Meta-analysis of individual participant data from observational studies

Simon Thompson

University of Cambridge, UK

3. Risk prediction metrics

Freiburg, March 2013

Purpose

- Combine information on predictive ability of risk models across studies
- Assess how much a new risk marker aids risk prediction

Rationale

- Greater precision
- Reduced over-fitting
- Increased generalizability

Risk prediction models for multiple prospective studies

Cox proportional hazards model, stratified by study (*s*) and sex (*k*)

$$\log(h_{ski}(t \mid \boldsymbol{X})) = \log(h_{0sk}(t)) + \sum_p \beta_p \boldsymbol{X}_{psi}$$

1-stage approach; ignores heterogeneity of β_p across studies

– focus is on estimates rather than their uncertainty

– easier to get absolute risk predictions than from a 2-stage random effects approach

Risk score = $\sum_p \hat{\beta}_p \boldsymbol{X}_{psi}$ for individual *i* in study *s*

Risk discrimination in a single prospective study (i)

Harrell's C-index = probability that the predicted order of events is correct in a randomly selected pair of participants

$C = n_c / (n_c+n_d)$ where

n_c = # concordant pairs; n_d = # discordant pairs

C ranges from 0.5 to 1.0

The risk score is sufficient when ranking predicted risk

ΔC = change in C on adding a new risk marker

Variance of C and ΔC calculated by efficient jackknife estimator

Harrell, Stat Med 1996

Risk discrimination in a single prospective study (ii)

Royston-Sauerbrei's D statistic
- – quantifies the association between ranked predicted risk and observed risk
- – interpreted as log hazard ratio comparing top to bottom halves of predicted risks

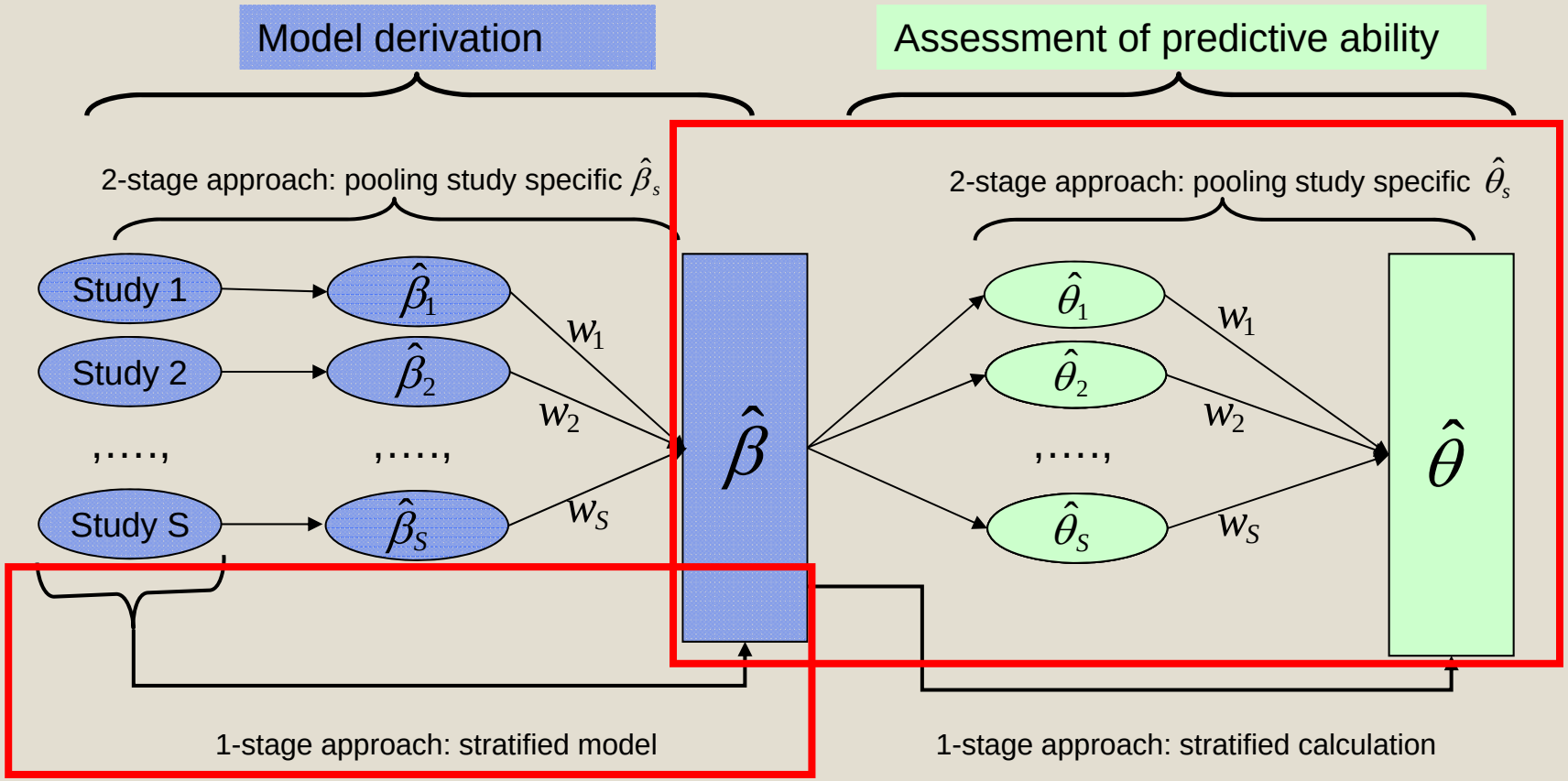
D ranges from 0 upwards

ΔD = change in D on adding a new risk marker
Variance of D obtained directly
Variance of ΔD obtained by bootstrapping

