# Meta-analysis of individual participant data from observational studies

Simon Thompson
University of Cambridge, UK

**Part 2**

UNIVERSITY OF CAMBRIDGE
Department of Public Health and Primary Care

IBC Victoria, July 2016

1

## **Aim:** To quantify exposure – disease associations

Challenges in meta-analysis of **published** observational studies

- Publication bias
  - since analysing associations is quick and easy, many probably do not get reported
- Searching for studies
  - not so much standard vocabulary (i.e. as for randomized trials)
- Variation in exposure measurement and outcome assessment
- Variation in analysis and reporting
- Dealing with adjustments for different covariates

2

## Advantages of individual participant data (IPD) meta-analysis

- Harmonisation of exposures and outcomes
- Updated follow-up information
- Reduction of publication and reporting biases

- Consistent analyses, e.g. adjustment for confounders
- Exploration of heterogeneity, e.g. resolve controversy
- Investigation of interactions (joint effects)
- Allowance for measurement error, using serial measurements of risk factors

- Increasingly common
- Provides reliable evidence-base

3

## Emerging Risk Factors Collaboration (ERFC)

Individual data collated from observational prospective epidemiological studies in Western populations

120 studies, 1.2m participants, 10 years average follow-up, 70,000 CVD events

Subsets available for specific risk factors:

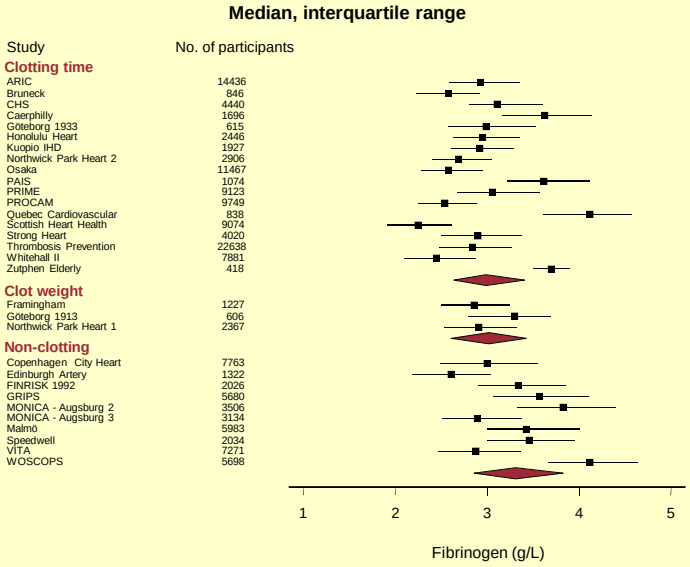
e.g. fibrinogen, related to inflammation and blood coagulation, 31 studies, 7000 CHD events

Purpose: To provide estimates of the associations of novel (or under-investigated) risk factors with CVD, which are reliable and detailed

Thompson et al, Int J Epidemiol 2010

4

## Heterogeneity of fibrinogen across studies




## Principal method of analysis used in ERFC

**Cox proportional hazards regression**

- separately in each study (s)
- stratified by sex (k=1,2)
- adjusted for entry age ($X_{si}$ for individual i in study s)

**In each study:**

$$\log(h_{ski}(t \mid E_{si}, \boldsymbol{X}_{si})) = \log(h_{0sk}(t)) + \beta_s E_{si} + \boldsymbol{\gamma}_s \boldsymbol{X}_{si}$$

for each study $s = 1,\ldots,S$, with strata $k = 1,2$, and individuals $i = 1,\ldots,n_s$ with exposure of interest $E_{si}$ and other covariates $\boldsymbol{X}_{si}$

t = time from baseline (i.e. duration)

