

Investigating Health Disparities and Key Factors in Medical Outcomes

Group 11

Nguyen Canh Huy - Can Ha An - Dao Chi Tuong



Table of contents

01

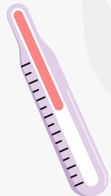
Dataset: Academic Literature on Racial and Ethnic Disparities in Reproductive Medicine in the US

02

Question 1: How are race and ethnicity included and categorized in racial/ethnic-related medical research?

03

Question 2: What health outcomes have been studied, and what disparities have been identified?





01

Dataset: Academic Literature on Racial and Ethnic Disparities in Reproductive Medicine in the US



The dataset

Racial and ethnic disparities in reproductive medicine in the United States: a narrative review of contemporary high-quality evidence

Ayodele G Lewis ¹, Divya K Shah ², Regina Leonis ³, John Rees ⁴, Katharine F B Correia ⁵

Affiliations + expand

PMID: 39059596 DOI: [10.1016/j.ajog.2024.07.024](https://doi.org/10.1016/j.ajog.2024.07.024)

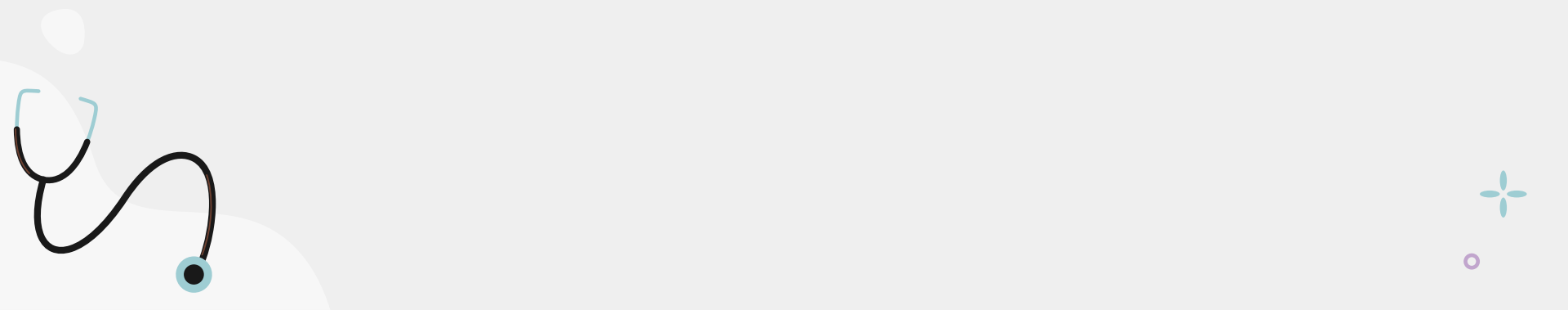
Free article





The dataset

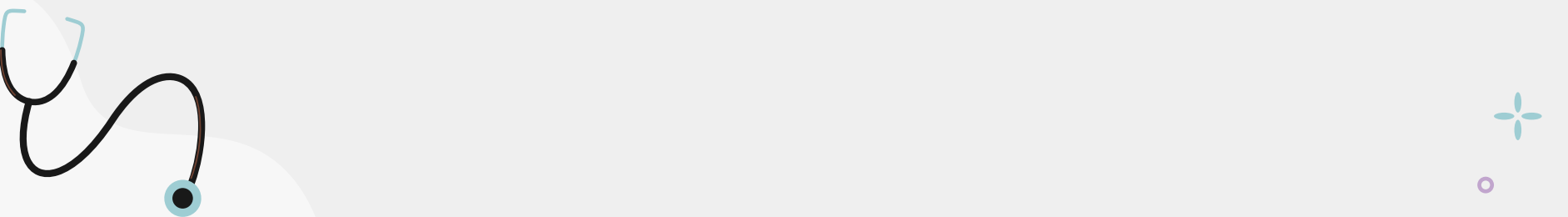
Article Data (`article_dat.csv`)

- Provenance: Metadata on medical research articles.
 - Dimensions: 318 rows × 65 columns.
 - Key Variables:
 - `doi`, `pmid`: Study identifiers.
 - `journal`, `year`, `title`, `abstract`: Study metadata.
 - `study_location`: Geographic study area.
 - `race1`, `eth1`: Reported primary racial and ethnic categories.
- 



