

**Paracetamol, paracetamol propensity and 7 year asthma/wheezing in ALSPAC**

Roger Newson (King's College, London, UK)

Seif Shaheen (King's College, London, UK)

and the ALSPAC team (Bristol University, UK)

- Problems with multiple confounders.
- The method of propensity scores
- Paracetamol exposures and propensity scores in ALSPAC.
- Advantages and limitations of propensity scores.

**Prenatal paracetamol and asthma/wheezing at 7 years in the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort (14060 children).**

- **2 primary outcomes:** Questionnaire-reported asthma and wheezing at 69-81 months. (Yes or no.)
- **1 secondary outcome:** Questionnaire-reported asthma frequency at 69-81 months. (Never, 1-2 times or 3+ times.)
- **4 primary exposures:** Maternal paracetamol and aspirin consumption at 18-21 weeks and 32 weeks gestation. ("Never", "Some days" or "Most days".)
- **14 "primary" (prenatal or perinatal) confounders:** Child's sex, mother's age, smoking in pregnancy, mother's education level, mother's housing tenure, mother's parity, mother's anxiety score, mother's ethnic origin, multiple pregnancy, child's birthweight, gestational age at birth, child's head circumference, maternal disease history, maternal infection history.
- **7 "secondary" (postnatal) confounders:** Younger siblings at 7 years, pets in first year, breast fed in first 6 months, day care in first 8 months, damp and/or mould in home, weekend tobacco exposure of child, child's body mass at 7 years.
- **2 "tertiary" confounders (subject to cautions about reverse causation):** Paracetamol use and antibiotic use during infancy.

## **General principles**

In general, a statistical analysis should have 2 stages:

1. **Deciding what parameters to measure.** This is the difficult part, and should be done before writing the Methods section.
2. **Measuring them.** This is the easy part, and should be done after writing the Methods section, to the specifications in the Methods section.

Previously, I have usually defined these stages as follows:

1. **Building a “confounders-only” regression model** to predict the outcome(s) from the confounders.
2. **Adding the exposures of primary interest** to this model as additional predictors.