Stratified versions of C-index and D can be calculated
- – e.g. stratifying by sex

Royston & Sauerbrei, Stat Med 2004

Overall scheme for modelling and assessment



w_i = study-specific weights for meta-analysis

$\hat{\theta}_s$ = C, D, ΔC, or ΔD

Primary ERFC prediction analyses

Choice of weights in meta-analysis (1 & 2)

1. Inverse-variance (fixed effect)
2. Inverse-variance (random effects)

Meta-analysis of C-index and D

C_1 for model 1, C_2 for model 2, ΔC for difference

D_1 for model 1, D_2 for model 2, ΔD for difference

Inverse-variance weights leads to inconsistency between

pooled estimates of $C_2–C_1$ and ΔC

pooled estimates of $D_2–D_1$ and ΔD

Choice of weights in meta-analysis (3 & 4)

3. Number of events

Well matched to intrinsic weighting in stratified Cox PH model

4. Estimators stratified by study

For D, equivalent to option 3
For C, weights depend on number of informative pairs – large studies can receive substantial weight despite few events

Prefer option 3

Example

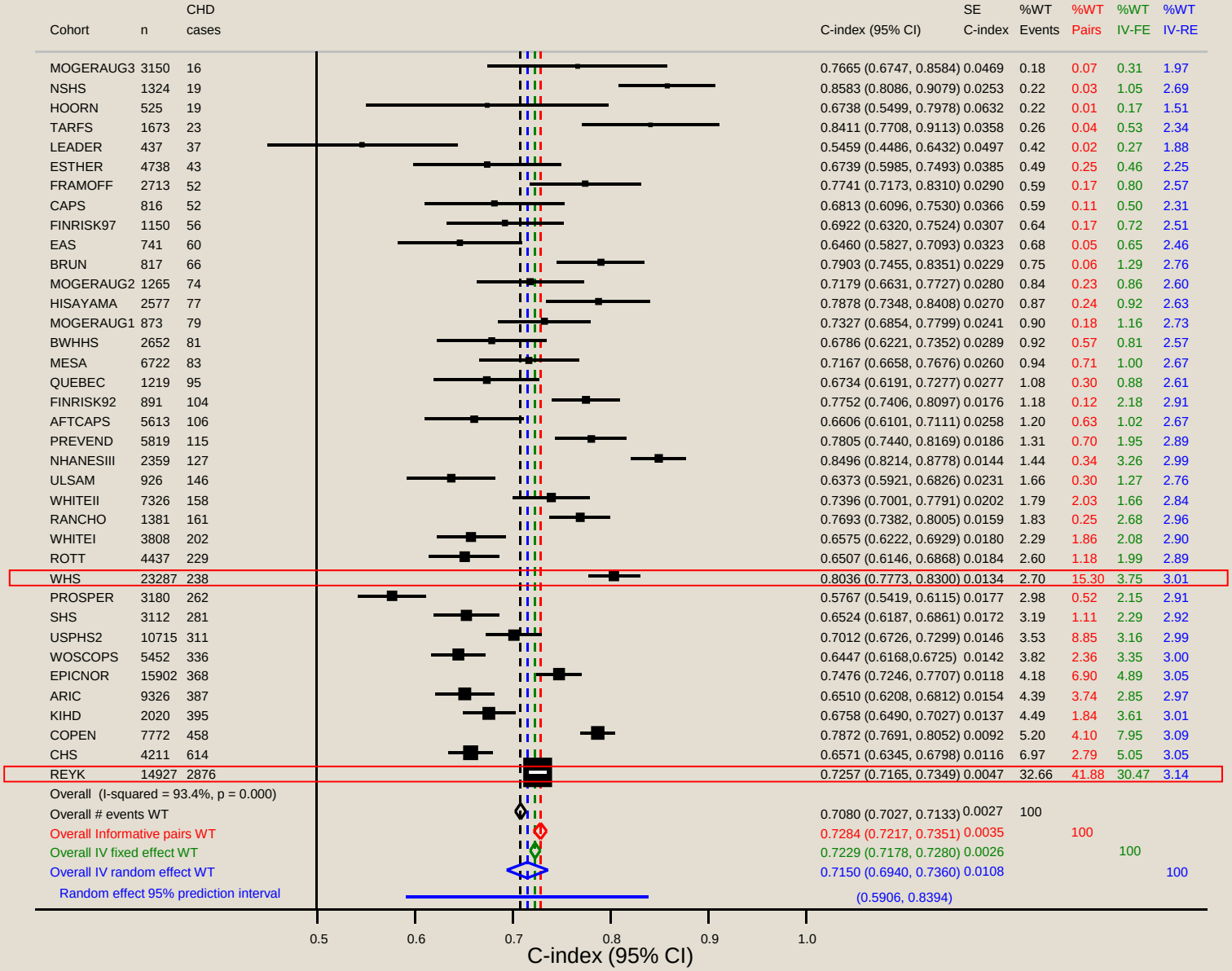
Prediction of CHD (fatal and non-fatal)

37 population-based prospective studies, 8806 CHD cases

Conventional risk factors: age, sex, smoking, systolic BP, diabetes, total chol, HDL chol

