# Supplementing risk ratios in sibling analysis: Estimating clinically useful measures from family-based analyses

*Viktor H. Ahlqvist, **Hugo Sjöqvist**, Arvid Sjölander, Daniel Berglind, Paul C. Lambert, Brian K. Lee, Paul Madley-Dowd*

# About me

PhD student at the Department of Global Public Health (GPH), at Karolinska Institutet.

- → ” Early life infections and neurodevelopmental conditions: a focus on familial confounding”
  - Examining the associations between early-life infections and neurodevelopmental conditions.

Part of the Biostatistics Team at the Department of Global Public Health.

# Example: Association between maternal infection and child autism or intellectual disability (ID)

Articles

## Maternal infection during pregnancy and likelihood of autism and intellectual disability in children in Sweden: a negative control and sibling comparison cohort study

*Martin Brynge*, Hugo Sjöqvist*, Renee M Gardner, Brian K Lee, Christina Dalman, Håkan Karlsson*

*Lancet Psychiatry* 2022; 9: 782–91

Published Online September 7, 2022 http://.doi.org/10.1016/S2215-0366(22)00264-4

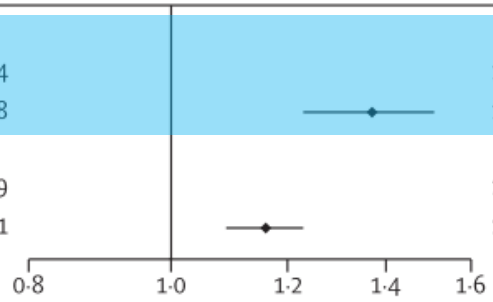
*Contributed equally

### Summary

**Background** Maternal infections during pregnancy are associated with intellectual disability and autism in exposed children. Whether these associations are causal, and therefore should be targets of preventive strategies, remains unknown. We aimed to investigate these associations, to determine whether there is a causal role of maternal infection during pregnancy for children's risk of autism and intellectual disability, by accounting for unmeasured familial factors.

# Example: Association between maternal infection and child autism or intellectual disability (ID)

| | Cases | Person-years | HR (95% CI) | | |
|---|---|---|---|---|---|
| | | | Model 1 | Model 2 | Model 3 |
| **Any intellectual disability diagnosis** | | | | | |
| No pregnancy infection | 5087 | 7022404 | 1 (ref) | 1 (ref) | 1 (ref) |
| Pregnancy infection | 445 | 384478 | 1·63 (1·48–1·81) | 1·37 (1·24–1·52) | 1·37 (1·23–1·51) |
| **Any autism diagnosis** | | | | | |
| No pregnancy infection | 13034 | 6994459 | 1 (ref) | 1 (ref) | 1 (ref) |
| Pregnancy infection | 1123 | 382251 | 1·31 (1·23–1·40) | 1·17 (1·10–1·24) | 1·16 (1·09–1·23) |




Model 1 was adjusted for child's sex and birth year.
Model 2 was adjusted for child's sex, birth year, birth order, and season of birth; maternal age and BMI; parental psychiatric history, income at birth, level of education at birth, and regions of origin; and paternal infection during pregnancy.
Model 3 was adjusted as model 2, but also adjusted for maternal pre-pregnancy infection.

