

# Non-inheritable risk factors during pregnancy for congenital heart defects in offspring: a matched case-control study<sup>1</sup>

Jun Cai

Ph.D. Candidate in Ecology  
Department of Earth System Science  
Tsinghua University

May 27, 2018



清华大学  
Tsinghua University



地球系统科学系  
Department of Earth System Science

<sup>1</sup>A collaboration study with Dr. Yu Feng

① Congenital heart defect (CHD)

② Subjects and questionnaires

③ Matching controls with cases

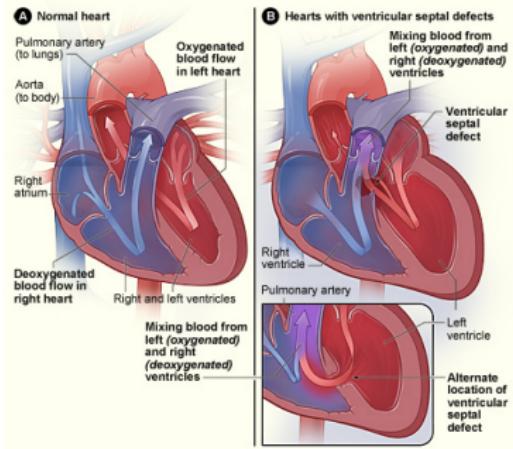
④ Conditional logistic regression

⑤ Multiple comparisions

- Pairwise multiple comparisions
- Dose-response analysis

# What is a congenital heart defect (CHD)?

A problem in the structure of the heart that is present at birth. Also known as a congenital heart anomaly or congenital heart disease.



The normal structure of the heart (left) in comparison to two common locations for a ventricular septal defect (VSD) (right), the most common form of CHD.

**Symptoms** Rapid breathing, bluish skin, poor weight gain, feeling tired

**Complications** Heart failure

**Causes** Often unknown

**Risk factors** Rubella infection during pregnancy, alcohol or tobacco, parents being closely related, poor nutritional status or obesity in the mother

**Treatment** None, catheter based procedures, **heart surgery**, heart transplantation

**Prognosis** Generally good (with treatment)

## Aetiology of CHDs

- multifactorial and a combination of **environmental** and genetic causes (e.g., Down syndrome etc.)
- difficult to establish the role of a single factor in CHDs
- aetiological heterogeneity due to several distinct subtypes of CHDs (e.g., atrial septal defect (ASD), patent ductus arteriosus (PDA, 动脉导管未闭), pulmonary stenosis (PS, 肺动脉瓣狭窄) and tetralogy of Fallot (TOF, 法洛四联症))

## Objective

- to evaluate the possible associations between exposure to environmental factors during pregnancy and the risk of isolated CHDs among offspring in Jiangsu and Anhui

## Subjects

- 6,568 cases with CHDs and 4,301 controls without birth defect
- **cases:** consecutively enrolled children who were diagnosed based on echocardiography, with some diagnoses further confirmed through cardiac catheterization and/or surgery.

To focus on **non-inheritable risk factors**, excluded patients with

- clinical features of developmental syndromes caused by gene mutations or chromosomal aberrations
  - a positive family history of CHDs in a first-degree relative
- **controls:** randomly selected from children who were undergoing routine physical examinations
  - recruited from the Children's Hospital of Nanjing Medical University between Nov. 2011 and Dec. 2017

## Questionnaires

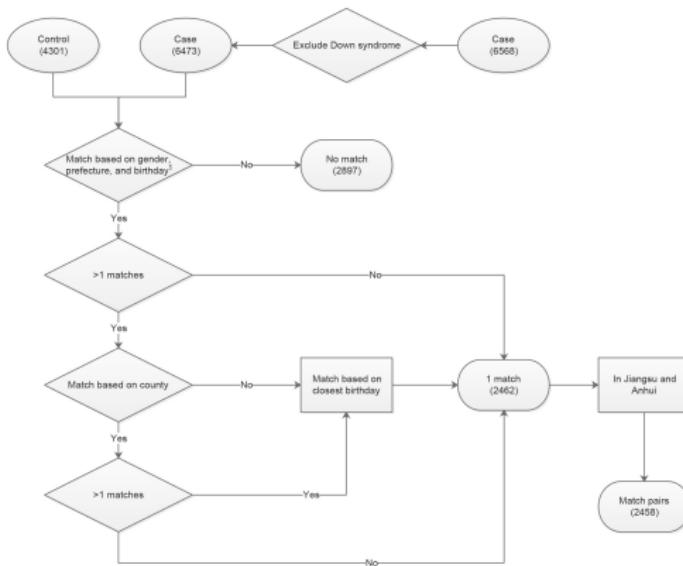
- ① A brief questionnaire to determine whether the potential study subjects were willing to participate in this research study
- ② A face-to-face interview to obtain demographic information from the subjects
- ③ A structured questionnaire consisting of 109 questions completed by the parents of subjects

### Aspects of the parents' lives

medical and family history, demographic characteristics, pregnancy history, use of medications, diet, occupation, exposure to possible toxic substances, smoking, drug and alcohol use

## Matching as a way to control for confounding

Controls were matched to cases at a rate of 1:1, according to the child's gender and birth date, and the parents' prefecture of residence.