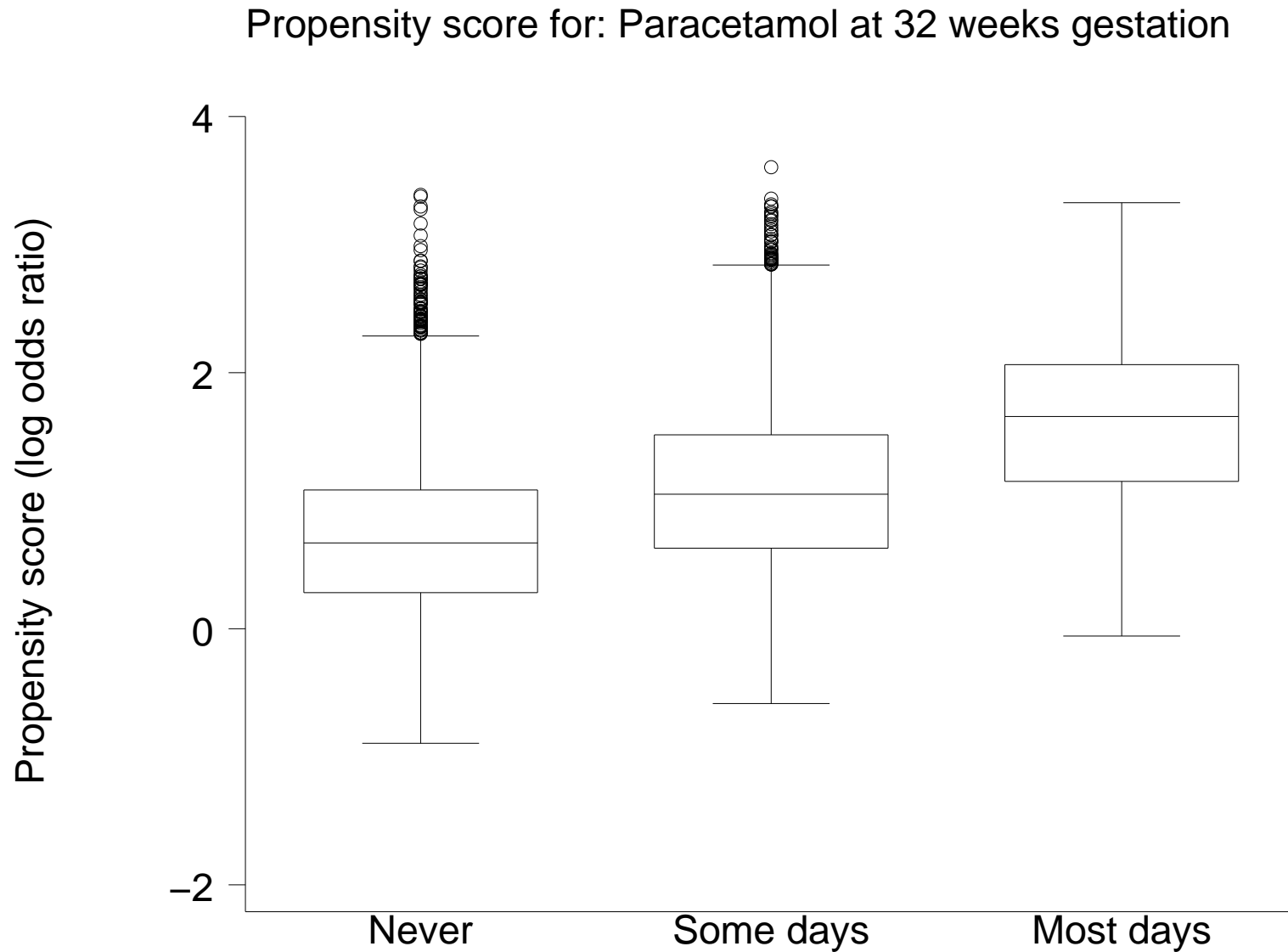
## **The method of propensity scores**

- Given an outcome  $Y$ , an exposure  $X$  and a list of confounders  $V_1, \dots, V_k$ , the **propensity score**  $W(V_1, \dots, V_k)$  is a measure of “exposure-proneness”, dependent on the confounder values.
- **Stage 1** of the method involves optimising a regression model with the exposure  $X$  as “outcome” and the confounders (or a subset of them) as “predictors”.
- The propensity score  $W(V_1, \dots, V_k)$  is defined as the fitted value of the exposure, given the confounders. (Or as some “transformation” of that fitted value, eg a linear predictor from a generalized linear model.)
- We may use this propensity score to define an ordered set of **propensity groups**.
- In **Stage 2** of the method, we carry out a “bivariate” regression analysis of  $Y$  with respect to  $X$  and the propensity group.
- So the propensity score is a single “summary confounder” that does the job of a multitude of confounders.

### **Propensity scores for prenatal paracetamol and aspirin exposure**

- The 4 primary exposures are maternal paracetamol and aspirin consumption at 18-21 weeks and 32 weeks gestation, each with 3 ordered levels ("Never", "Some days" and "Most days").
- For each exposure, we fitted an ordinal logistic regression model, using `ologit` in Stata. The exposure was the "outcome", and the 14 "primary" and 7 "secondary" confounders were "predictors".
- We defined the propensity score as the log odds ratio, calculated using the Stata `predict` command.
- We then grouped the values of each propensity score into 8 similarly-sized propensity groups, using the Stata `xtile` command.