# Controlling for the unobservable

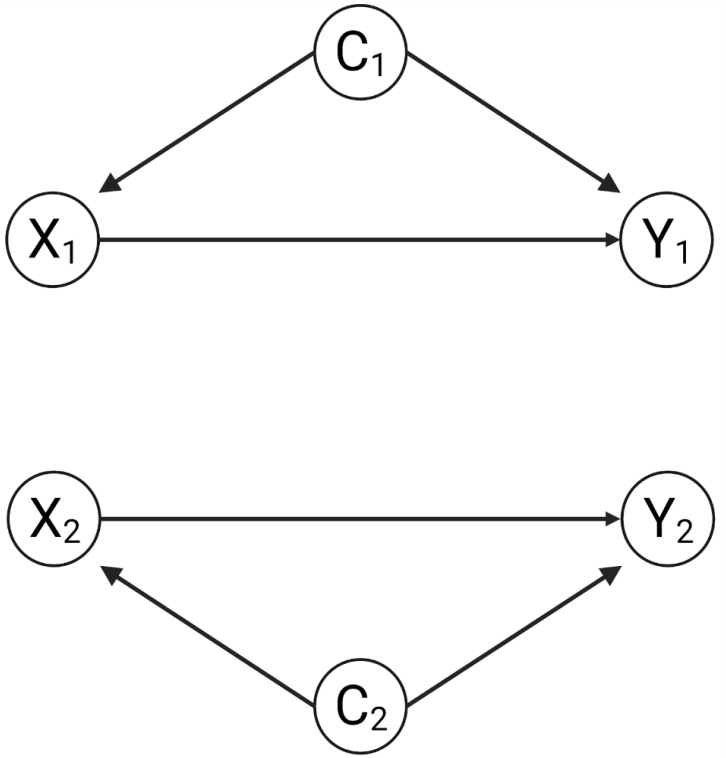
Families' structures offer a unique opportunity.

- → We can typically assume that they share a large (or some) part of the unobserved environmental confounders.
- → Family members also share genes, and comparing family members may control for genetic confounding.

**The principal idea:** family members are more similar in unobserved (inc. unobservable and unknown) characteristics than two people in the general population, they, therefore, are more “exchangeable” than two people out of the general population.

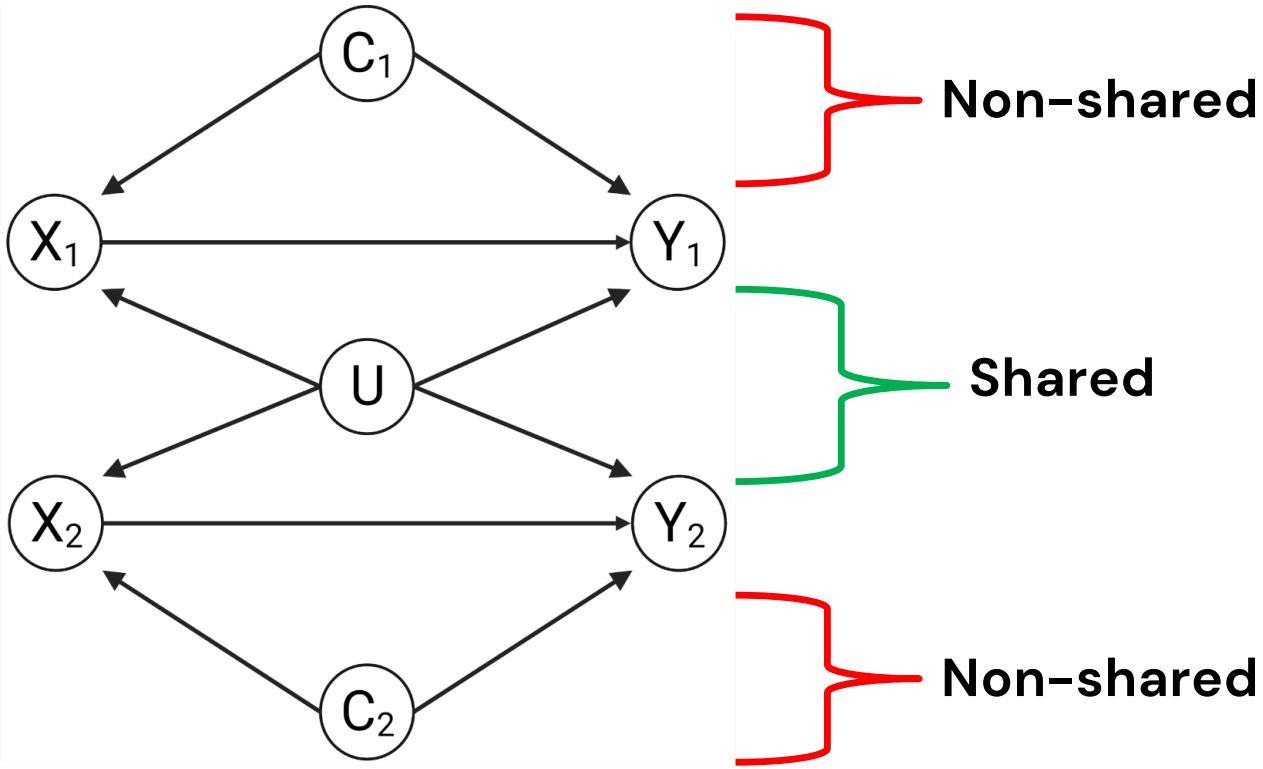
- → This principal idea is the strength of family-designs.