The dataset

Model Data (`model_dat.csv`)

- Provenance: Extracted from medical research studies.
 - Dimensions: 6,804 rows × 16 columns.
 - Key Variables:
 - `doi`: Unique study identifier.
 - `model_number`: Statistical model reference.
 - `outcome`: Outcome measure for the model as reported in the article.
 - `measure`: Estimated measure from the model: incidence, prevalence, odds ratio (OR), risk ratio (RR), hazards ratio (HR), absolute difference, percent, other.
 - `point`, `lower`, `upper`: Estimate and confidence intervals.
 - `stratgrp`, `subgrp`: Stratification variables (e.g., demographics).
- 

Preprocessing

- Analyze missing data using missingno matrix => imputation
- Created custom race_standardization mapping + fuzzywuzzy to handle compound races

```
# Lowercased standardized mapping
race_standardization = {
    'african american': 'Black or African American',
    'african american or black': 'Black or African American',
    'african american, non-hispanic': 'Black or African American',
    'black': 'Black or African American',
    'black or african american': 'Black or African American',
    'black, non-hispanic': 'Black or African American',
    'non african american': 'Black or African American',
    'non-african american': 'Black or African American',
    'non-hispanic black': 'Black or African American',

    'caucasian': 'White',
    'white': 'White',
    'non hispanic white': 'White',
    'non-hispanic white': 'White',
    'non-latina white': 'White',
    'white (non-hispanic)': 'White',
    'white non-hispanic': 'White',
    'white/caucasian': 'White',
    'whites': 'White',
    'u.s.-born, non-hispanic white;': 'White',
    'non-hispanic, white': 'White',
    'white donor / white recipient': 'White',
    'european american': 'White',
    'non hispanic whites': 'White',
    'white, non-hispanic': 'White',

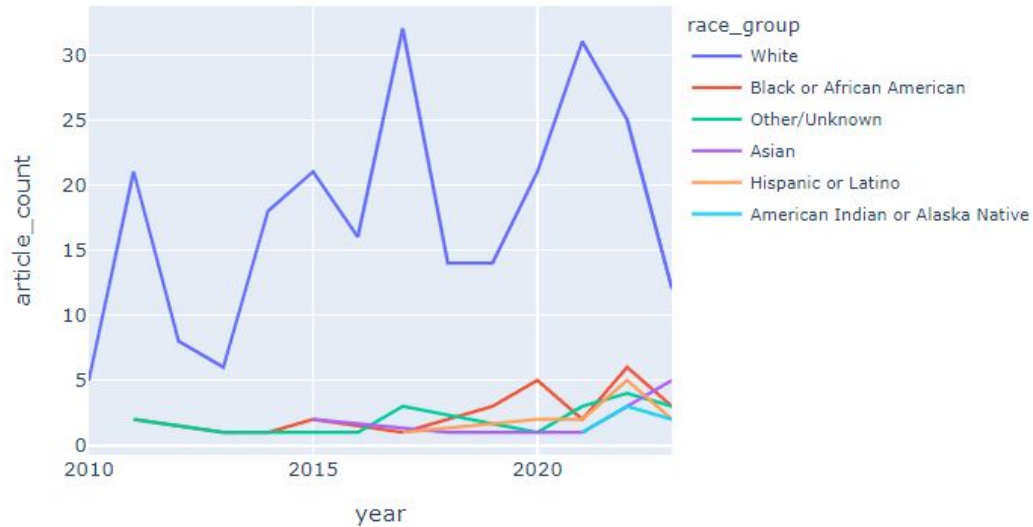
    'hispanic': 'Hispanic or Latino',
    'latina': 'Hispanic or Latino'.
```