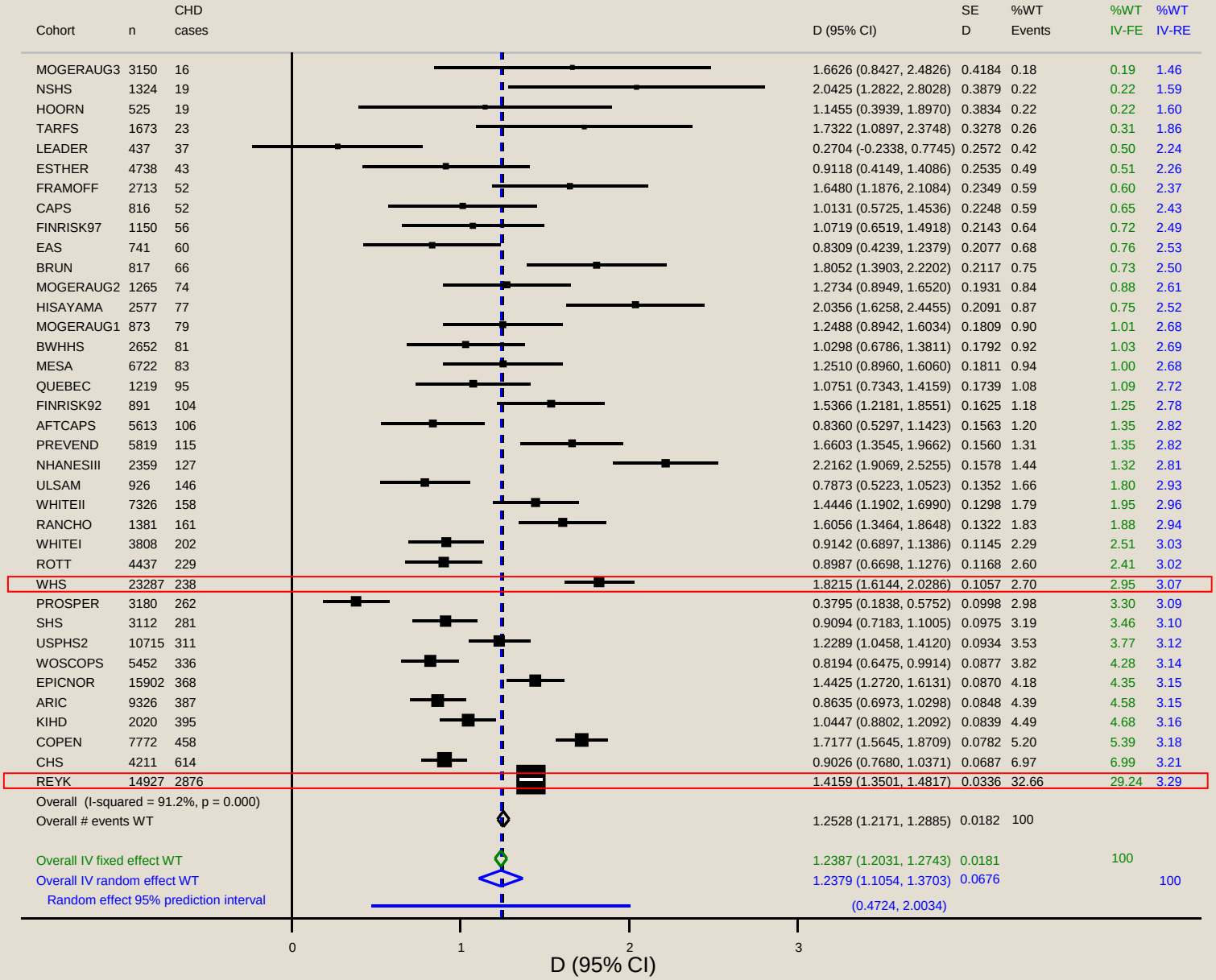
Investigate value of adding of log C-reactive protein (CRP), an inflammatory marker

Meta-analysis of C-index



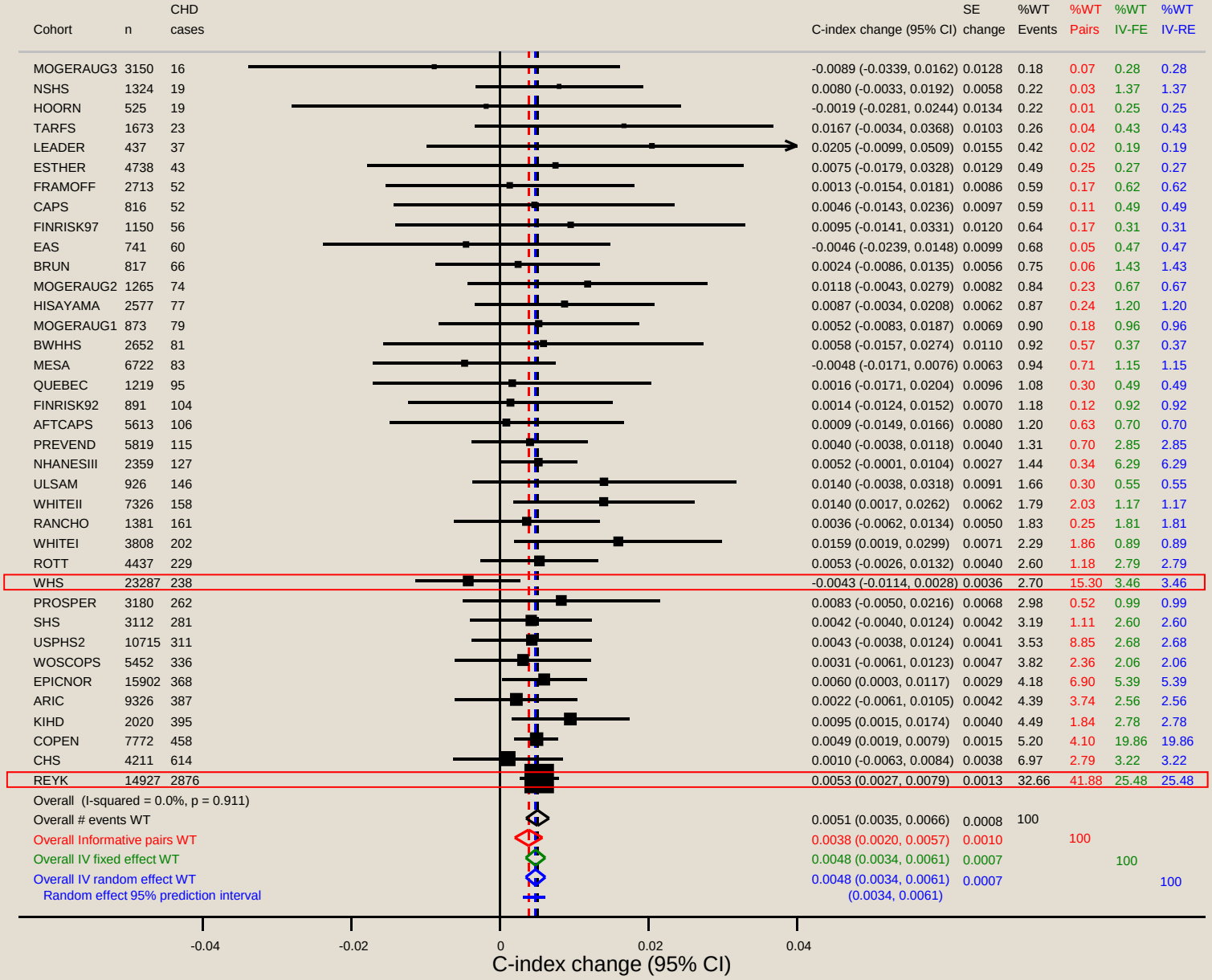
C-index with conventional CHD risk factors: age (sex), smoking, systolic BP, diabetes, total chol, HDL chol