# Directed acyclic graphs – per family member




**X = Exposure**
**C = Cofounder**
**Y = Outcome**
**For individual 1 and 2**

# Directed acyclic graphs – 'family-set'




**X = Exposure**
**C = Cofounder**
**Y = Outcome**
**For individual 1 and 2**
**U = Shared factors**

# For binary outcomes

The conditional logistic regression (i$^{th}$ individual, j$^{th}$ family):

$$P\left(Y_{ij}\middle|X_{ij}\right) = \frac{exp\left(\alpha_j + \beta_1 X_{ij}\right)}{1 + exp\left(\alpha_j + \beta_1 X_{ij}\right)}$$

Estimated using conditional likelihood, so alpha $\alpha_j$ cancels out.

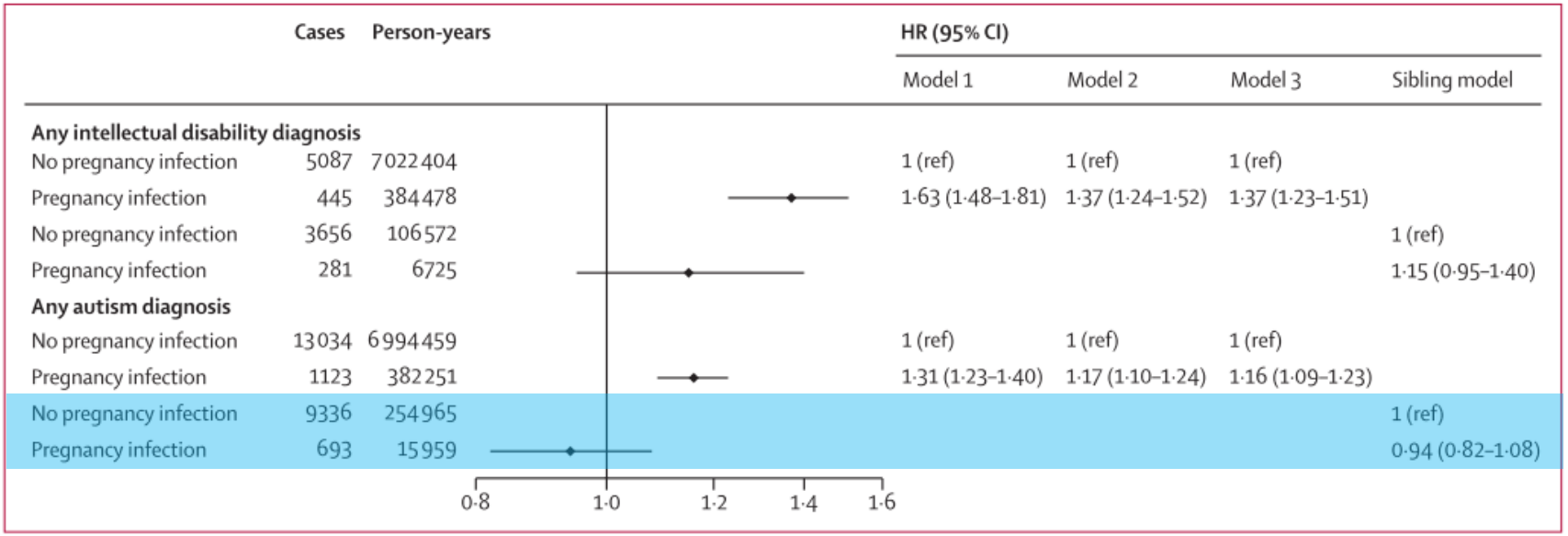
The conditional between-within model:

$$P\left(Y_{ij}\middle|X_{ij}, \bar{X}_j\right) = \frac{exp\left(\alpha_j + \beta_1 X_{ij} + \theta_1 \bar{X}_j\right)}{1 + exp\left(\alpha_j + \beta_1 X_{ij} + \theta_1 \bar{X}_j\right)}$$

Where $\bar{X}_j$ represents the j$^{th}$ family's average of the exposure, and $\alpha_j = \alpha_j^{\dagger} + \gamma_1 \bar{X}_j$ the cluster-specific intercept

Under both models, $\beta_1$ measures the conditional association of the outcome with a 1 unit increase in X (aka. the within-effect).

# Example: Association between maternal infection and child autism or intellectual disability (ID)

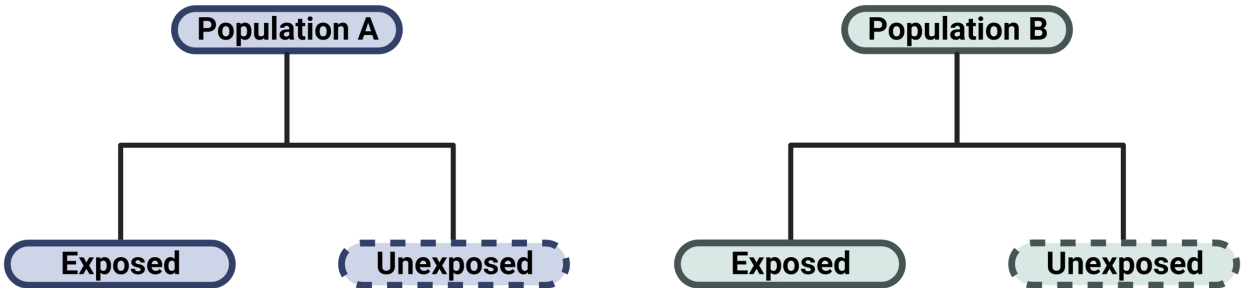


| | Cases | Person-years | HR (95% CI) Model 1 | Model 2 | Model 3 | Sibling model |
|---|---|---|---|---|---|---|
| **Any intellectual disability diagnosis** | | | | | | |
| No pregnancy infection | 5087 | 7022404 | 1 (ref) | 1 (ref) | 1 (ref) | |
| Pregnancy infection | 445 | 384478 | 1·63 (1·48–1·81) | 1·37 (1·24–1·52) | 1·37 (1·23–1·51) | |
| No pregnancy infection | 3656 | 106572 | | | | 1 (ref) |
| Pregnancy infection | 281 | 6725 | | | | 1·15 (0·95–1·40) |
| **Any autism diagnosis** | | | | | | |
| No pregnancy infection | 13034 | 6994459 | 1 (ref) | 1 (ref) | 1 (ref) | |
| Pregnancy infection | 1123 | 382251 | 1·31 (1·23–1·40) | 1·17 (1·10–1·24) | 1·16 (1·09–1·23) | |
| No pregnancy infection | 9336 | 254965 | | | | 1 (ref) |
| Pregnancy infection | 693 | 15959 | | | | 0·94 (0·82–1·08) |

0·8 1·0 1·2 1·4 1·6

The sibling model is not “statistically significant” for ID, but its point estimate still show 15% increased risk.

# The uninformative sole odds ratio/hazard ratio



# Solutions for the odds/hazard ratios: Absolute measurements

Absolute measures of occurrence (risk, risk difference, survival curves, numbers needed to treat, etc.), together with the relative measures.