Box plot of 32-week paracetamol propensity score by 32-week paracetamol exposure



**Table of 32-week paracetamol exposure by 32-week paracetamol propensity group**

The propensity score was grouped into 8 approximately equal groups using `xtile`, and tabulated against paracetamol exposure as shown:

8 quantiles of PS_para32g	Paracetamol at 32 weeks gestation			Total
	Never	Some days	Most days	
1	1181	333	2	1516
2	1059	455	2	1516
3	1000	511	5	1516
4	890	617	9	1516
5	833	673	10	1516
6	751	749	16	1516
7	634	856	26	1516
8	456	992	67	1515
Total	6804	5186	137	12127

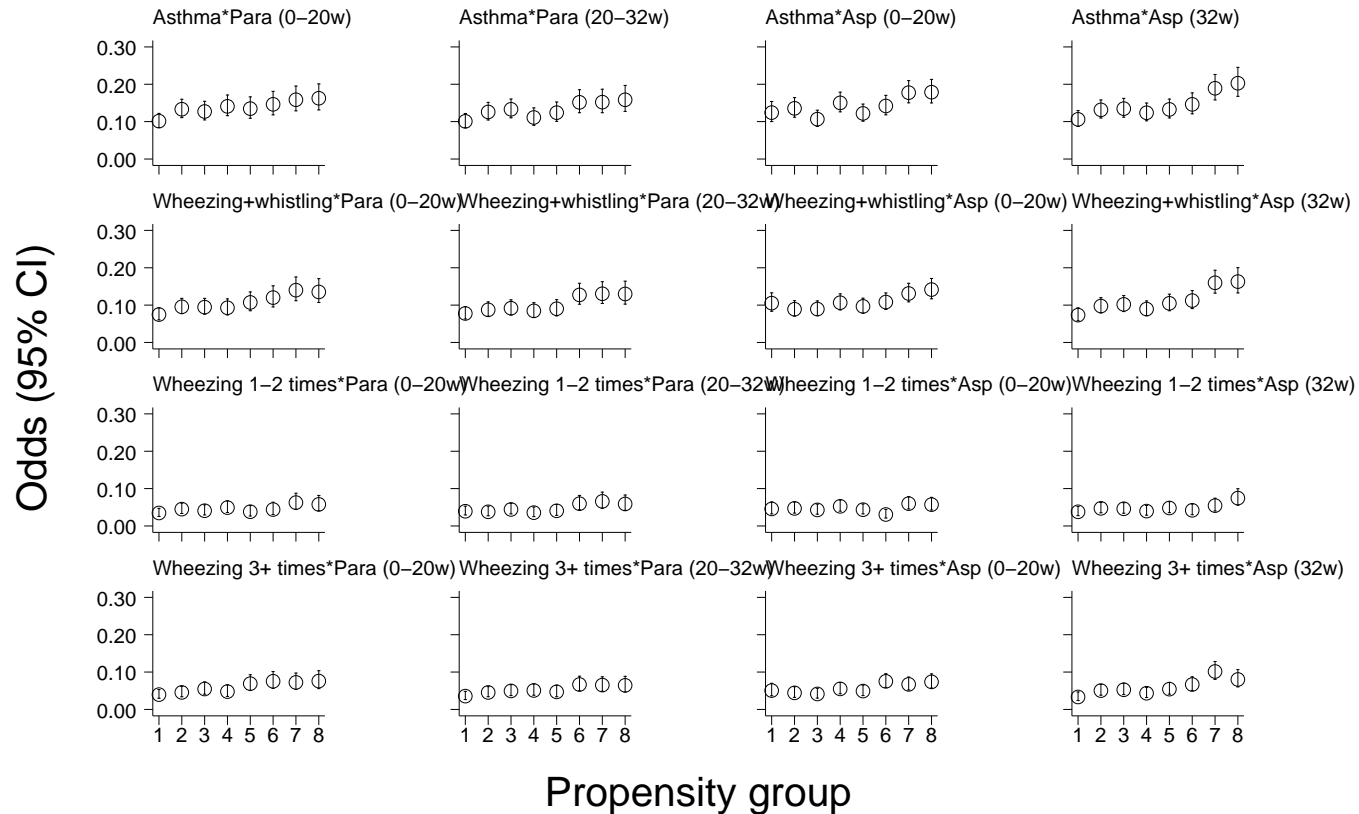
Note that higher propensity groups predict higher paracetamol exposures, but not too reliably. This suggests that we should be able to disentangle the effect of paracetamol exposure on asthma from the effect of “paracetamol-proneness” on asthma by using a “bivariate” regression model. (If paracetamol-proneness predicts paracetamol exposure “too well”, then this will widen the confidence limits for the propensity-adjusted exposure effect.)