Meta-analysis of D



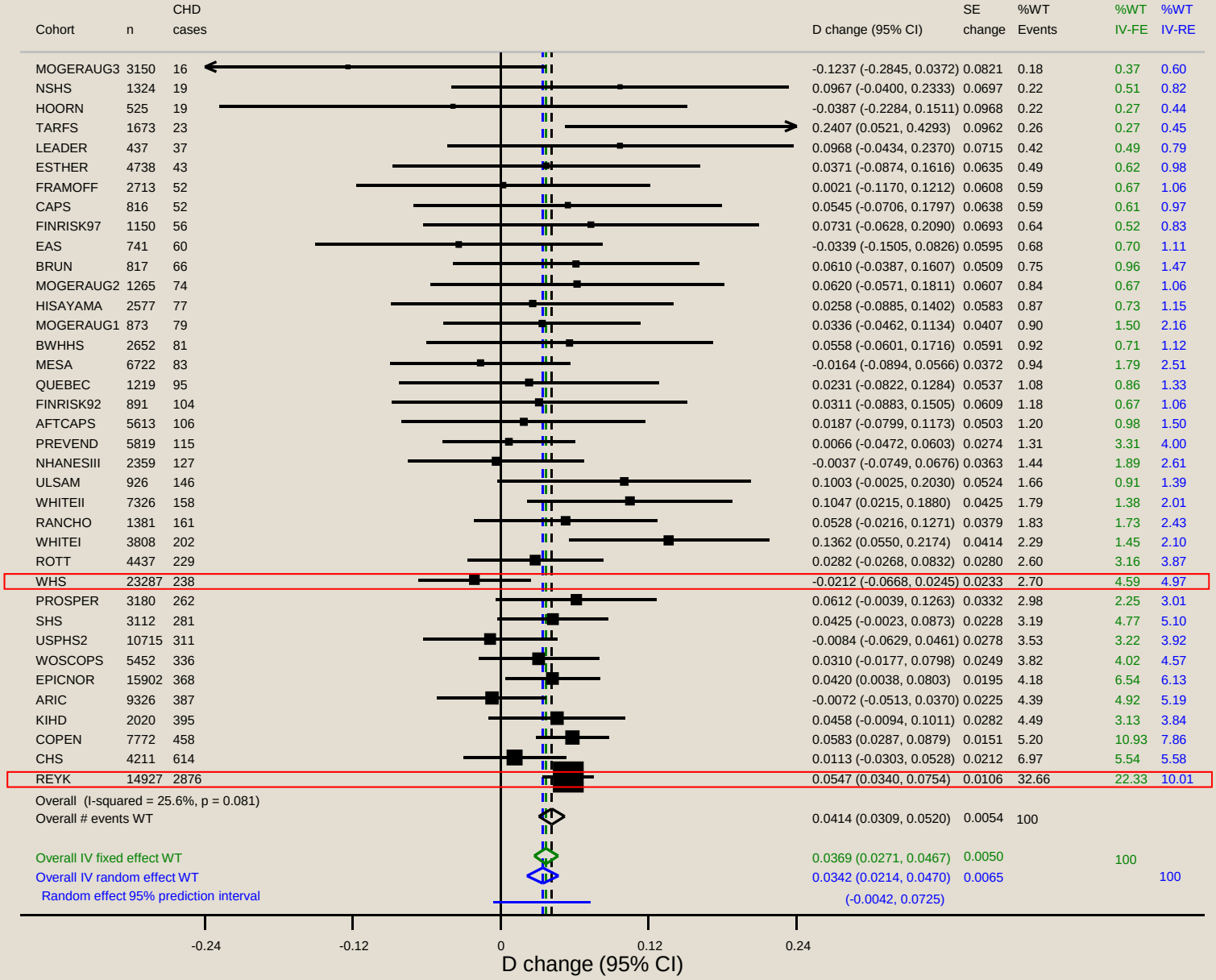
D with conventional CHD risk factors: age (sex), smoking, systolic BP, diabetes, total chol, HDL chol