For absolute measures we need a stable baseline (e.g., a baseline risk). However:

1. The baseline cancels out under conditional logistic regression and stratified Cox regression.
2. The conditional between-within models use cluster-specific intercepts, or a fixed distribution – computationally horrible/impossible.

This means that absolute measures can not be estimated from traditional family-based analysis.

# The Marginal Between-Within model

*Sjölander. *Estimation of marginal causal effects in the presence of confounding by cluster*. Biostatistics. 2019.

Recall, that the conditional between-within model:

$$P\left(Y_{ij}\middle|X_{ij},\bar{X}_j\right)=\frac{exp\left(\boldsymbol{\alpha_j}+\beta_1X_{ij}+\theta_1\bar{X}_j\right)}{1+exp\left(\boldsymbol{\alpha_j}+\beta_1X_{ij}+\theta_1\bar{X}_j\right)}$$

The marginal between-within model*:

$$P\left(Y_{ij}\middle|X_{ij},\bar{X}_j\right)=\frac{exp\left(\boldsymbol{\alpha}+\beta_1X_{ij}+\theta_1\bar{X}_j\right)}{1+exp\left(\boldsymbol{\alpha}+\beta_1X_{ij}+\theta_1\bar{X}_j\right)}$$

With a fixed intercept **α** that is common to all clusters.

With this, we are able to get around previous limitations.

# Estimating clinically useful measurements for infections during pregnancy and risk of ID

```
local variables exposure co_exposure male cnum2 cnum3 seas2 seas3 seas4 zbyear
foreach var of local variables {
    egen between_`var' = mean(`var'), by(fam_id)
}

stcox i.exposure co_exposure male cnum2 cnum3 seas2 seas3 seas4 zbyear ///
    between_* ///
    if sibling==1,  basesurv(S0)
```

# Estimating clinically useful measurements for infections during pregnancy and risk of ID

# Maternal smoking and infant mortality
# Differences between conditional and BW model

| | Binary analysis | | Survival analysis | |
|---|---|---|---|---|
| | Conditional logistic | Marginal between-within logistic | Stratified Cox | Marginal between-within Cox |
| **Conditional odds ratio, (95% CI)** | 1.63 (1.42-1.87) | 1.68 (1.46-1.94) | n/a | n/a |
| **Conditional hazard ratio, (95% CI)** | n/a | n/a | 1.60 (1.40-1.83) | 1.68 (1.46-1.93) |
| **Risk among unexposed, % (95% CI)** | - | 0.31 (0.30-0.32) | - | 0.31 (0.30-0.32) |
| **Risk among exposed, % (95% CI)** | - | 0.52 (0.46-0.58) | - | 0.52 (0.46-0.58) |
| **Risk difference, % (95% CI)** | - | 0.21 (0.14-0.28) | - | 0.21 (0.14-0.28) |
| **Attributable fraction, % (95% CI)** | - | 8.62 (6.68-10.56) | - | 8.62 (6.68-10.56) |
| **Number Needed to Harm, N (95% CI)** | - | 476 (320-631) | - | 476 (320-632) |

[a]All absolute measures were calculated up until the 365$^{th}$ day of life.

# Pre-print available

Preprint
Made publicly accessible prior to peer review

NIH National Institutes of Health
Turning Discovery Into Health

▸ medRxiv [Preprint]. 2025 May 16:2025.05.16.25327702. [Version 1] doi: 10.1101/2025.05.16.25327702

## Moving beyond risk ratios in sibling analysis: estimating clinically useful measures from family-based analysis

Viktor H Ahlqvist [1,2], Hugo Sjöqvist [3], Arvid Sjölander [4], Daniel Berglind [3], Paul C Lambert [4,5], Brian K Lee [3,6,7], Paul Madley-Dowd [8,9]

▸ Author information ▸ Copyright and License information

PMCID: PMC12132141 PMID: 40463571


Karolinska
Institutet