**Stage 2 analysis: “Bivariate” regressions on exposure and propensity**

- For all 4 exposures (paracetamol and aspirin exposure at 18-20 and 32 weeks), we found that propensity group predicted exposure, but not “too well”.
- We therefore went on to Stage 2, and fitted a “bivariate” logistic regression model of each of the outcomes (asthma, wheezing, and wheezing frequency) with respect to each of the 4 exposures (with its propensity grouping).
- For each combination of exposure and outcome, we fitted a model for the outcome, with 8 unexposed odds (one for each propensity group), and 2 common odds ratios (associated with pre-natal analgesic exposure on “some days” and on “most days”).



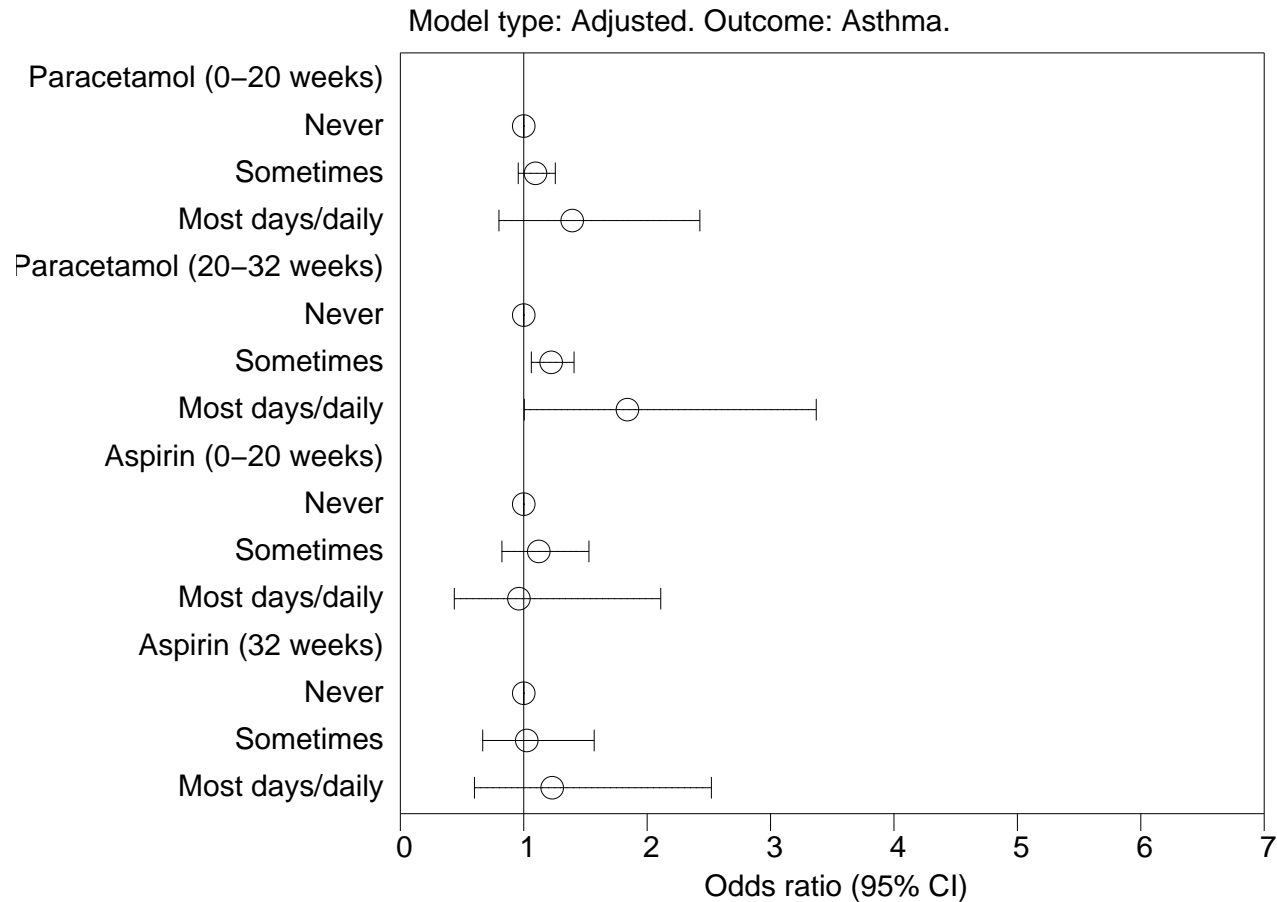
Adjusted unexposed disease odds by propensity group for the 4 exposures