6

## Random effects meta-analysis

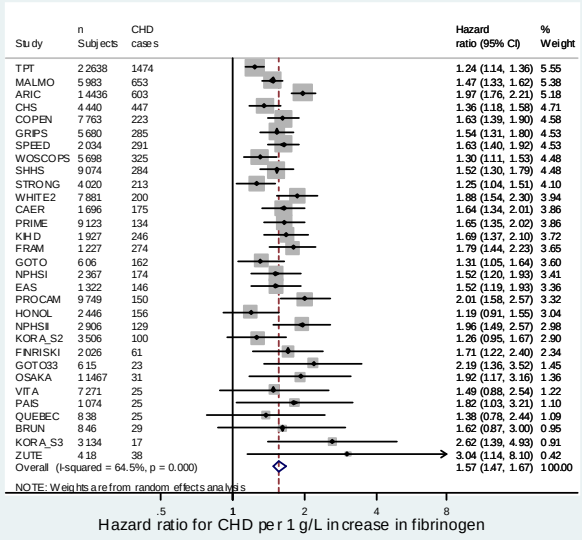
$$\hat{\beta}_s = \beta_s + \varepsilon_s; \text{ where } \varepsilon_s \sim N(0, v_s)$$
$$\beta_s = \beta + \eta_s; \text{ where } \eta_s \sim N(0, \tau^2)$$

This is a 2-stage method (e.g. using a moment estimator of $\tau^2$)

Provides a summary of within-study associations

7

## Hazard ratios of CHD per 1 g/L fibrinogen increase, adjusted for age and sex




Fibrinogen Studies Collaboration, JAMA 2005

8

## Log hazard ratios of CHD per 1 g/L increase in fibrinogen

| Meta-analysis | Log hazard ratio (SE) | $\tau$ | $I^2$ |
|---|---|---|---|
| Random effects | 0.450 (0.033) | 0.134 | 64% |

Hazard ratio from RE meta-analysis

Overall 1.57 (95% CI 1.47 to 1.67)

Prediction interval for a new study, based on $t_{s-2}$ distribution:

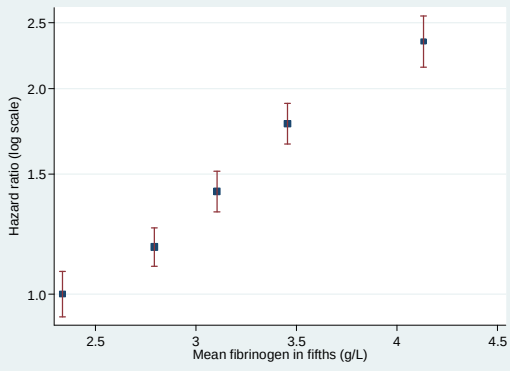
(95% range 1.18 to 2.08)

9

## Is the risk relationship linear?

Visual approach

- Divide fibrinogen into study-specific fifths
- Estimate log hazard ratios in each fifth in each study
- Combine these using multivariate RE meta-analysis
- Plot these against the mean fibrinogen level in each fifth




Analytical options

- Floating absolute risks (Plummer, Stat Med 2004)
- Quadratic term
- Fractional polynomials
- Allowance for measurement error

10

## Choice of exposure scale

Does the log hazard ratio increase:

- Linearly with fibrinogen?
- Linearly with log fibrinogen?
- Linearly with study-specific SD score fibrinogen?

| | Log hazard ratios per SD increase (SE) | $I^2$ |
|---|---|---|
| Untransformed fibrinogen | 0.294 (0.022) | 64% |
| Log fibrinogen | 0.325 (0.025) | 65% |
| Study-specific SD score fibrinogen | 0.292 (0.021) | 63% |

11

## Covariate adjustment

$$\log(h_{ski}(t \mid E_{si}, \boldsymbol{X}_{si})) = \log(h_{0sk}(t)) + \beta_s E_{si} + \boldsymbol{\gamma}_s \boldsymbol{X}_{si}$$

The confounding effects $\boldsymbol{\gamma}_s$ are different in each study (2-stage meta-analysis)