Meta-analysis of C-index change



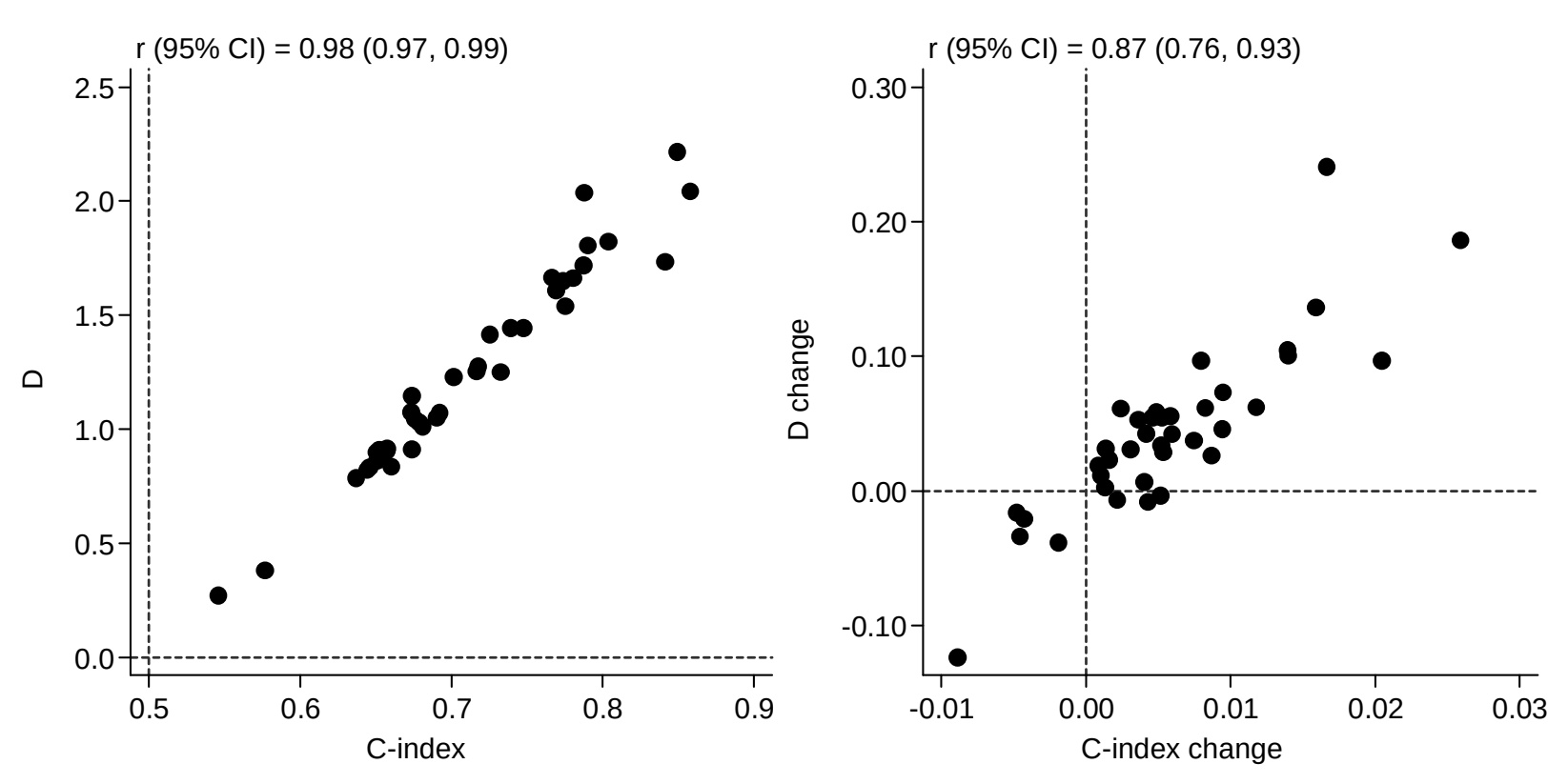
C-index change with addition of log CRP

Meta-analysis of D change

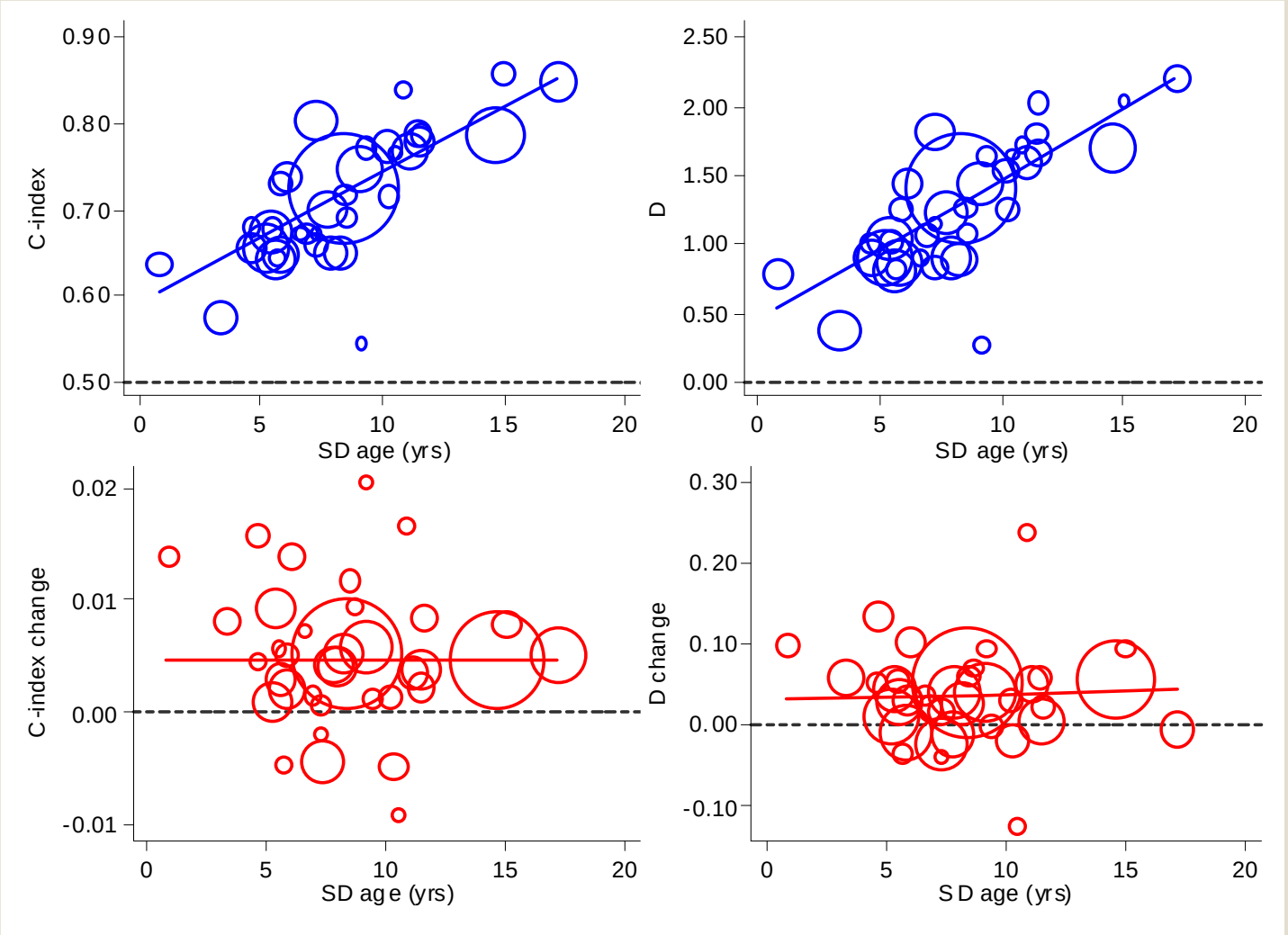


D change with addition of log CRP