<sup>†</sup>The birthday of case (control) was calculated by subtracting age years from date of surgery (date of survey). The difference in birthdays of case and control was matched with no more than 15 days.

```

# for a given control, find the matched case from cases
match.single <- function(control, cases, birthYMD.threshold = 15) {
  control.id <- control$id
  x <- subset(cases, sex == control$sex & province == control$province & prefecture == control$prefecture &
    abs(birthYMD - control$birthYMD) <= birthYMD.threshold)
  if (nrow(x) == 0) {
    case.id <- NA
  } else if (nrow(x) == 1) {
    case.id <- x$id
  } else {
    x1 <- subset(x, !is.na(county) & county == control$county)
    if (nrow(x1) == 0) {
      idx <- which.min(abs(x$birthYMD - control$birthYMD))
      case.id <- x[idx, "id"]
    } else if (nrow(x1) == 1) {
      case.id <- x1$id
    } else {
      idx <- which.min(abs(x1$birthYMD - control$birthYMD))
      case.id <- x1[idx, "id"]
    }
  }
  return(data.frame(control.id, case.id))
}

# match controls with cases
match <- function(controls, cases, birthYMD.threshold = 15) {
  id.lookup <- data.frame(control.id = NULL, case.id = NULL)
  for (i in 1:nrow(controls)) {
    control.idx <- controls[i, ]
    cases.remain <- subset(cases, !id %in% id.lookup$case.id)
    ret <- match.single(control.idx, cases.remain, birthYMD.threshold = 15)
    id.lookup <- rbind(id.lookup, ret)
  }
  return(id.lookup)
}

```

```

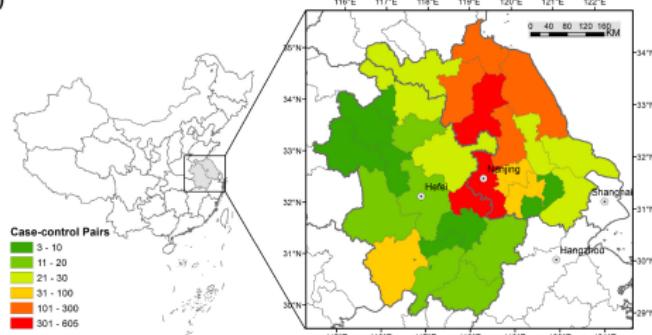
# 剔除先天愚型:Down, 先天愚型, 21-三体综合征
idx <- grep("Down|愚|三体锁", ignore.case = TRUE, case$diagnosis)
case <- case[!idx, ]

# match controls with cases
id.lookup <- match(control, case, birthYMD.threshold = 15)
id.lookup1 <- subset(id.lookup, !is.na(case.id))

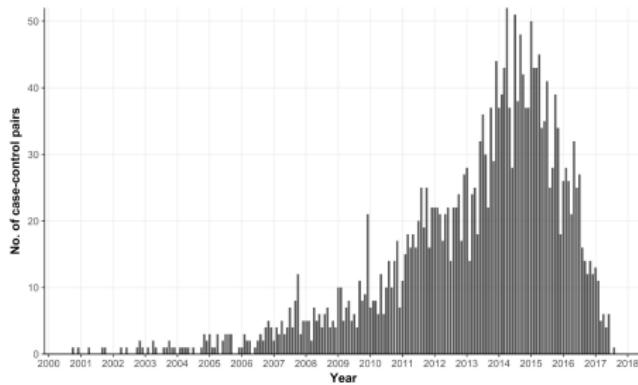
case.m <- subset(case, id %in% id.lookup1$case.id)
table(case.m$province, useNA = "ifany")
# 只分析江苏省和安徽省的病例，其它省份的病例较少
case.m <- subset(case.m, province %in% c("江苏省", "安徽省"))
control.m <- subset(control, id %in% id.lookup1$control.id)
control.m <- subset(control.m, province %in% c("江苏省", "安徽省"))
id.lookup <- subset(id.lookup1, control.id %in% control.m$id)
id.lookup$pair.id <- 1:nrow(id.lookup)
case.m <- case.m %>%
  left_join(id.lookup, by = c("id" = "case.id")) %>%
  dplyr::select(-control.id)
control.m <- control.m %>%
  left_join(id.lookup, by = c("id" = "control.id")) %>%
  dplyr::select(-case.id)

```

(a)



(b)



Geographical and temporal distributions of 2,458 matched case-control pairs.

## Conditional logistic regression

- An extension of logistic regression that allows one to take into account stratification and matching.
- Designed by Breslow et al. 1978.
- Mainly applied in observational studies and in particular epidemiology.
- The most flexible and general procedure for matched data.
- Available in R as the function *clogit* in the **survival** package. Because the log likelihood of a conditional logistic model = loglik from a Cox model with a particular data structure.

Source: wikipedia

## Application

To assess the potential risk factors for CHDs by computing and comparing the exposure ratios overall and within different CHD groups.