Unexposed odds for each outcome and exposure

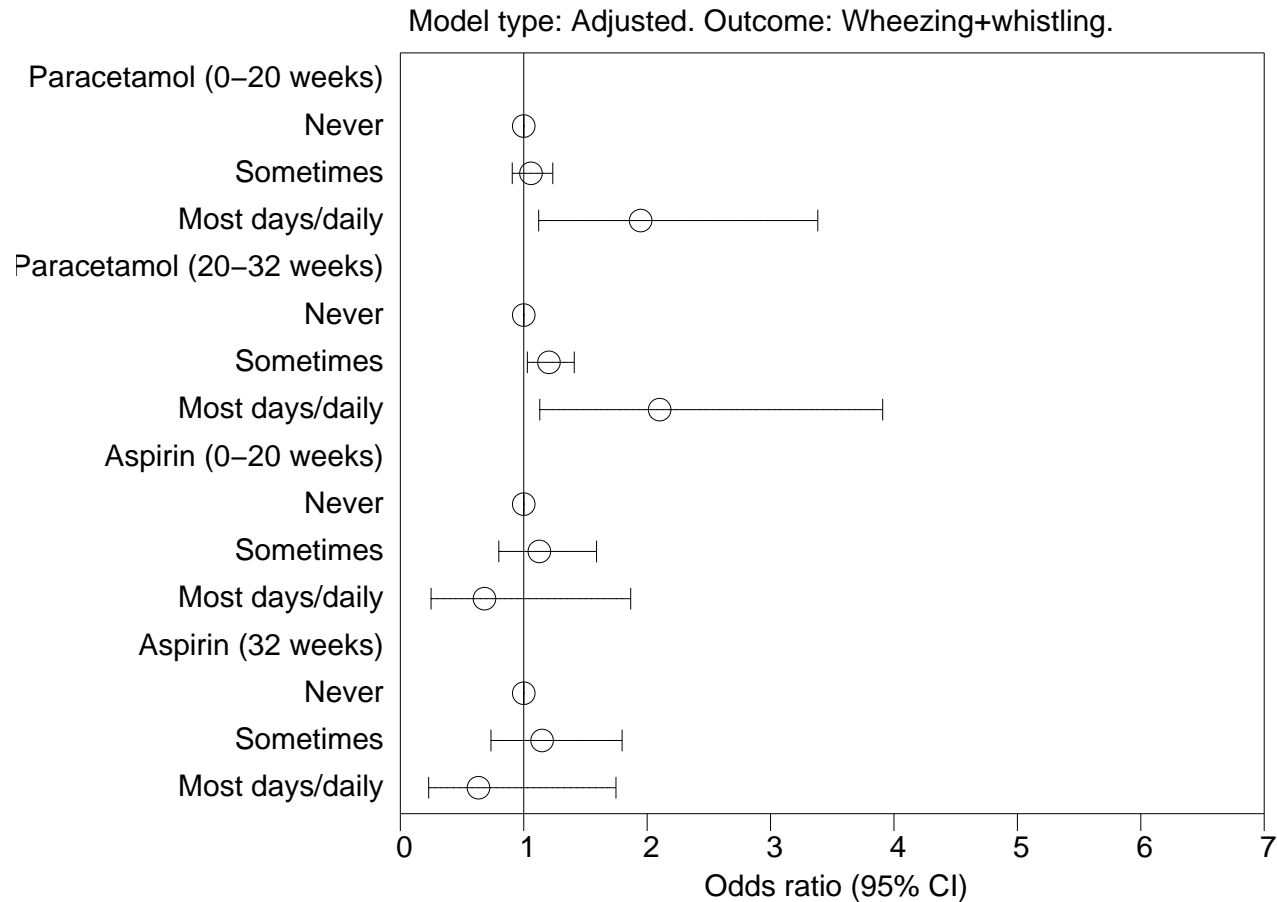
Note that children whose mothers are more “paracetamol-prone” or “aspirin-prone” are also more asthma-prone and wheeze-prone, even if unexposed before birth.