02

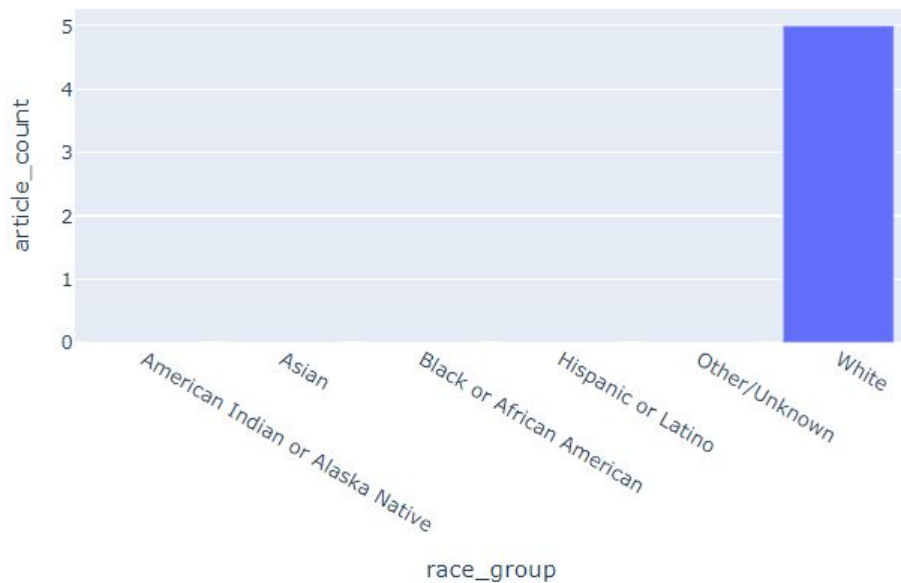
Question 1: How are race and ethnicity included and categorized in racial/ethnic-related medical research?

Articles Over Time by Race Group



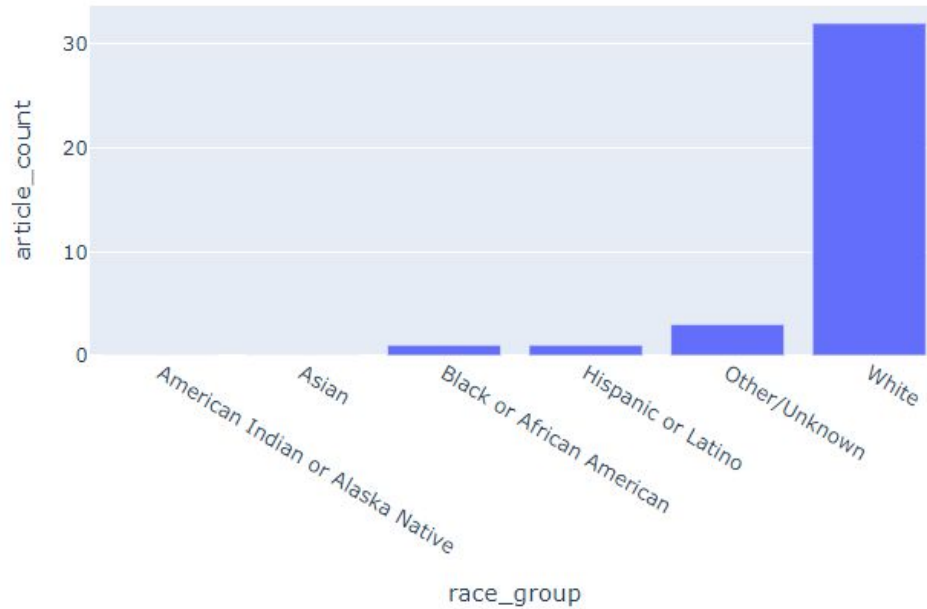
The majority of published articles focus on White populations, with consistent underrepresentation of other racial groups over time. A spike in recent years may reflect increased attention to disparities, but gaps remain.

Number of Articles by Race Group (2010)



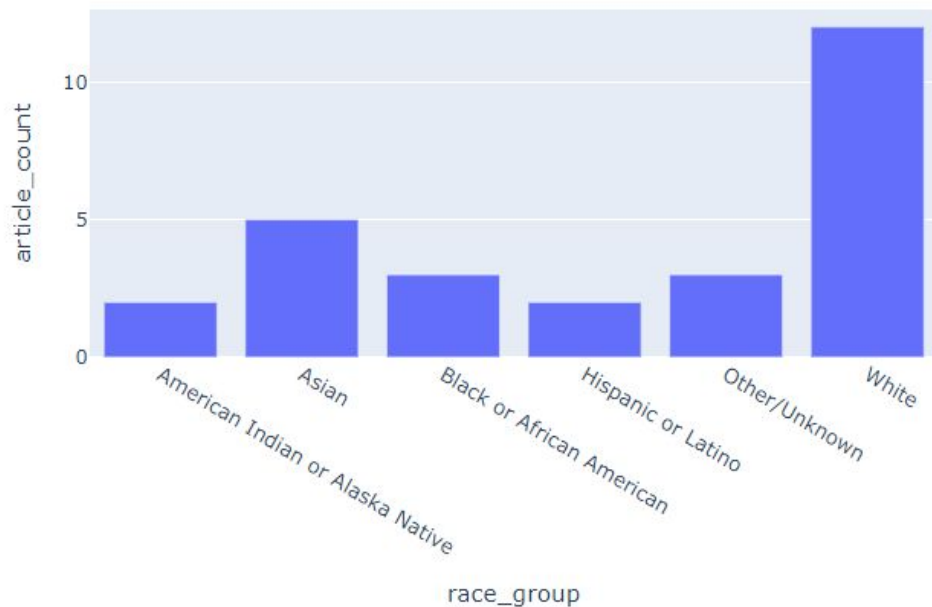
The majority of published articles focus on White populations, with consistent underrepresentation of other racial groups over time. A spike in recent years may reflect increased attention to disparities, but gaps remain.