## Univariate conditional logistic regression

To screen potential demographic confounders

```
# data preparation
cases <- case.m %>%
  mutate(CHD = 1) %>%
  dplyr::select(pair.id, CHD, age.y, sex, tube.baby, M.production.age, term,
    production.mode, abortion, parity, gravidity, M.edu, F.production.age, F.edu,
    M.toxic.exposure, M.radioactive.exposure, M.smoke, M.smoked.years,
    M.smoke.freq, M.pregnancy.smoke, M.pregnancy.passive.smoke, F.toxic.exposure,
    F.radioactive.exposure, F.smoke, F.smoked.years, F.smoke.freq, M.drink,
    F.drink, decoration, HV.cable, chemical.plant, M.pregnancy.flu,
    M.pregnancy.flu.time, M.pregnancy.complication, M.med, M.pregnancy.med,
    M.pregnancy.med.time, M.pregnancy.med.name, M.pregnancy.folic.acid,
    M.oral.contraceptive) %>%
  arrange(pair.id)
# same manipulation for controls except for CHD = 0
mydata <- rbind(cases, controls)
```

```
# univariate conditional logistic regression
library(survival)
# case和control母亲教育水平
x <- xtabs(~ CHD + M.edu, data = mydata)
round(x / rowSums(x) * 100, 2)

mydata$M.edu <- factor(mydata$M.edu, levels = c("4", "3", "2", "1"))
mylogit <- clogit(CHD ~ M.edu + strata(pair.id), data = mydata)
summary(mylogit)

## Call:
## coxph(formula = Surv(rep(1, 4916L), CHD) ~ M.edu + strata(pair.id),
##       data = mydata, method = "exact")
##
##    n= 4916, number of events= 2458
##
##          coef exp(coef) se(coef)   z Pr(>|z|)
## M.edu3  0.36931   1.44674  0.07221 5.114 3.15e-07 ***
## M.edu2  0.30237   1.35306  0.07938 3.809 0.000139 ***
## M.edu1  0.12417   1.13221  0.25134 0.494 0.621274
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## M.edu3     1.447      0.6912    1.2558     1.667
## M.edu2     1.353      0.7391    1.1581     1.581
## M.edu1     1.132      0.8832    0.6918     1.853
##
## Rsquare= 0.006  (max possible= 0.5 )
## Likelihood ratio test= 27.84 on 3 df,  p=4e-06
## Wald test           = 27.44 on 3 df,  p=5e-06
## Score (logrank) test = 27.71 on 3 df,  p=4e-06
```

**Table 1**

Demographic characteristics of the matched case and control participants.

Variables	Control participants (N = 2458), n (%)	Case participants (N = 2458), n (%)	cORs (95% CI) <sup>a</sup>
Age (years) at interview (mean (SD))	2.69 (2.56)	2.20 (2.28)	
Gender			
Male	1262 (51.34%)	1262 (51.34%)	
Female	1196 (48.66%)	1196 (48.66%)	
IVF (in-vitro fertilization)			
Yes	18 (0.73%)	15 (0.61%)	0.83 (0.42, 1.65)
No	2440 (99.27%)	2443 (99.39%)	1
Maternal age (years) (mean (SD))	24.17 (4.19)	24.38 (4.08)	
>30	252 (10.25%)	263 (10.70%)	1.03 (0.85, 1.24)
<20	332 (13.51%)	285 (11.59%)	0.84 (0.71, 1.00)*
20–30	1874 (76.24%)	1910 (77.71%)	1
Gestational age			
Preterm birth	923 (37.55%)	921 (37.47%)	1.00 (0.89, 1.12)
Full-term birth	1535 (62.45%)	1537 (62.53%)	1
Fertility			
C-section	1045 (42.51%)	1026 (41.74%)	0.97 (0.87, 1.09)
Spontaneous delivery	1413 (57.49%)	1432 (58.26%)	1
History of miscarriage			
Yes	202 (8.22%)	226 (9.19%)	1.13 (0.93, 1.38)
No	2256 (91.78%)	2232 (90.81%)	1
Parity			
1	1830 (74.45%)	1848 (75.18%)	1
2	628 (25.55%)	610 (24.82%)	0.96 (0.85, 1.0)
Gravidity			
1	1669 (67.90%)	1657 (67.41%)	1
2	668 (27.18%)	668 (27.18%)	1.01 (0.89, 1.14)
≥3	121 (4.92%)	133 (5.41%)	1.11 (0.86, 1.43)
Maternal education status			
Illiteracy	37 (1.51%)	32 (1.30%)	1.13 (0.69, 1.85)
Middle school or below	690 (28.07%)	722 (29.37%)	1.35 (1.16, 1.58)***
High school	1033 (42.03%)	1160 (47.19%)	1.45 (1.26, 1.67)***
College or above	698 (28.40%)	544 (22.13%)	1
Paternal age (years) (mean (SD))	26.82 (4.22)	27.01 (4.12)	
>30	461 (18.76%)	467 (19.00%)	1.00 (0.86, 1.15)
<22	335 (13.63%)	302 (12.29%)	0.89 (0.75, 1.05)
22–30	1662 (67.62%)	1689 (68.71%)	1
Paternal education status			
Illiteracy	9 (0.37%)	8 (0.33%)	0.94 (0.36, 2.45)
Middle school or below	708 (28.80%)	665 (27.05%)	1.00 (0.85, 1.18)
High school	1249 (50.81%)	1322 (53.78%)	1.12 (0.97, 1.31)
College or above	492 (20.02%)	463 (18.84%)	1

\*(P &lt; 0.05) or \*\*\*(P &lt; 0.001) indicates a significant difference between the cases and controls based on a univariate conditional logistic regression analysis.

<sup>a</sup> CI, confidence interval; cORs, crude odds ratios.