Propensity-adjusted asthma odds ratios for the 4 exposures



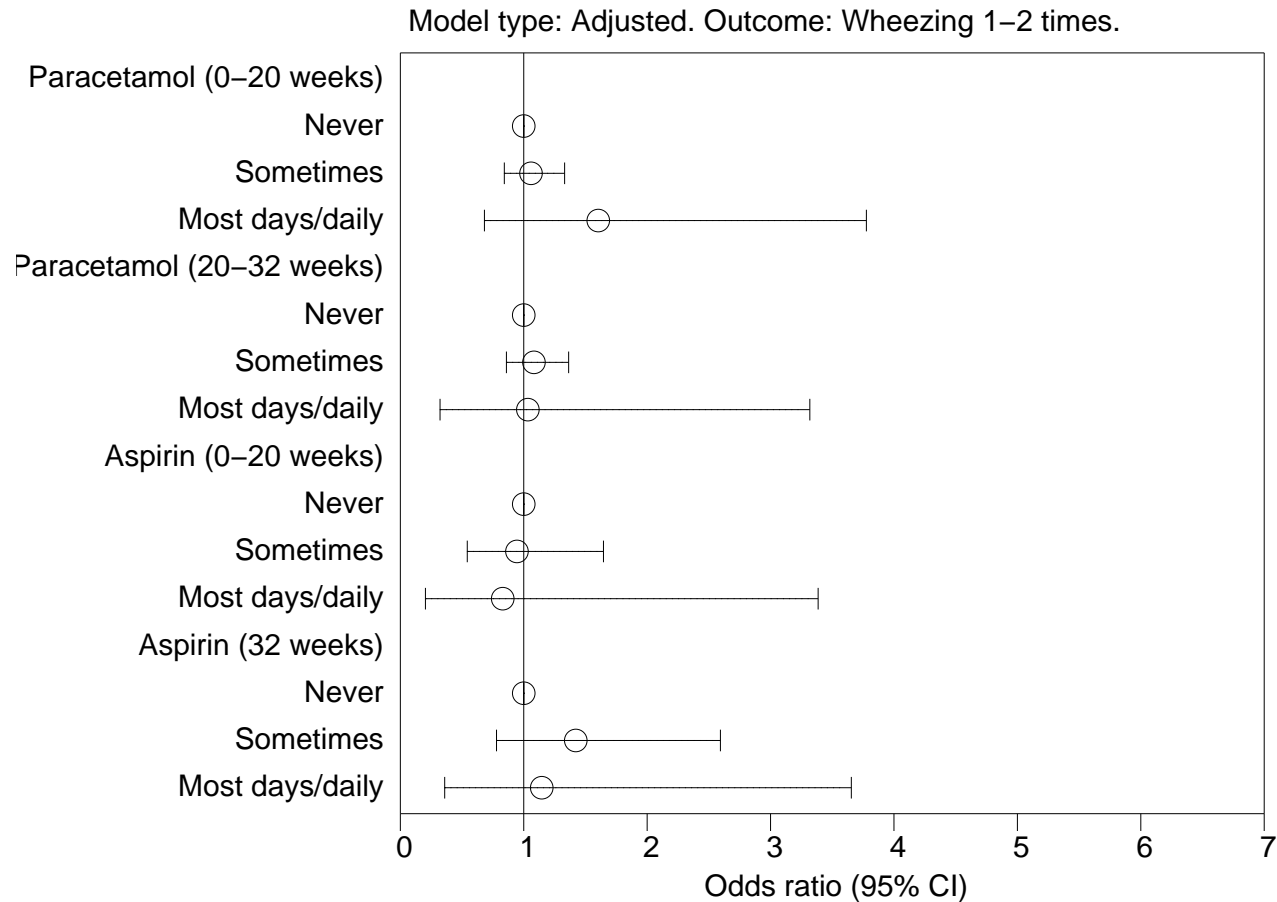
The adjusted odds ratios are similar to those from the conventional confounder-adjusted analysis involving the same confounders.