Study-level relationship between C-index & D, ΔC & ΔD



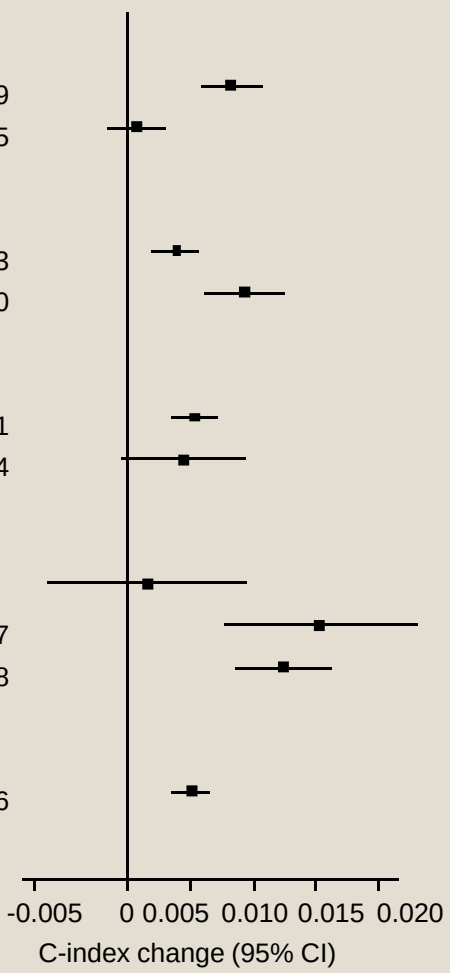
Between-study heterogeneity in absolute values of C-index or D is largely explained by age distribution