## Multivariate conditional logistic regression

- To evaluate the potential risk factor's association with CHD while simultaneously controlling for maternal age, parity, gravidity and educational status.
- Examine the collinearity of covariates by Generalized Variance Inflation Factor (GVIF). VIFs exceeding 4 warrant further investigation, while VIFs exceeding 10 are signs of serious multicollinearity requiring correction.

```
# case和control母亲孕期是否感染流感
x <- xtabs(~ CHD + M.pregnancy.flu, data = mydata)
round(x / rowSums(x) * 100, 2)

mydata$M.pregnancy.flu <- factor(mydata$M.pregnancy.flu)
mylogit <- clogit(CHD ~ M.pregnancy.flu + M.production.age + parity +
                     gravidity + M.edu + strata(pair.id), data = mydata)
summary(mylogit)
# examine the collinearity of covariates
library(car)
vif(mylogit)
```

**Table 2**

Analysis of the parental lifestyle exposures during maternal pregnancy among the matched case and controls.

Maternal characteristics	Control participants (N = 2458), n (%)	Case participants (N = 2458), n (%)	aORs (95% CI) <sup>a</sup>	VSD (N = 1493)	ASD (N = 786)	PDA (N = 362)	PS (N = 129)	TOF (N = 193)	VSD & ASD (N = 272)	Multiple defects (N = 257)
	aORs (95% CI)									
Maternal toxic exposure	58 (2.36%)	46 (1.87%)	0.76 (0.51, 1.12)	0.68 (0.40, 1.14)	0.62 (0.33, 1.18)	0.95 (0.30, 2.98)	0.21 (0.02, 1.85)	8.66 (1.05, 71.70)*	1.27 (0.45, 3.61)	1.67 (0.53, 5.32)
Maternal radioactive exposure	40 (1.63%)	40 (1.63%)	1.03 (0.66, 1.61)	1.36 (0.78, 2.37)	0.76 (0.34, 1.67)	0.21 (0.04, 1.02)	0.64 (0.11, 3.61)	1.46 (0.31, 6.85)	1.46 (0.45, 4.70)	1.50 (0.41, 5.43)
Maternal smoking	87 (3.54%)	115 (4.68%)	1.37 (1.02, 1.83)*	1.47 (1.01, 2.14)*	1.57 (0.97, 2.56)	0.99 (0.47, 2.13)	0.91 (0.25, 3.29)	1.73 (0.67, 4.47)	1.99 (0.93, 4.29)	1.02 (0.40, 2.61)
Smoking (years) (mean (SD))	9.68 (4.62)	10.31 (4.73)								
≤10	49 (1.99%)	58 (2.36%)	1.25 (0.85, 1.86)	1.44 (0.87, 2.37)	2.02 (1.00, 4.05)*	0.71 (0.20, 2.54)	0.66 (0.15, 2.88)	0.98 (0.28, 3.42)	3.56 (1.07, 11.82)*	0.62 (0.19, 2.08)
>10	38 (1.55%)	55 (2.24%)	1.44 (0.94, 2.20)	1.46 (0.82, 2.58)	1.23 (0.62, 2.42)	1.21 (0.46, 3.18)	2.32 (0.21, 25.40)	2.45 (0.62, 9.79)	1.21 (0.43, 3.41)	1.73 (0.36, 8.23)
Cigarettes/day										
≤5	35 (1.42%)	47 (1.91%)	1.38 (0.88, 2.14)	1.50 (0.86, 2.59)	1.15 (0.55, 2.41)	0.85 (0.30, 2.38)	0.82 (0.16, 4.32)	-	1.81 (0.57, 5.76)	0.63 (0.15, 2.67)
>5	52 (2.12%)	68 (2.77%)	1.36 (0.94, 1.99)	1.45 (0.88, 2.39)	1.97 (1.03, 3.74)*	1.20 (0.39, 3.70)	1.07 (0.15, 7.87)	0.87 (0.29, 2.63)	2.14 (0.77, 5.97)	1.50 (0.42, 5.40)