Propensity-adjusted wheezing+whistling odds ratios for the 4 exposures



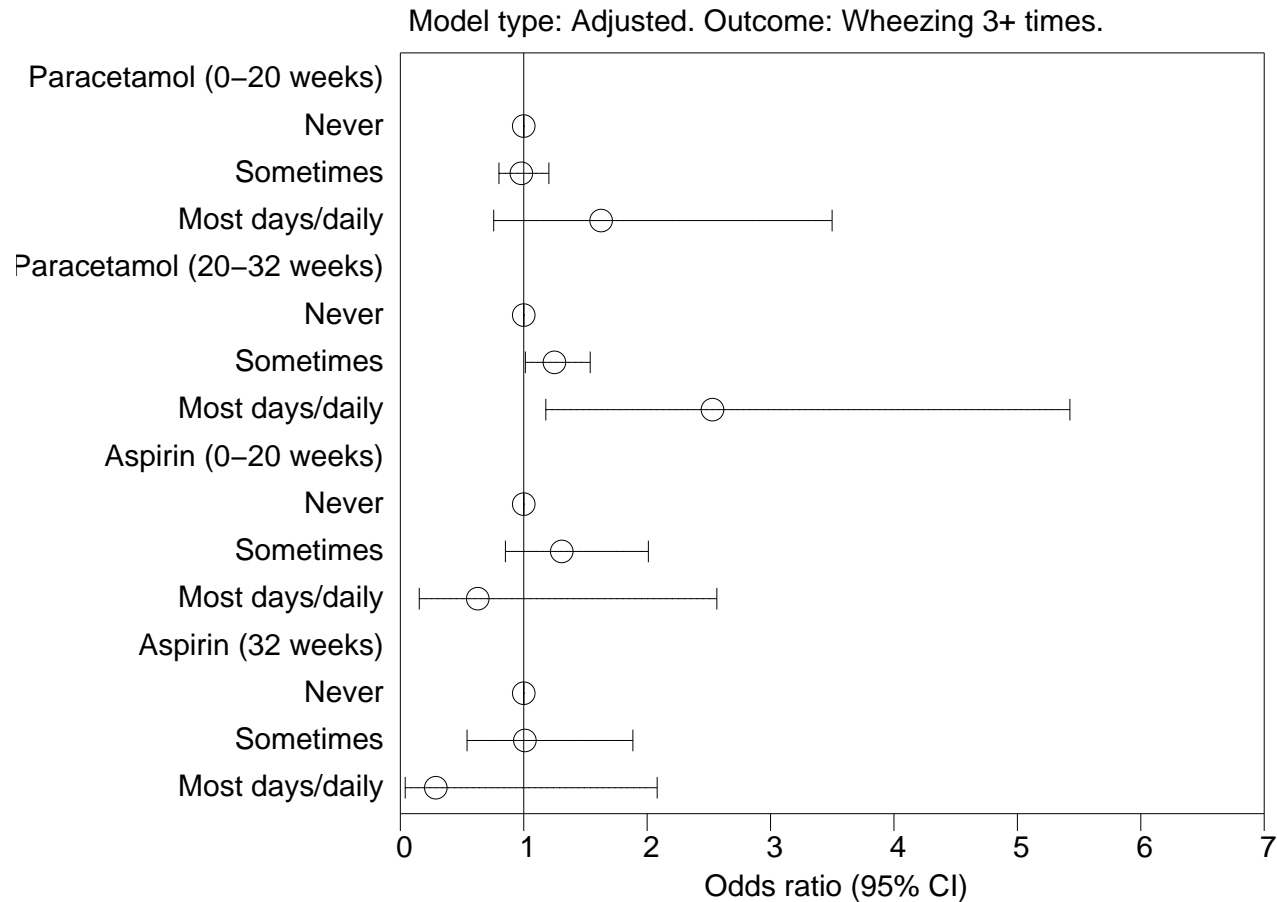
The adjusted odds ratios are similar to those from the conventional confounder-adjusted analysis involving the same confounders.