Alternatives: $\gamma_s = \gamma$ unrealistic

$\gamma_s \sim N(\gamma, \sigma^2_\gamma)$ unnecessary

12

## Overall log hazard ratios of CHD per 1 g/L increase in fibrinogen: **Adjusted for covariates**

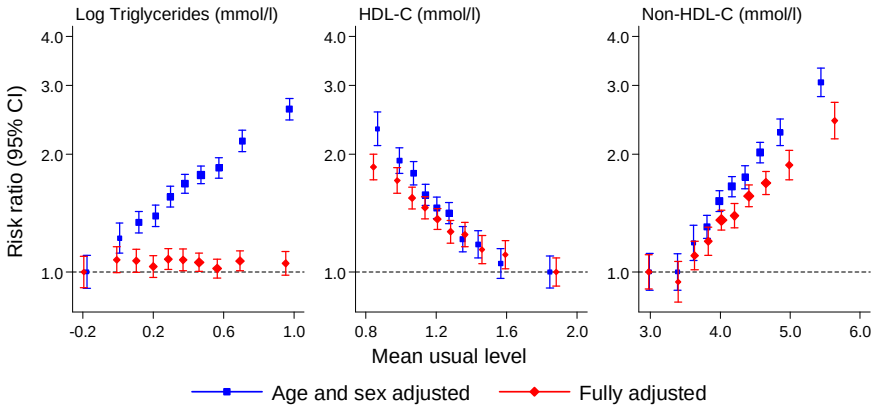
| | Adjustment | |
|---|---|---|
| | Age | Age, smoking, chol, SBP, BMI |
| Overall log HR | 0.450 | 0.320 |
| (SE) | (0.033) | (0.026) |
| $I^2$ | 64% | 35% |
| (95% CI) | (48 to 76%) | (0 to 58%) |

BUT residual confounding from other covariates remains

13

## Usefulness of consistent adjustment for confounders (Lipids)

Associations with CHD (68 studies, 302,430 participants, 12,785 CHD cases)




Fully adjusted: Age, sex, smoking, systolic BP, BMI, diabetes, log TG, HDL, non-HDL

Emerging Risk Factors Collaboration, JAMA 2009

14

## Confounders missing in some studies

Hazard ratios of CHD per 1 g/L increase in fibrinogen

| Adjustment | 31 studies (n ≈ 150,000) | 14 studies (n ≈ 72,000) |
|---|---|---|
| (1) Age | 1.57 | 1.61 |
| (2) + smok, chol, SBP, BMI | 1.38 | 1.44 |
| (3) + HDL, LDL, alcohol, TG, diabetes | – | 1.35 |

Bivariate random effects meta-analysis of (2) and (3) to 'fill in the gap':

Hazard ratio 1.30 (95% CI 1.23 to 1.36)

Estimating within-study correlations: bootstrapping; analytical approximations; record stacking.

Jackson et al, Stat Med 2009

15

## Interactions / joint effects / subgroups

Does the relationship of fibrinogen with risk:

- vary with age or level of BMI?
- depend on how fibrinogen is measured?
- differ between men and women?

16

## (1) Individual level variables

e.g. Does the association of fibrinogen (*E*) with risk vary with the level of BMI (*X*)?

$$\log(h_{ski}(t \mid E_{si}, X_{si})) = \log(h_{0sk}(t)) + \beta_s E_{si} + \gamma_s X_{si} + \delta_s E_{si} X_{si}$$

$$\delta_s = \delta_W + \eta_s; \text{ where } \eta_s \sim N(0, \tau^2)$$

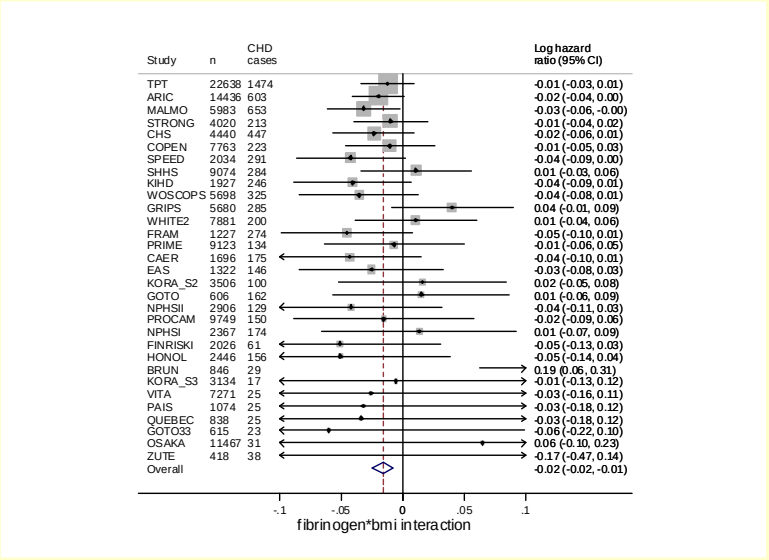
Summarises within-study information on the interaction

Expresses change in [ log hazard ratio of CHD per 1 g/L increase in fibrinogen ] per 1 kg/m$^2$ increase in BMI

Model can be extended to include other covariates, and their interactions with the exposure

17

## Within-study fibrinogen / BMI interaction




18

## (2) Study level variables

e.g. Does the association of fibrinogen (*E*) with risk vary with the assay method (*Z*)?

