing to rounding. Population prevalence is shown for the exposure entering the formula (metabolic composite and preterm birth); the access and racial disparity reduction scenarios enter through effect quantities estimated on the pooled Black/White sample, so no exposure prevalence is listed. For the metabolic composite, outcome-specific 95% confidence intervals (200 bootstrap resamples) were 0.66–1.05 (severe SMM), 12.23–13.30 (CRITICAL) and 12.41–13.40 pp (composite), with analytic samples of 475,055, 475,421 and 475,676 respectively; the adjustment set is the union of the PGDM and PGHTN back-door sets excluding the exposure components themselves (so the composite exposure is not adjusted for its own constituent diseases). Ranges in Table 4 of the main text correspond to the 20–80% effectiveness ladder applied to these inputs. The SMM, CRITICAL and Composite columns report the outcome-specific risk differences (pp) entering the burden formula, not case counts.

Supplementary Table S6. Software environment

Component	Version
Python	3.12.7
pgmpy	0.1.25
EconML	0.16.0
scikit-learn	1.5.1
statsmodels	0.14.2
matplotlib	3.10.8
NumPy	1.26.4
pandas	2.2.2
NetworkX	3.3

All analyses were run with these versions under a fixed random seed (42). The complete reproducibility pipeline (scripts 01–09 and the mode-imputation sensitivity check) is provided in the code repository (see Code availability in the main text).

Supplementary Table S7. Model sample sizes by exposure (Black women, 2023)

Exposure	Restricted sample	Analytic sample	Excluded	Exposed	Events
PGDM	491,753	475,676	16,077	8,047	77,972
PGHTN	491,753	475,676	16,077	27,471	77,972
Gestational hypertension[a]	462,540	449,657	12,883	56,172	70,335
ART	491,753	490,362	1,391	4,027	81,484
GDM[b]	482,589	469,159	13,430	32,005	75,247
Age ≥35	491,753	491,157	596	96,244	81,676

[a]Restricted to women without PGHTN; [b]restricted to women without PGDM (NCHS mutual-exclusivity coding). The restricted sample is the race-restricted Black 2023 cohort after any disease-restriction; the analytic sample is the complete-case model sample (non-missing exposure, outcome and adjustment covariates); excluded is the difference. Exposed and event counts are within the analytic sample; events are composite-endpoint (SMM or CRITICAL) cases. ART additionally drops the infertility-treatment marker from adjustment. Adjustment sets are the minimal DAG-derived back-door sets listed in Supplementary Table S2: PGDM, age ≥35, obesity, PGHTN and smoking; PGHTN, age ≥35, obesity, PGDM and smoking; gestational hypertension and GDM, obesity; ART, age ≥35; age ≥35, none.

Supplementary Table S8. Outcome rates restricted to singleton births (Black vs White, 2023)

Outcome (singletons only)	NH Black	NH White
SMM (5-item proxy)	1.39%	1.53%
Severe SMM (4-item proxy)	0.95%	0.66%
CRITICAL (neonatal)	14.17%	11.32%
Composite morbidity	15.10%	12.53%
Composite severe morbidity (4-item)	14.72%	11.75%

Because a multiple-gestation delivery contributes one record per infant, a single maternal morbidity event can appear on more than one record; the singleton-restricted rates above show the Black–White disparity is not an artefact of multiple gestation.

Supplementary Table S9. Singleton-restricted adjusted risk differences (Black women, 2023)

Exposure	Outcome	Risk difference (pp)	95% CI
PGHTN	Severe SMM (4-item)	0.71	0.56–0.98
PGDM	Severe SMM (4-item)	0.35	0.22–0.72
ART	Severe SMM (4-item)	0.16	0.10–1.81
PGDM	Composite morbidity	16.27	15.31–17.27
PGHTN	Composite morbidity	9.30	8.82–9.77

Restricting to singleton births removes the one-record-per-infant duplication of a maternal morbidity event. Severe-SMM attenuation was modest for PGHTN (0.94 to 0.71 pp) and PGDM (0.53 to 0.35 pp) but marked for ART (1.80 to 0.16 pp, 91%), consistent with multiple gestation being a major pathway for ART-related maternal morbidity rather than a general driver of the metabolic-outcome findings.

Supplementary Table S10. DML heterogeneity and interaction tests within Black women

Subgroup	PGDM interaction (pp)	PGDM raw p	PGDM BH q	PGHTN interaction (pp)	PGHTN raw p	PGHTN BH q
Education (below HS)	−2.6	0.35	0.52	−0.7	0.63	0.76
Education (college+)	−2.3	0.30	0.52	−0.8	0.46	0.62
WIC	+0.4	0.82	0.90	−1.0	0.27	0.52
Age ≥35	+0.1	0.96	0.96	+2.5	**0.009**	**0.036**
Obesity	−6.1	**0.002**	**0.017**	−1.6	0.093	0.28
Foreign-born	−6.3	**0.004**	**0.023**	+1.7	0.21	0.49
DML ATE		**18.9 pp**			**10.7 pp**	

Interaction terms from linear-probability models with robust standard errors. Both raw p-values and Benjamini–Hochberg FDR-adjusted q-values across the 12 tests are shown; q-values were computed from the unrounded p-values before any rounding for display, so recomputation from the table is not expected to reproduce the three-decimal q-values exactly (the displayed q=0.017, 0.023 and 0.036 correspond to unrounded interaction p-values of 0.00145, 0.00378 and 0.00890, respectively). Significant modifiers (bold) are obesity and foreign-born nativity for PGDM and age ≥35 for PGHTN; all are modest in magnitude. DML ATE broadly consistent with DAG ACE (16.3 and 10.1 pp). ART was not estimable in the DML heterogeneity analysis within Black women (0.82% prevalence).

Supplementary Table S11. Alternative-DAG sensitivity: relaxing PGDM–PGHTN conditional independence

Exposure	Model	Risk difference (pp)	95% CI
PGDM	Main DAG	16.25	15.17–17.33
PGDM	+ PGDM → PGHTN edge	19.50	18.44–20.64
PGHTN	Main DAG	10.08	9.53–10.60
PGHTN	+ PGHTN → PGDM edge	11.51	10.98–12.00

The primary DAG treats pre-gestational diabetes and pre-gestational hypertension as conditionally independent given their shared preconception causes. Adding a direct association between the two (in either direction)

moves each estimate by at most 3.25 percentage points, indicating the conditional-independence specification is not the sole driver of the findings.

Supplementary Table S12. Complete-case versus mode-imputed adjusted risk differences (Black women, 2023)

Exposure	Complete-case	Mode-imputed	Difference
PGDM	16.25	16.47	+0.22
PGHTN	10.08	10.47	+0.39
ART	12.53	12.55	+0.02

Risk differences are in percentage points (pp). All estimates are DAG-adjusted average causal effects on the composite (SMM or CRITICAL) morbidity endpoint. Mode imputation replaces each missing value in the exposure, outcome and adjustment-set covariates with the column mode before g-computation; the complete-case comparator is the primary estimate from Table 3 of the main text. Differences are all < 0.5 percentage points, indicating that the low missingness does not drive the primary estimates. Point estimates only (no bootstrap).

Supplementary Table S13. Constrained-network temporal-topology comparison (Black women)

Metric	Value
2022 learned edges	82
2023 learned edges	84
Structural Hamming distance (SHD)	32
Jaccard similarity	0.677

Networks were learned independently per year using hill-climbing with Bayesian information criterion scoring under a temporal-ordering blacklist (80,000-record subsamples). The comparison is a data-informed robustness check on the *a priori* DAG, not independent validation of the causal estimates.