Propensity-adjusted infrequent wheezing odds ratios for the 4 exposures



The adjusted odds ratios are similar to those from the conventional confounder-adjusted analysis involving the same confounders.

Propensity-adjusted frequent wheezing odds ratios for the 4 exposures



The adjusted odds ratios are similar to those from the conventional confounder-adjusted analysis involving the same confounders.

## **Advantages of propensity scores**

The advantages of using propensity scores, rather than conventional confounder adjustment, are twofold:

- **Statistically rigorous separation of Stages 1 and 2.**
  - The outcomes are not used in Stage 1 (which deals only with exposures and confounders).
  - The user can therefore use a lot of “trial and error” methods in Stage 1 to select a confounder set and to choose the right number of propensity groups. (*Possibly* even stepwise regression!)
  - This will not invalidate the confidence intervals and/or *P*-values calculated in Stage 2 for presentation in the Results, because the formulae for calculating these are based on the *conditional* distribution of the outcome, *given* the values of the exposure and the confounders.
- **Bivariate regressions are simpler to plan than multivariate regressions.**
  - In particular, in Stage 1, we can carry out diagnostics such as plotting the propensity score and/or tabulating the propensity group against the exposure, to check that the propensity score does not predict the exposure “too well”.

## Limitations of propensity scores

Propensity scores are a good way of controlling for multiple *observed* confounders. However, they are subject to the following limitations:

- **Propensity scores cannot control for *unobserved* confounders.**
  - Therefore, a propensity-adjusted effect may still be explained by residual confounding by unobserved confounders, or even by imperfect measurement of the “observed” confounders.
  - This limitation does *not* apply to a *large* randomized trial.
  - However, it *does* apply to conventional confounder-adjusted multivariate regressions.
- **Propensity scores cannot tell us which individual confounders are “doing the confounding”.**
  - *However*, in our case, we have a large number of weak confounders, which together are expected to act as “markers” for hidden causal variables.
  - Therefore, asking which confounder is “doing the confounding” is probably like asking which straw broke the camel’s back. Or like asking which study made a meta-analysis “significant”.