Number of Articles by Race Group (2017)

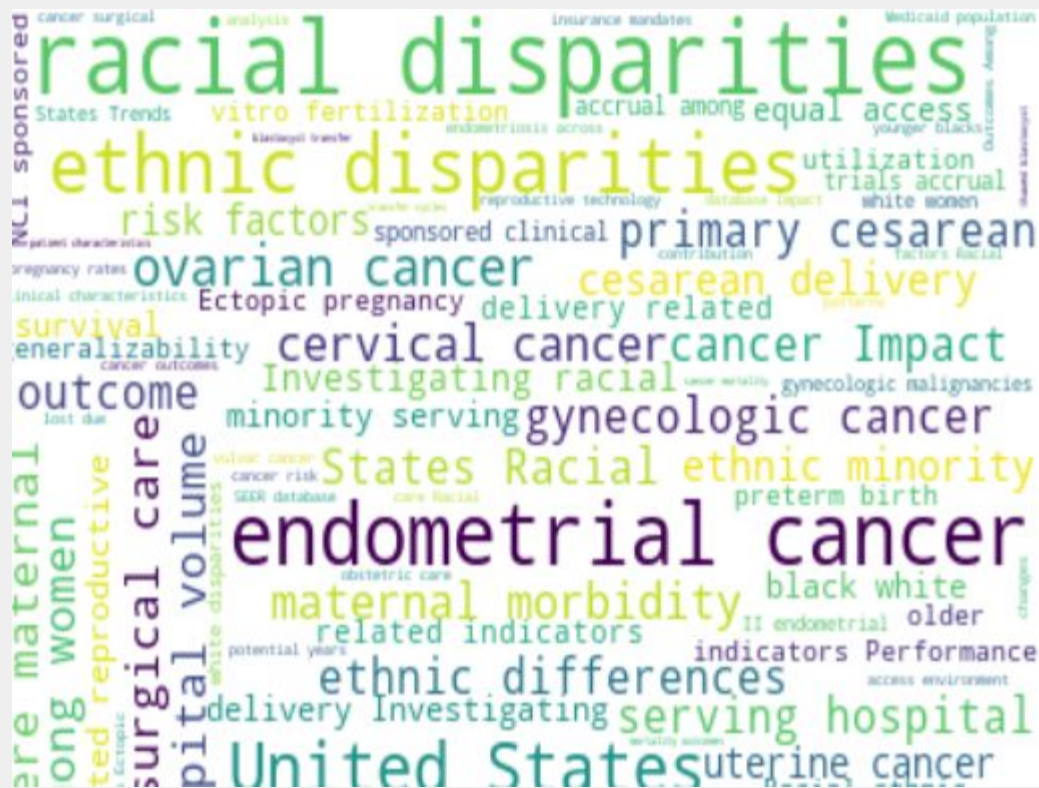


The majority of published articles focus on White populations, with consistent underrepresentation of other racial groups over time. A spike in recent years may reflect increased attention to disparities, but gaps remain.

Number of Articles by Race Group (2023)



The majority of published articles focus on White populations, with consistent underrepresentation of other racial groups over time. A spike in recent years may reflect increased attention to disparities, but gaps remain.



03

Question 2: What health outcomes have been studied, and what disparities have been identified?