Smoking during pregnancy	57 (2.32%)	104 (4.23%)	1.89 (1.35, 2.64)***	2.07 (1.34, 3.21)**	2.45 (1.38, 4.37)**	1.81 (0.70, 4.65)	3.81 (0.40, 36.46)	1.86 (0.68, 5.12)	2.94 (1.19, 7.28)*	1.60 (0.52, 4.90)
Passive smoking during pregnancy	734 (29.86%)	738 (30.02%)	1.01 (0.89, 1.14)	0.98 (0.83, 1.15)	0.95 (0.77, 1.19)	0.94 (0.68, 1.30)	1.15 (0.63, 2.11)	1.01 (0.65, 1.57)	1.06 (0.71, 1.57)	0.99 (0.67, 1.47)
Paternal toxic exposure	22 (0.90%)	28 (1.14%)	1.28 (0.72, 2.28)	1.01 (0.48, 2.15)	1.52 (0.48, 4.83)	0.65 (0.15, 2.79)	-	4.13 (0.40, 42.45)	0.51 (0.03, 7.47)	0.93 (0.18, 4.66)
Paternal radioactive exposure	38 (1.55%)	49 (1.99%)	1.28 (0.83, 1.96)	1.48 (0.84, 2.62)	1.12 (0.49, 2.57)	0.63 (0.20, 2.01)	2.03 (0.46, 8.90)	2.23 (0.41, 12.03)	0.80 (0.11, 5.89)	0.62 (0.17, 2.24)
Paternal smoking	1222 (49.72%)	1260 (51.26%)	1.08 (0.96, 1.20)	1.14 (0.99, 1.31)	1.07 (0.87, 1.30)	1.10 (0.84, 1.45)	1.23 (0.74, 2.03)	1.05 (0.71, 1.58)	1.20 (0.86, 1.67)	0.95 (0.67, 1.35)
Smoking (years) (mean (SD))	15.88 (4.58)	16.08 (4.43)								
≤10	172 (7.00%)	153 (6.22%)	0.91 (0.72, 1.15)	0.92 (0.68, 1.24)	1.03 (0.68, 1.57)	0.89 (0.45, 1.75)	0.85 (0.32, 2.24)	1.46 (0.71, 2.97)	1.41 (0.70, 2.85)	0.76 (0.34, 1.69)
>10	1050 (42.72%)	1106 (45.00%)	1.10 (0.98, 1.24)	1.17 (1.01, 1.36)*	1.07 (0.87, 1.32)	1.14 (0.86, 1.52)	1.30 (0.77, 2.19)	0.98 (0.64, 1.50)	1.17 (0.83, 1.65)	0.98 (0.68, 1.40)
Cigarettes/day										
≤5	822 (33.44%)	819 (33.32%)	1.05 (0.92, 1.18)	1.14 (0.97, 1.34)	0.99 (0.80, 1.24)	1.05 (0.77, 1.44)	1.11 (0.62, 1.97)	1.12 (0.72, 1.72)	1.23 (0.85, 1.79)	0.94 (0.64, 1.39)
>5	399 (16.23%)	440 (17.90%)	1.15 (0.98, 1.34)	1.14 (0.94, 1.40)	1.26 (0.94, 1.69)	1.20 (0.80, 1.80)	1.45 (0.73, 2.89)	0.90 (0.50, 1.62)	1.13 (0.71, 1.82)	0.98 (0.58, 1.65)
Maternal binge drinking <sup>b</sup>	383 (15.58%)	442 (17.98%)	1.18 (1.01, 1.38)*	1.19 (0.97, 1.44)	1.03 (0.78, 1.35)	1.02 (0.67, 1.54)	1.19 (0.56, 2.53)	1.83 (1.05, 3.18)*	0.92 (0.58, 1.47)	0.79 (0.50, 1.26)
Paternal binge drinking	956 (38.89%)	982 (39.95%)	1.05 (0.94, 1.18)	1.07 (0.92, 1.23)	1.01 (0.82, 1.23)	0.93 (0.68, 1.26)	1.14 (0.68, 1.93)	1.17 (0.76, 1.79)	1.14 (0.80, 1.63)	1.09 (0.76, 1.56)
Decoration	40 (1.63%)	49 (1.99%)	1.24 (0.82, 1.89)	1.37 (0.80, 2.33)	0.94 (0.42, 2.10)	1.08 (0.34, 3.42)	-	0.77 (0.20, 2.95)	1.14 (0.33, 3.95)	0.47 (0.15, 1.48)
High voltage cable nearby	59 (2.40%)	62 (2.52%)	1.05 (0.73, 1.52)	1.32 (0.84, 2.08)	0.85 (0.42, 1.73)	1.57 (0.66, 3.75)	-	0.99 (0.24, 4.02)	1.35 (0.40, 4.55)	2.11 (0.73, 6.16)
Chemical factory nearby	18 (0.73%)	19 (0.77%)	1.04 (0.52, 2.08)	1.65 (0.67, 4.04)	0.57 (0.14, 2.43)	0.66 (0.11, 4.09)	0.87 (0.05, 14.74)	0.47 (0.04, 5.56)	3.05 (0.30, 30.50)	-