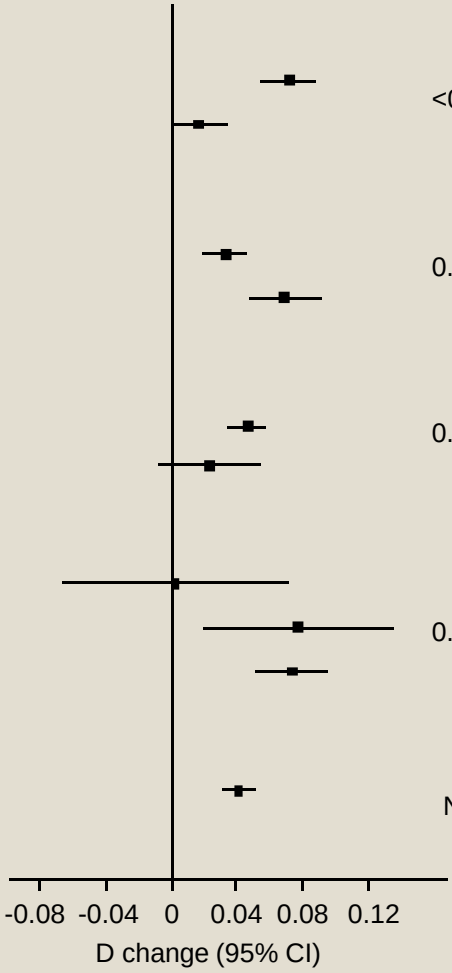


Size of circle is proportional to 1/variance of estimate

Examination of subgroup effects – restricted to studies with all levels observed

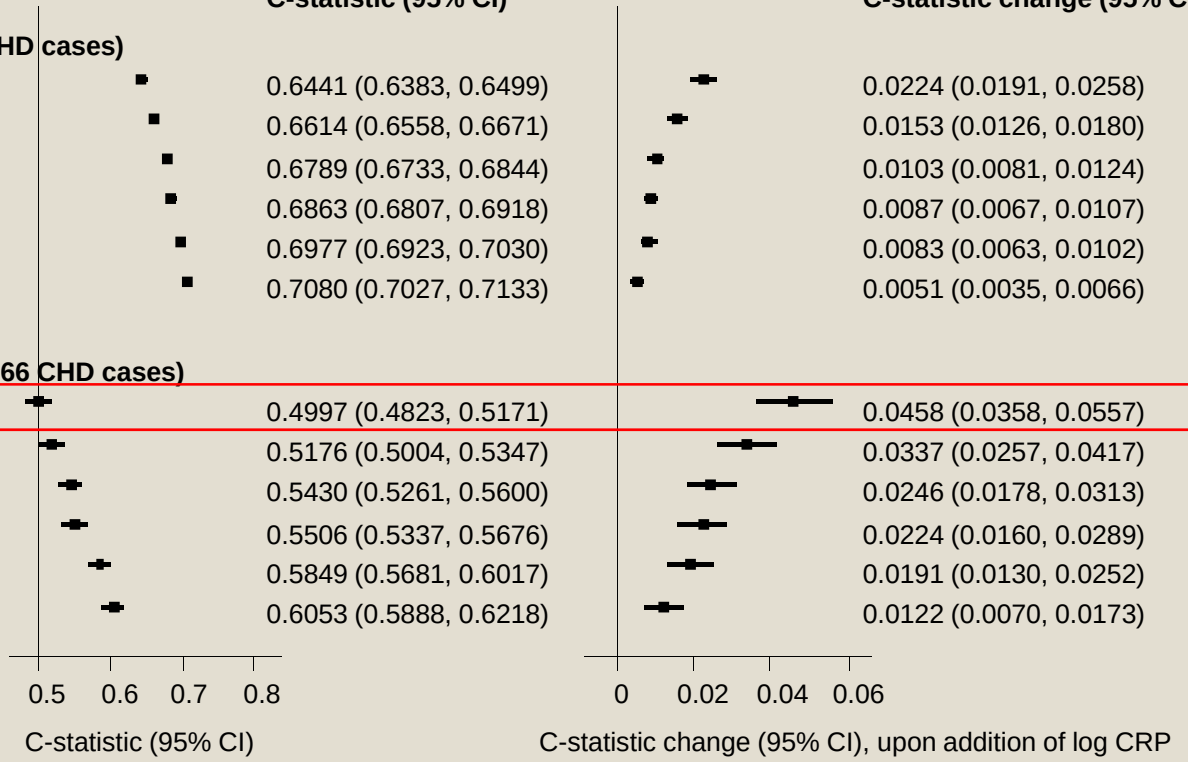
Variable\ Subgroup	No. of studies	No. of participants	CHD cases	P-value for heterogeneity	P-value for heterogeneity
Sex					
Male	24	53037	4189	<0.0001	<0.0001
Female	24	58199	2515		
Smoking status					
Other	36	128676	5403	0.004	0.004
Current	36	35507	3380		
History of diabetes					
No	36	144353	7491	0.752	0.173
Yes	36	10788	1004		
Framingham 2008 10-yr CVD risk					
0%-<10%	30	62715	932	0.027	0.147
10%-<20%	30	47161	2527		
>=20%	30	47069	4808		
Overall	37	165856	8806	NA	NA