Comparison between studies

Limited by number of studies

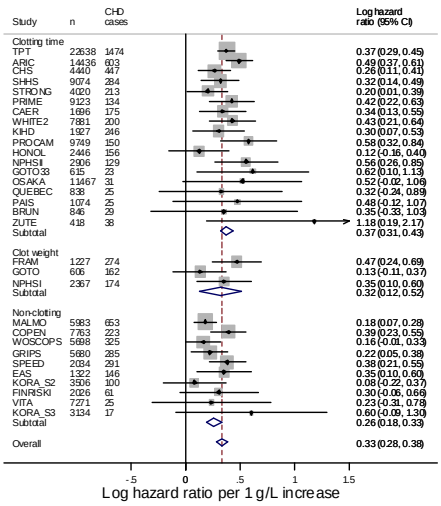
Susceptible to between-study confounding

Subgroup analysis (categorical variable Z)

Meta-regression (continuous variable Z)

19

## Study-specific log hazard ratios by assay method




20

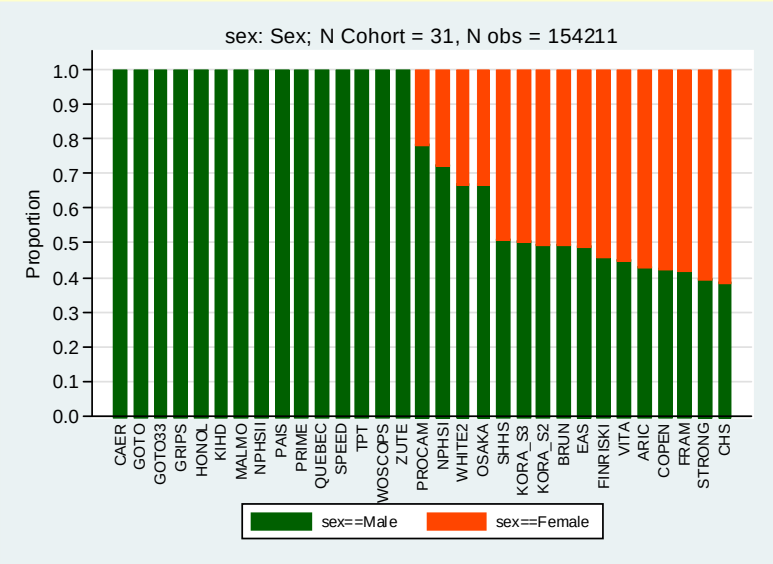
## Fibrinogen – CHD association by assay method

| *Potential effect modifier** | *n cohorts* | *n subjects* | *Estimate β (SE)* | *$\chi^2$ test $\chi^2$ (df)* | *p* |
|---|---|---|---|---|---|
| **Assay methods** | | | | **10 (2)** | **0.006** |
| **Clotting time** | **18** | **105594** | **0.373 (0.030)** | | |
| **Clot weight** | **3** | **4200** | **0.319 (0.102)** | | |
| **Non-clotting** | **10** | **44417** | **0.257 (0.037)** | | |

***Adjusted for age, smoking, total cholesterol, systolic BP, and body mass index**

21

## (3) Mixed variables




22

## Interaction with sex

(a) Within-study information ($\delta_W$)