<sup>a</sup> CI, confidence interval; aORs, adjusted odds ratios; multivariate conditional logistic regression was used to compute the adjusted ORs while simultaneously controlling for maternal age, parity, gravidity, and education status.<sup>b</sup> Binge drinking: an intake of ≥5 drinks on a single occasion.\* Statistically significant ( $P < 0.05$ ).\*\* Statistically significant ( $P < 0.01$ ).\*\*\* Statistically significant ( $P < 0.001$ ).

Table 3

Analyses of maternal diseases, medicine consumptions and pregnancy supplementations among the matched cases and controls.

Maternal characteristics	Control participants (N = 2458), n (%)	Case participants (N = 2458), n (%)	aORs <sup>a</sup> (95% CI)	VSD (N = 1493)	ASD (N = 786)	PDA (N = 362)	PS (N = 129)	TOF (N = 193)	VSD & ASD (N = 272)	Multiple defects (N = 257)
	aORs (95% CI)									
Influenza	446 (18.14%)	554 (22.54%)	1.31 (1.14, 1.52) <sup>***</sup>	1.36 (1.14, 1.63) <sup>***</sup>	1.37 (1.05, 1.79) <sup>*</sup>	1.31 (0.87, 1.97)	1.29 (0.68, 2.45)	1.60 (0.96, 2.68)	1.56 (1.01, 2.42) <sup>*</sup>	1.65 (1.04, 2.60) <sup>*</sup>
Influenza infection time <sup>b</sup>										
1st trimester	109 (4.43%)	114 (4.64%)	1.11 (0.84, 1.45)	1.20 (0.86, 1.67)	1.12 (0.68, 1.85)	1.38 (0.54, 3.52)	1.24 (0.39, 3.92)	1.30 (0.43, 3.94)	1.79 (0.74, 4.31)	1.31 (0.41, 4.15)
2nd trimester	256 (10.41%)	331 (13.47%)	1.38 (1.15, 1.65) <sup>***</sup>	1.49 (1.18, 1.87) <sup>***</sup>	1.44 (1.04, 1.98) <sup>*</sup>	1.33 (0.81, 2.17)	1.10 (0.44, 2.78)	1.47 (0.81, 2.67)	1.74 (0.99, 3.03)	1.74 (0.98, 3.08)
3rd trimester	81 (3.30%)	109 (4.43%)	1.42 (1.05, 1.91) <sup>*</sup>	1.24 (0.86, 1.80)	1.52 (0.87, 2.63)	1.22 (0.52, 2.86)	2.00 (0.47, 8.53)	3.51 (0.72, 17.21)	1.10 (0.48, 2.54)	1.68 (0.70, 4.02)
Complication	371 (15.09%)	407 (16.56%)	1.11 (0.95, 1.29)	1.04 (0.86, 1.27)	1.13 (0.85, 1.49)	1.37 (0.91, 2.08)	1.07 (0.60, 1.94)	1.41 (0.76, 2.61)	1.02 (0.62, 1.68)	0.61 (0.38, 0.98) <sup>*</sup>
Medicine	510 (20.75%)	752 (30.59%)	1.70 (1.49, 1.95) <sup>***</sup>	1.63 (1.37, 1.95) <sup>***</sup>	1.62 (1.29, 2.05) <sup>***</sup>	2.16 (1.50, 3.11) <sup>***</sup>	2.37 (1.17, 4.80) <sup>*</sup>	1.57 (0.98, 2.53)	1.73 (1.14, 2.62) <sup>*</sup>	2.29 (1.43, 3.68) <sup>***</sup>
Medicine intake time										
1st trimester	144 (5.86%)	268 (10.90%)	2.11 (1.70, 2.61) <sup>***</sup>	2.08 (1.58, 2.75) <sup>***</sup>	1.95 (1.34, 2.82) <sup>***</sup>	3.63 (1.87, 7.05) <sup>***</sup>	5.84 (1.76, 19.31) <sup>**</sup>	1.37 (0.63, 2.97)	2.08 (1.06, 4.08) <sup>*</sup>	5.57 (2.40, 12.91) <sup>***</sup>
2nd trimester	60 (2.44%)	106 (4.31%)	1.94 (1.40, 2.68) <sup>***</sup>	1.64 (1.09, 2.47) <sup>*</sup>	2.29 (1.27, 4.13) <sup>**</sup>	1.46 (0.62, 3.46)	1.33 (0.23, 7.83)	2.84 (0.74, 10.98)	1.78 (0.65, 4.89)	4.12 (0.87, 19.53)
3rd trimester	94 (3.82%)	147 (5.98%)	1.76 (1.34, 2.32) <sup>***</sup>	1.72 (1.20, 2.47) <sup>*</sup>	1.87 (1.15, 3.04) <sup>*</sup>	1.42 (0.73, 2.77)	4.21 (0.74, 23.90)	1.83 (0.77, 4.36)	1.94 (0.87, 4.32)	1.50 (0.57, 3.92)
Antibiotics	155 (6.31%)	272 (11.07%)	2.00 (1.61, 2.47) <sup>***</sup>	1.95 (1.48, 2.59) <sup>***</sup>	1.63 (1.12, 2.36) <sup>*</sup>	2.12 (1.24, 3.61) <sup>**</sup>	4.31 (1.42, 13.04) <sup>**</sup>	3.24 (1.39, 7.57) <sup>**</sup>	1.88 (0.99, 3.56)	4.00 (1.70, 9.41) <sup>*</sup>
Diet pills	24 (0.98%)	38 (1.55%)	1.62 (0.95, 2.75)	1.67 (0.80, 3.47)	2.42 (0.99, 5.94)	2.66 (0.60, 11.78)	-	0.33 (0.07, 1.71)	2.49 (0.24, 25.52)	7.11 (0.76, 66.54)
Antidepressant	16 (0.65%)	25 (1.02%)	1.70 (0.89, 3.22)	1.06 (0.46, 2.40)	2.53 (0.78, 8.13)	1.85 (0.43, 8.08)	3.04 (0.18, 51.08)	2.14 (0.11, 40.63)	1.02 (0.17, 5.92)	2.96 (0.55, 15.86)
Antitumour drug	3 (0.12%)	6 (0.24%)	2.74 (0.52, 14.45)	2.22 (0.39, 12.52)	-	-	-	-	-	-
Tocolytic agent	42 (1.71%)	78 (3.17%)	2.03 (1.39, 2.99) <sup>***</sup>	1.63 (1.00, 2.66) <sup>*</sup>	2.58 (1.29, 5.16) <sup>**</sup>	3.26 (0.85, 12.44)	-	2.08 (0.56, 7.72)	1.22 (0.37, 4.08)	2.91 (0.72, 11.75)
Analgesics	32 (1.30%)	67 (2.73%)	2.31 (1.49, 3.58) <sup>***</sup>	2.74 (1.57, 4.78) <sup>***</sup>	2.78 (1.31, 5.91) <sup>**</sup>	1.30 (0.42, 4.04)	-	2.56 (0.41, 16.09)	8.06 (1.67, 38.89) <sup>**</sup>	2.68 (0.73, 9.94)
Tranquillizer	25 (1.02%)	35 (1.42%)	1.66 (0.98, 2.81)	1.55 (0.81, 2.99)	1.57 (0.51, 4.86)	1.81 (0.27, 12.33)	7.82 (0.82, 74.98)	0.38 (0.04, 3.93)	0.47 (0.04, 5.99)	3.34 (0.32, 34.82)
Contraceptive	248 (10.09%)	305 (12.41%)	1.29 (1.07, 1.55) <sup>**</sup>	1.24 (0.97, 1.58)	1.26 (0.92, 1.73)	1.84 (1.11, 3.04) <sup>*</sup>	0.59 (0.21, 1.65)	1.25 (0.64, 2.41)	1.47 (0.85, 2.54)	1.14 (0.58, 2.25)
Folate	1560 (63.47%)	1080 (43.94%)	0.46 (0.41, 0.52) <sup>***</sup>	0.45 (0.39, 0.53) <sup>***</sup>	0.45 (0.36, 0.55) <sup>***</sup>	0.59 (0.44, 0.79) <sup>***</sup>	0.42 (0.25, 0.73) <sup>**</sup>	0.35 (0.22, 0.56) <sup>***</sup>	0.48 (0.33, 0.69) <sup>***</sup>	0.78 (0.53, 1.12)