Health Outcome Categories by Race

Select Racial Group:

All

Prevalence of Health Outcome Categories

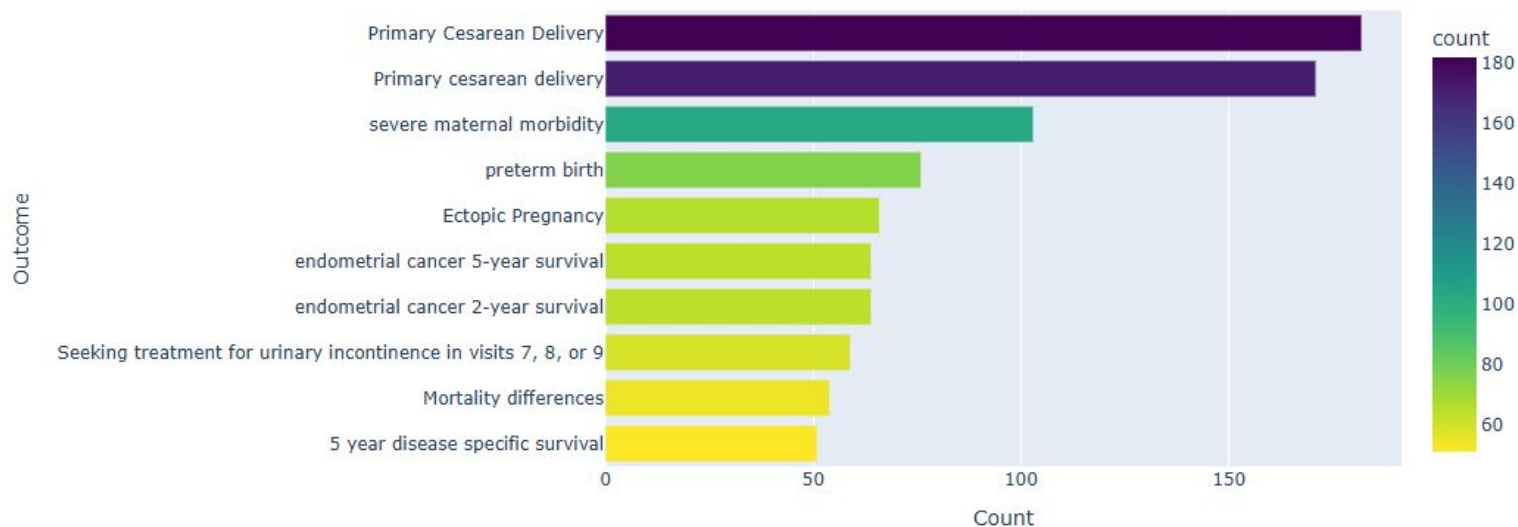


Maternal and cancer-related health outcomes dominate the literature across racial groups. Neonatal and mental health categories are less studied, indicating potential areas for further research.



Top 10 Health Outcomes in Selected Category

Top 10 Health Outcomes



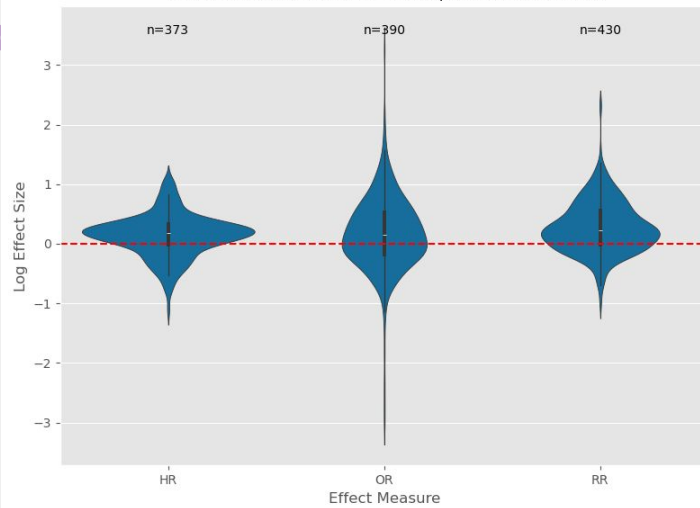
Across all racial groups, cesarean delivery and maternal morbidity are the most frequently studied outcomes. Many of these outcomes reflect critical gaps in reproductive healthcare.