Pooling interaction terms as before

All-male / all-female studies do not contribute

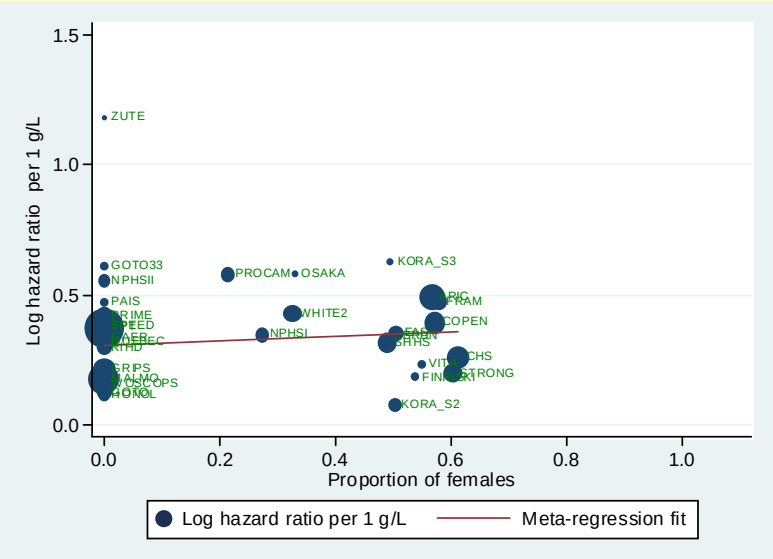
(b) Between-study information ($\delta_B$)

Meta-regression on proportion of women in each study

All studies contribute

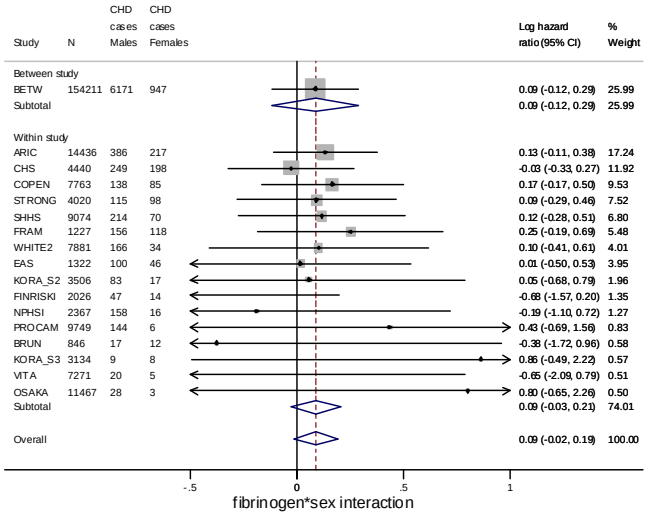
23

## Between-study fibrinogen / sex interaction




24

## Fibrinogen / sex interaction




25

## Some comments

$\delta_B$ is susceptible to between-study confounding / ecological bias

$\delta_W$ can be susceptible to regression dilution bias, through measurement error

Usually ignore between-study information when there is lots of within-study information

Inspecting weight given to information within- and between-studies is useful

26

## Assessing proportional hazards

Non-proportional hazards can be considered simply as an interaction with time, based on within-study information alone.

Each study estimates this interaction term, a non-PH parameter:

$$\log(h_{ski}(t \mid E_{si}, X_{si})) = \log h_{0sk}(t) + \beta_s E_{si} + \xi_s t E_{si}$$

**Fibrinogen and CHD risk**

Random effects meta-analysis of study-specific non-PH parameters:
Estimate 0.0016 per year (SE 0.0045), chi-squared = 0.12 (df 1)

27

## Computational methods

See: www.phpc.cam.ac.uk/ceu/research/erfc/stata/

Relevant Stata programs (author: Stephen Kaptoge) are available to install within a Stata session by typing:

net from http://ceu.phpc.cam.ac.uk/software/erfc/

net describe <package>

net install <package>

28

## Summary

IPD has enabled:

- Consistent method of analysis across studies
- Consistent covariate adjustment
- Assessment of interactions

Further topics (Part 4)

Adjusting for measurement error

Deriving measures of public health impact

Are risk factor associations causal?

29

## References


Thompson SG et al. Statistical methods for the time-to-event analysis of individual participant data from multiple epidemiological studies. *Int J Epidemiol* 2010; **39**: 1345-1359.

Jackson D et al. Systematically missing confounders in individual participant data meta-analysis of observational cohort studies. *Stat Med* 2009; **28**: 1218-1237.

Jackson D, White IR, Thompson SG. Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. *Stat Med* 2010; **29**: 1282-1297.

Fibrinogen Studies Collaboration. Plasma fibrinogen and the risk of major cardiovascular diseases and nonvascular mortality: an individual participant meta-analysis. *JAMA* 2005; **294**: 1799-1809.

Emerging Risk Factors Collaboration. Major lipids, apolipoproteins and risk of vascular disease. *JAMA* 2009; **302**: 1993-2000.


30