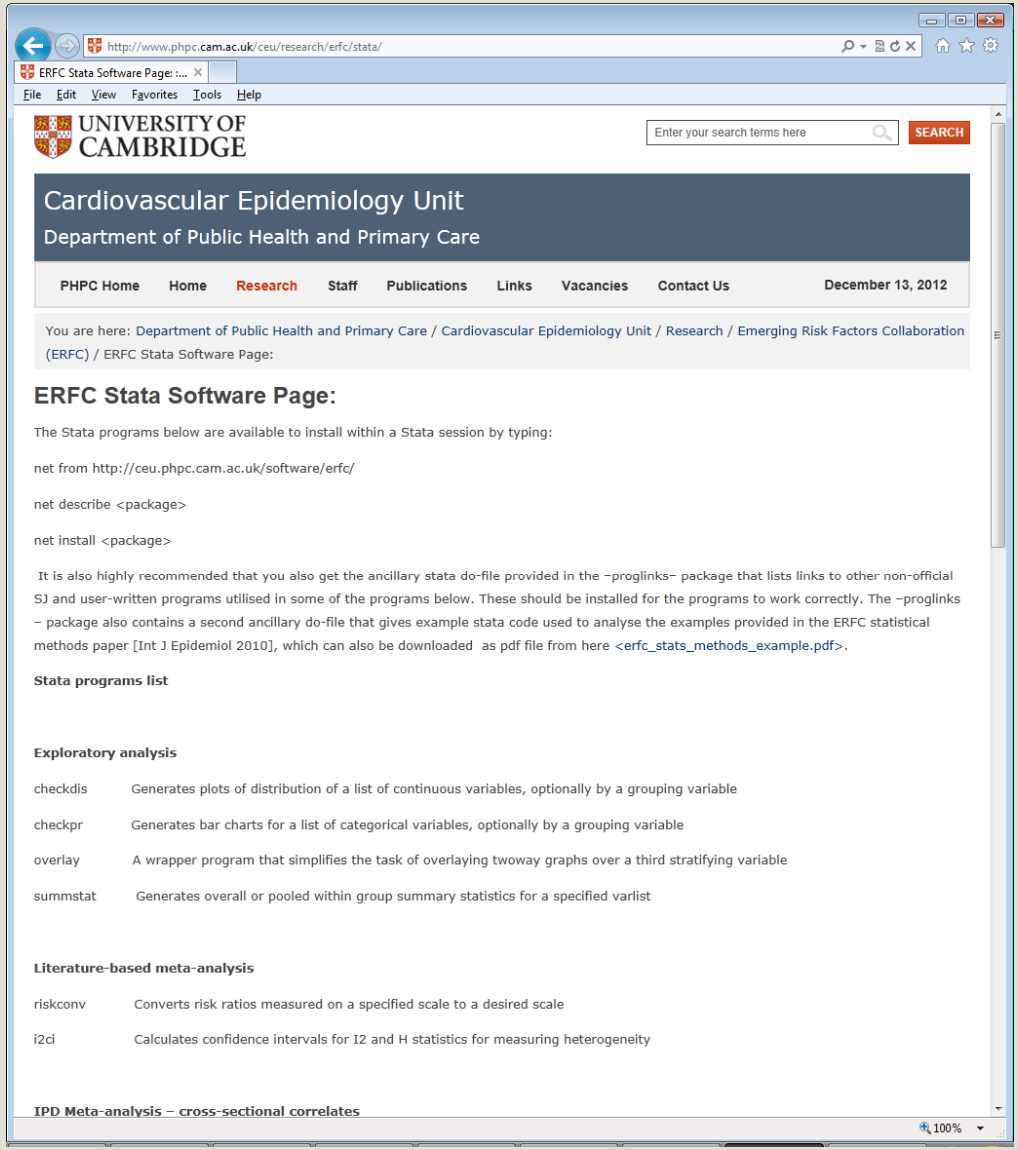
Case-control studies are particularly problematic because of matching on predictive covariates

Study design \ Base model variables	C-statistic (95% CI)	C-statistic change (95% CI)	Z	P-value
Cohort (37 studies, 8806 CHD cases)				
Age	0.6441 (0.6383, 0.6499)	0.0224 (0.0191, 0.0258)	13.2	<0.0001
Above + smoking status	0.6614 (0.6558, 0.6671)	0.0153 (0.0126, 0.0180)	11.1	<0.0001
Above + systolic BP	0.6789 (0.6733, 0.6844)	0.0103 (0.0081, 0.0124)	9.2	<0.0001
Above + history of diabetes	0.6863 (0.6807, 0.6918)	0.0087 (0.0067, 0.0107)	8.5	<0.0001
Above + total cholesterol	0.6977 (0.6923, 0.7030)	0.0083 (0.0063, 0.0102)	8.4	<0.0001
Above + HDL cholesterol	0.7080 (0.7027, 0.7133)	0.0051 (0.0035, 0.0066)	6.3	<0.0001
Case-control (8 studies, 1566 CHD cases)				
Age	0.4997 (0.4823, 0.5171)	0.0458 (0.0358, 0.0557)	9.0	<0.0001
Above + smoking status	0.5176 (0.5004, 0.5347)	0.0337 (0.0257, 0.0417)	8.2	<0.0001
Above + systolic BP	0.5430 (0.5261, 0.5600)	0.0246 (0.0178, 0.0313)	7.1	<0.0001
Above + history of diabetes	0.5506 (0.5337, 0.5676)	0.0224 (0.0160, 0.0289)	6.8	<0.0001
Above + total cholesterol	0.5849 (0.5681, 0.6017)	0.0191 (0.0130, 0.0252)	6.1	<0.0001
Above + HDL cholesterol	0.6053 (0.5888, 0.6218)	0.0122 (0.0070, 0.0173)	4.6	<0.0001



Stata software for IPD meta-analysis

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



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

ERFC Stata Software Page: :...

File Edit View Favorites Tools Help

UNIVERSITY OF CAMBRIDGE

Enter your search terms here SEARCH

Cardiovascular Epidemiology Unit

Department of Public Health and Primary Care

PHPC Home | Home | Research | Staff | Publications | Links | Vacancies | Contact Us | December 13, 2012

You are here: Department of Public Health and Primary Care / Cardiovascular Epidemiology Unit / Research / Emerging Risk Factors Collaboration (ERFC) / ERFC Stata Software Page:

ERFC Stata Software Page:

The Stata programs below 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>

It is also highly recommended that you also get the ancillary stata do-file provided in the –proglinks– package that lists links to other non-official SJ and user-written programs utilised in some of the programs below. These should be installed for the programs to work correctly. The –proglinks – package also contains a second ancillary do-file that gives example stata code used to analyse the examples provided in the ERFC statistical methods paper [Int J Epidemiol 2010], which can also be downloaded as pdf file from here <erfc_stats_methods_example.pdf>.

Stata programs list

Exploratory analysis

checkdis	Generates plots of distribution of a list of continuous variables, optionally by a grouping variable
checkpr	Generates bar charts for a list of categorical variables, optionally by a grouping variable
overlay	A wrapper program that simplifies the task of overlaying twoway graphs over a third stratifying variable
summstat	Generates overall or pooled within group summary statistics for a specified varlist

Literature-based meta-analysis

riskconv	Converts risk ratios measured on a specified scale to a desired scale
i2ci	Calculates confidence intervals for I2 and H statistics for measuring heterogeneity

IPD Meta-analysis – cross-sectional correlates

100%