<sup>a</sup> CI, confidence interval; aORs, adjusted odds ratios; multivariate logistic regression analysis was used to compute the adjusted ORs while simultaneously controlling for maternal age, parity, gravidity, and education status.<sup>b</sup> 1st trimester refers to the first 3 months of pregnancy, 2nd trimester refers to the 4th–6th months of pregnancy, and the remaining 4 months of pregnancy are defined as the 3rd trimester.<sup>\*</sup> Statistically significant ( $P < 0.05$ ).<sup>\*\*</sup> Statistically significant ( $P < 0.01$ ).<sup>\*\*\*</sup> Statistically significant ( $P < 0.001$ ).

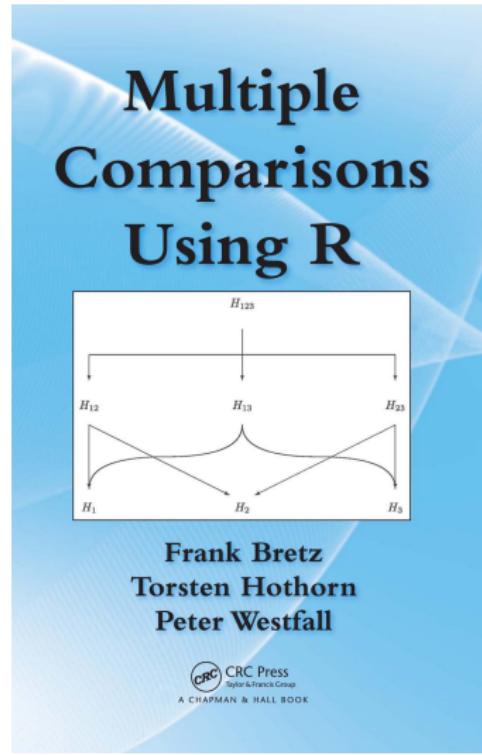
## Summary

Significant risk factors include a lower maternal educational level, maternal smoking, binge drinking, medicine consumption, and influenza infection during pregnancy, whereas significant protective factor includes maternal folic acid supplementation.

## Further question

To assess whether the number of exposure factors (0, 1, and  $\geq 2$ ) showed a dose-response effect on the occurrence of CHDs.

- All pairwise comparisions using the well-known Tukey test, which is the standard procedure in this situation.
- Trend test for a dose-response effect, which can borrow strength from neighboring dose levels.



Details refer to Chapter 4 of Bretz, Westfall, and Hothorn 2010. Demos reproducing analyzes presented in this book are included in **multcomp** package.

```

# data preparation
cases <- case.m %>%
  mutate(CHD = 1) %>%
  dplyr::select(pair.id, CHD, M.production.age, parity, gravidity, M.edu,
    M.smoke, M.drink, M.pregnancy.flu, M.med) %>%
  arrange(pair.id)
# same manipulation for controls except for CHD = 0
mydata <- rbind(cases, controls)

# recode a lower maternal educational level into 1
mydata$M.edu1 <- ifelse(mydata$M.edu %in% c(1, 2, 3), 1, 0)

# number of exposures
mydata$n.exp <- rowSums(mydata[, c("M.edu1", "M.smoke", "M.drink",
  "M.pregnancy.flu", "M.med")])

mydata <- mydata %>%
  mutate(n.exp = as.character(n.exp)) %>%
  mutate(n.exp = case_when(
    .n.exp == "0" ~ "0",
    .n.exp == "1" ~ "1",
    TRUE ~ ">=2"
  ))
mydata$n.exp <- factor(mydata$n.exp, levels = c("0", "1", ">=2"))

# case和control暴露风险因子个数
x <- xtabs(~ CHD + n.exp, data = mydata)
round(x / rowSums(x) * 100, 2)

# fit a conditional logistic regression model
mylogit <- clogit(CHD ~ n.exp + M.production.age + parity +
  gravidity + strata(pair.id), data = mydata)
summary(mylogit)
vif(mylogit)

```

- The input model for `glht` is a conditional logistic regression model rather than an analysis of variance model.

