Analyzing Vaccine Study Bias

https://gemini.google.com/app/2a2a26a8922af53b

User prompt: can you predict whether the OR should be <1 or >1 or 1?

Response: Based on the study design provided, I can confidently predict that under a null assumption, the Odds Ratio (OR) will be substantially **< 1**.

This study design relies on a case-only cohort of decedents. By restricting the analysis entirely to children who died before age 3 and lacking a healthy, living control group, the design introduces massive structural confounding—specifically, the **Healthy Vaccinee Effect (HVE)**.

Here is the detailed breakdown and prediction for the analysis.

Part A: Predictions Assuming Zero Causal Effect

1. Predicted Odds Ratio for Individual Vaccines

- **Prediction:** The OR will be substantially **< 1** (likely between **0.2 and 0.6**).
- **Reasoning:** Infants who receive their 2-month vaccines on schedule (days 60–90) are generally healthy enough to attend a routine well-child visit. Conversely, the "unvaccinated" group will be highly enriched with infants who are severely ill, hospitalized in the NICU, or suffering from severe congenital anomalies—conditions that both delay vaccination and drastically increase the risk of early death (days 90–120). Therefore, the vaccinated group will artificially appear protected against early death relative to the highly vulnerable unvaccinated group.

2. Variation Across Vaccine Types

- **Prediction:** There will be **minimal meaningful variation** across the core 2-month vaccines (DTaP, HIB, polio, pneumococcal, rotavirus).
- **Reasoning:** In the US, these vaccines are almost exclusively administered together at the same 2-month visit. Any infant receiving one is highly likely receiving the others.
- **Highest OR:** Hepatitis B might show the highest OR (closest to 1) because its scheduling is slightly more flexible (a birth dose is standard), meaning its administration or delay might be less strictly tied to the infant's health status exactly at day 60–90.

3. Gradient Across 1 → 5 → 6 Vaccines

- **Prediction:** The OR will decrease (move further < 1) as the number of vaccines increases.
- **Reasoning:** Receiving all 6 vaccines perfectly on schedule is a very strong proxy for an infant being in optimal health with good pediatric care. Receiving only 1 vaccine could indicate an off-schedule catch-up or a specific medical contraindication to the others. Thus, the HVE will be strongest in the "all 6" group.

4. Sex Differences in the ORs

- **Prediction:** The OR for males will likely be slightly higher (closer to 1) than for females.

- **Reasoning:** SIDS has a 60:40 male-to-female predominance and peaks at 1-4 months. In the *vaccinated* arm (the generally healthy infants), early deaths (days 90-120) will be heavily driven by SIDS. Because males die of SIDS at higher rates, the male vaccinated arm will have a proportionally higher early-death rate than the female vaccinated arm, pushing the male OR slightly closer to the null.

5. Race Differences in the ORs

- **Prediction:** The OR for Black infants will likely be lower (further < 1) than for White infants.
- **Reasoning:** The baseline infant mortality for Black infants is roughly double the national average (≈ 10.5 per 1,000), largely driven by higher rates of prematurity and complications. Higher prematurity leads to a larger proportion of severely ill Black infants in the unvaccinated arm, amplifying the Healthy Vaccinee Effect and skewing the OR further below 1.

6. Fraction of Contrasts Pointing in the Same Direction

- **Prediction: ~100%** of the ~55 contrasts will point in the same direction (OR < 1).
- **Reasoning:** Because the unvaccinated reference group (n = 343) is the exact same across all individual contrasts, the structural HVE bias is baked into every single calculation.

7. Identified Biases

- **Healthy Vaccinee Effect / Confounding by Indication:** Sick infants do not get vaccinated on schedule but die early. (Sign: Toward OR < 1. Magnitude: Very Large).
- **Index Event Bias (Collider Bias):** Conditioning the entire dataset only on children who died before age 3. By forcing the total deaths in each group to be 100% over the 3-year window, anything that alters the *timing* of death acts as a direct confounder. (Sign: Toward OR < 1. Magnitude: Large).
- **Immortal Time Bias (misaligned windows):** To die between 90-120 days, a child must survive to day 90. Very sick unvaccinated children might die before day 90 and be excluded from the numerator, altering the denominator pools. (Sign: Unpredictable, likely toward OR < 1. Magnitude: Moderate).

Part B: Predictions Assuming Vaccines DO Cause Increased Early Mortality

8. How Answers 1–6 Would Change

- **1 (OR):** The massive structural HVE (pushing OR < 1) would compete against the true causal effect (pushing OR > 1). The observed OR would shift upward (closer to 1, or potentially > 1 if the vaccine toxicity is exceptionally high), but because the HVE is so strong in case-only designs, it might still mask the hazard, resulting in an OR that is simply *less protective* (e.g., 0.8 instead of 0.3).
- **2 (Variation):** If a specific vaccine component (e.g., DTaP) is the causal driver, its specific contrast would spike higher than the others.
- **3 (Gradient):** The 1 → 5 → 6 gradient would likely **reverse**. If the vaccines cause mortality, a higher cumulative dose (6 vaccines) would result in higher mortality (pushing the OR higher) compared to just 1 vaccine.

- **4 & 5 (Sex/Race):** Changes here would depend entirely on whether the hypothetical fatal physiological response to the vaccine disproportionately affects a specific sex or race.
- **6 (Consistency):** If only one vaccine is the culprit, contrasts including that vaccine would shift, while others might remain heavily < 1, reducing the uniformity from 100% to a mixed distribution.

Part C: Discriminating Between Hypotheses

9. Which Observations Discriminate?

- **Discriminates well:** The **1 → 5 → 6 gradient**. Under the null (HVE), more vaccines equate to a healthier baby (lower OR). Under the causal hypothesis, more vaccines equate to higher toxicity (higher OR). A reversal of this gradient is a strong signal.
- **Does not discriminate well:** The **raw absolute OR**. An OR of 0.8 could mean vaccines are slightly protective, or it could mean a severe causal mortality risk (OR > 1) is being suppressed by a massive Healthy Vaccinee Effect (OR = 0.2). You cannot disentangle them from the raw point estimate.

10. Most Informative Additional Table

- A cross-tabulation of **Cause of Death (ICD-10 codes) by Exposure Group**.
- **Why:** If the "unvaccinated" early deaths are overwhelmingly coded for extreme prematurity, congenital heart defects, and chromosomal abnormalities, it proves the HVE is driving the effect. If the "vaccinated" early deaths show a distinct, anomalous spike in specific codes (e.g., anaphylaxis, myocarditis) immediately following the 60-90 day window, it provides mechanistic evidence for a causal effect.