Heatmap of Health Category Prevalence by Race

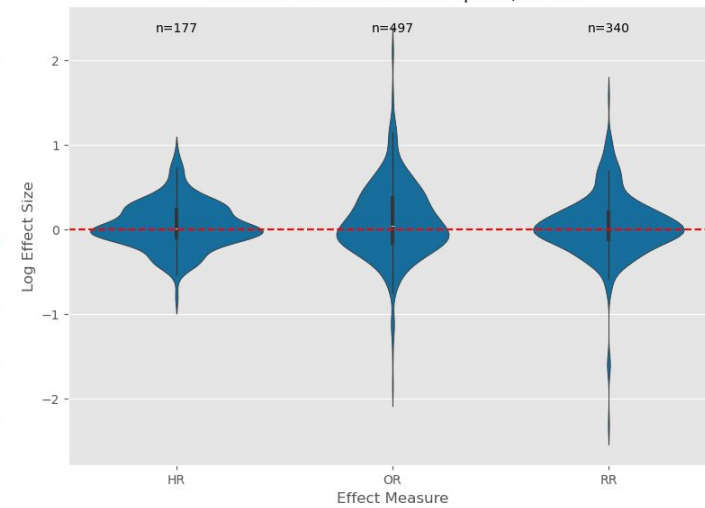


White populations are overrepresented in nearly every health category. The disparity is especially large in maternal and cancer-related research.

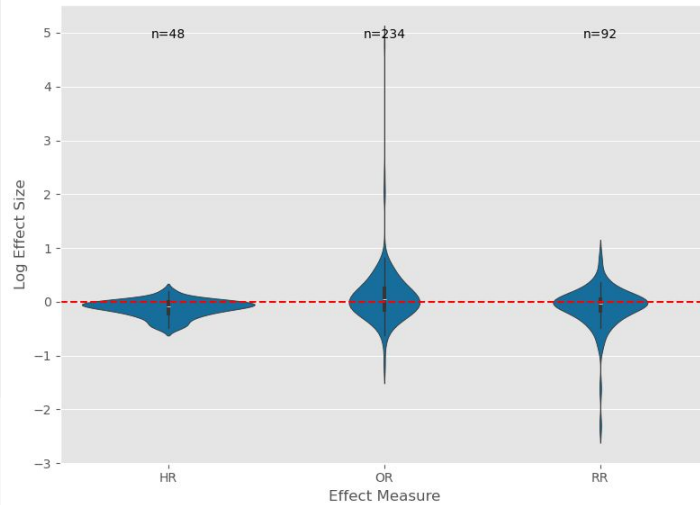
Effect Size Distribution - Black/African American



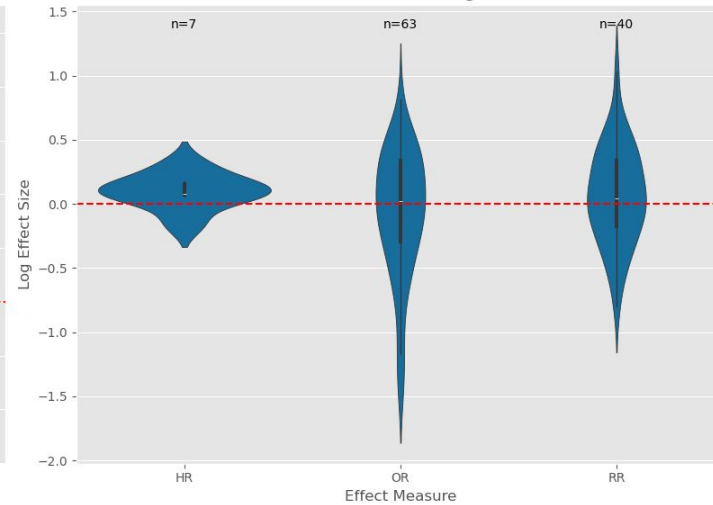
Effect Size Distribution - Hispanic/Latino



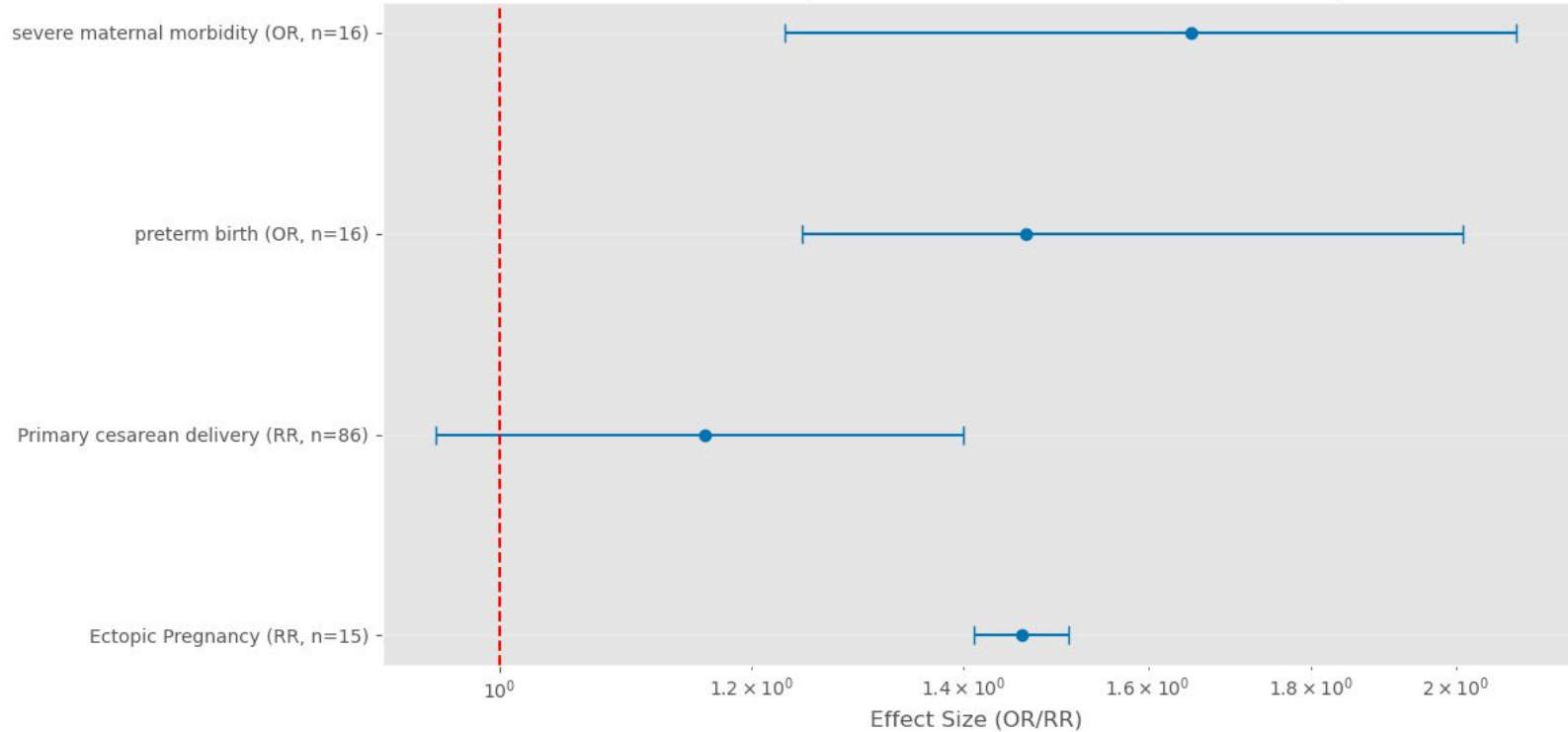
Effect Size Distribution - Asian/Pacific Islander



Effect Size Distribution - Indigenous/Native



Forest Plot: Health Disparities for Black or African American Population



For Black/African American populations, the strongest disparities were observed in severe maternal morbidity and cesarean delivery. These findings reinforce the urgency of addressing inequities in reproductive healthcare outcomes.

Related works

- [1] Lewis, Ayodele G. et al (2025), Racial and ethnic disparities in reproductive medicine in the United States: a narrative review of contemporary high-quality evidence American Journal of Obstetrics & Gynecology, Volume 232, Issue 1, 82 - 91.e44
- [2] <https://github.com/rfordatascience/tidytuesday/blob/main/data/2025/2025-02-25/readme.md>



Thanks

Do you have any questions?



CREDITS: This presentation template was created by [Slidesgo](#), including icons by [Flaticon](#) and infographics & images by [Freepik](#)

Please keep this slide for attribution