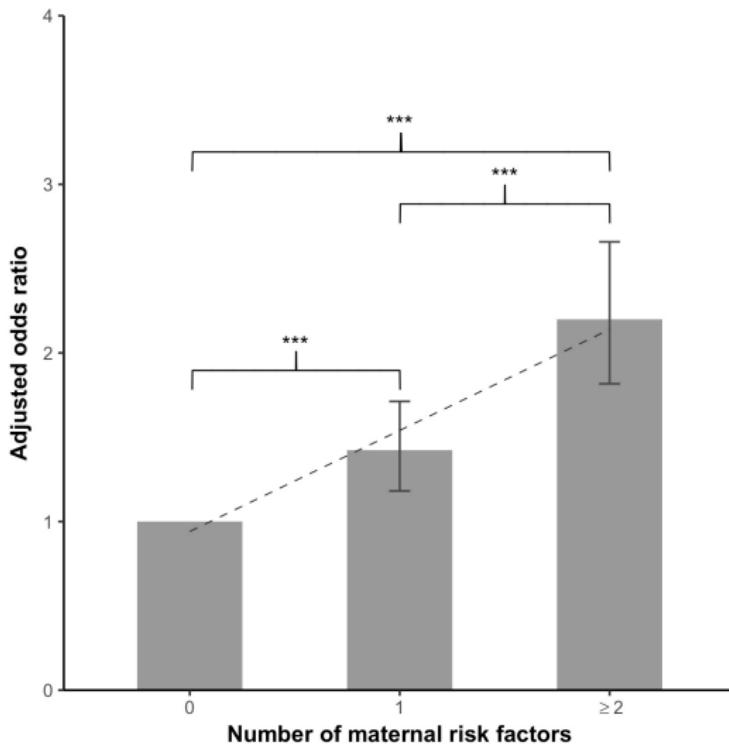
```
library(multcomp)
# all pairwise comparison
CHD.mc <- glht(mylogit, linfct = mcp(n.exp = "Tukey"), alternative = "greater")
summary(CHD.mc, test = adjusted(type = "bonferroni"))

##
## ^^^I Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
## Fit: coxph(formula = Surv(rep(1, 4916L), CHD) ~ n.exp + M.production.age +
##       parity + gravidity + strata(pair.id), data = mydata, method = "exact")
##
## Linear Hypotheses:
##             Estimate Std. Error z value Pr(>z)
## 1 - 0 <= 0    0.35230   0.09473  3.719   3e-04 ***
## >=2 - 0 <= 0   0.78757   0.09706  8.114 6.66e-16 ***
## >=2 - 1 <= 0   0.43527   0.06328  6.879 9.05e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- bonferroni method)
```

- Use Williams contrasts and the second contrast allows comparison of the weighted average of 1 and  $\geq 2$  groups with 0 group.

```
# dose-response analysis
CHD.mc2 <- glht(mylogit, linfct = mcp(n.exp = "Williams"), alternative = "greater")
summary(CHD.mc2, test = adjusted(type = "bonferroni"))

##
## ^^^I Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Williams Contrasts
##
## Fit: coxph(formula = Surv(rep(1, 4916L), CHD) ~ n.exp + M.production.age +
##       parity + gravidity + strata(pair.id), data = mydata, method = "exact")
##
## Linear Hypotheses:
##             Estimate Std. Error z value   Pr(>z)
## C 1 <= 0    0.78757   0.09706   8.114 4.44e-16 ***
## C 2 <= 0    0.56684   0.09052   6.262 3.80e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- bonferroni method)
```



The dose-response effect is significant. The higher the number of maternal exposures, the greater the risk of developing CHDs in offspring.

## Details

More details on results, particular discussion on biological mechanisms and relevant evidence for significant exposure factors and comparisons with other studies, refer to Feng et al. 2018.

## Reproducible research

Data and analysis code for this study are publicly available at

 [github.com/caijun/CHD](https://github.com/caijun/CHD)

# References

-  N. E. Breslow et al. "Estimation of multiple relative risk functions in matched case-control studies". In: *American Journal of Epidemiology* 108.4 (Oct. 1978), pp. 299–307. ISSN: 0002-9262. DOI: [10.1093/oxfordjournals.aje.a112623](https://doi.org/10.1093/oxfordjournals.aje.a112623).
-  Frank Bretz, Peter Westfall, and Torsten Hothorn. *Multiple comparisons using R*. Chapman and Hall/CRC, 2010.
-  Yu Feng et al. "Non-inheritable risk factors during pregnancy for congenital heart defects in offspring: A matched case-control study". In: *International Journal of Cardiology* 264 (Aug. 2018), pp. 45–52. ISSN: 0167-5273. DOI: [10.1016/j.ijcard.2018.04.003](https://doi.org/10.1016/j.ijcard.2018.04.003).

# Questions?



# Contact Information

✉ cai-j12@mails.tsinghua.edu.cn caijunthu@gmail.com

GitHub [github.com/caijun](https://github.com/caijun)

RSS [blog.tonytsai.name](http://blog.tonytsai.name)

To download this slides, please scan following QR code:

