Systematic reviews of prognosis studies III

Meta-analytical approaches in systematic reviews of prognostic studies

Thomas Debray, Johanna Damen, Karel Moons

for the Cochrane Prognosis Methods Group

Conflict of interest

We have no actual or potential conflict of interest in relation to this presentation

Prediction

- Risk prediction = foreseeing / foretelling
  … (probability) of something that is yet unknown

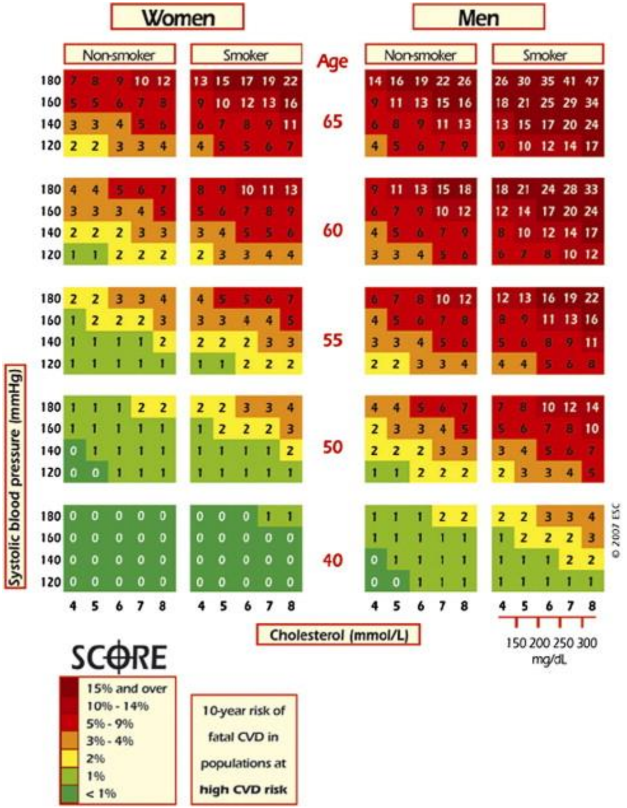
- Turn available information (predictors) into a statement about the probability:
  … diagnosis
  … prognosis

What is the big difference between diagnostic and prognostic 'prediction'?

Prediction models

APGAR
Test Scoring

	Score 0	Score 1	Score 2
Appearance	Blue all over	Blue only at extremities	No blue coloration
Pulse	No pulse	<100 beats/min.	>100 beats/min.
Grimace	No response to stimulation	Grimace or feeble cry when stimulated	Sneezing, coughing, or pulling away when stimulated
Activity	No movement	Some movement	Active movement
Respiration	No breathing	Weak, slow, or irregular breathing	Strong cry



Three phases of Prediction Modelling

1. Developing a prediction model
2. Validate (+update) the model in other subjects
3. Quantify model’s impact on doctor’s decision making and patient outcome (cost-effectiveness)

What is big difference between 3 versus 1-2?

Focus on 1-2

Two types of prediction modes: diagnostic and prognostic

Reviews of prognosis studies

Focus on MA of prognostic prediction models

Everything also applies to MA of diagnostic prediction models

Numerous prognostic models for same target population + outcomes

- >350 models for predicting cardiovascular disease
- >100 models for brain trauma patients
- >100 diabetes type 2 models
- > 60 models for breast cancer prognosis

Need for systematic reviews

Abundance of CPMs, with poor understanding of

- The comparative performance of these CPMs
- The consistency of accuracy and predictions across CPMs
- The clinical impact of these CPMs

Systematic review and MA validation studies of one or more certain models may help to identify promising models and evaluate the need for further improvements of these models.

Why do we need meta-analysis?

Quantitative synthesis (meta-analysis) may help

- To summarize the predictive performance of a certain CPM across multiple validation studies
- To evaluate whether a certain CPM yields consistently good performance across different populations, outcomes, etc.
- To establish boundaries of applicability and generalizability
- To identify possible improvements of CPMs

Is MA even possible?

You need multiple validation studies of same model!

Ex. Prognostic prediction models for cardiovascular disease

Top 5 validated models	N
Framingham (Wilson 1998)	80
Framingham (Anderson 1991 Am H J)	73
SCORE (Conroy 2003)	63
Framingham (D'Agostino 2008)	44
Framingham (no reference)	32

Is MA of prediction models even possible?

- Model validation studies are increasingly common!
  E.g. Framingham, EuroSCORE, Gail, …

- Reporting of model validation studies is steadily improving!
  E.g. due to reporting guidelines (TRIPOD)

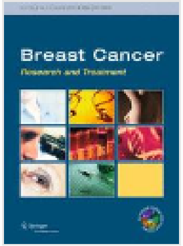
Annals of Internal Medicine RESEARCH AND REPORTING METHODS

Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration

Ann Intern Med. 2015;162:W1-W73. doi:10.7326/M14-0698

Karel G.M. Moons, PhD; Douglas G. Altman, DSc; Johannes B. Reitsma, MD, PhD; John P.A. Ioannidis, MD, DSc;
Petra Macaskill, PhD; Ewout W. Steyerberg, PhD; Andrew J. Vickers, PhD; David F. Ransohoff, MD; and Gary S. Collins, PhD

Is MA even possible?



Breast Cancer Research and Treatment
April 2012, Volume 132, Issue 2, pp 365–377

A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance

Authors | Authors and affiliations

Catherine Meads, Ikhlaaq Ahmed, Richard D. Riley

Review

First Online: 22 October 2011

DOI: 10.1007/s10549-011-1818-2

Cite this article as:
Meads, C., Ahmed, I. & Riley, R.D. Breast Cancer Res Treat (2012) 132: 365. doi:10.1007/s10549-011-1818-2

Is MA even possible?

Go to old article view

PDF

Info

journal of thrombosis and haemostasis™

ISTH International Society on Thrombosis and Haemostasis

Explore this journal >

Original Article - Cardiovascular Medicine

Predictive performance of the CHA2DS2-VASc rule in atrial fibrillation: a systematic review and meta-analysis

Sander van Doorn, Thomas P.A. Debray, Femke Kaasenbrood, Arno W. Hoes, Frans H. Rutten, Karel G.M. Moons, Geert-Jan Geersing

Accepted manuscript online: 4 April 2017 Full publication history

DOI: 10.1111/jth.13690 View/save citation

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/jth.13690

Accepted Articles

Browse Accepted Articles

Accepted, unedited articles published online and citable. The final edited and typeset version of record will appear in future.

Guidance papers

RESEARCH METHODS AND REPORTING

CrossMark
click for updates

A guide to systematic review and meta-analysis of prediction model performance

Thomas P A Debray,[1,2] Johanna A A G Damen,[1,2] Kym I E Snell,[3] Joie Ensor,[3] Lotty Hooft,[1,2]
Johannes B Reit

Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–19

Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0962280218785504
journals.sagepub.com/home/smm
SAGE

A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes

Thomas PA Debray,[1,2] Johanna AAG Damen,[1,2]
Richard D Riley,[3] Kym Snell,[3] Johannes B Reitsma,[1,2]
Lotty Hooft,[1,2] Gary S Collins[4] and Karel GM Moons[1,2]

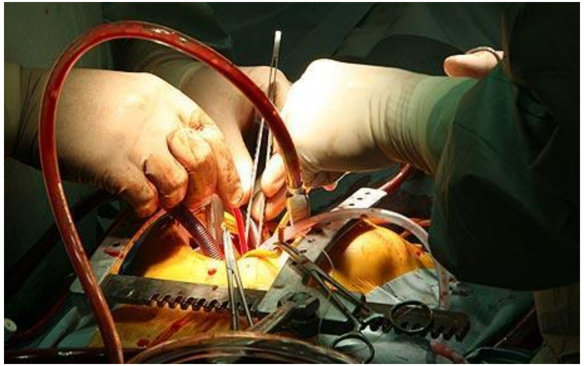
Required steps of the SR

1. Formulating the review question
2. Formulating the search strategy
3. Critical appraissal (CHARMS & PROBAST)
4. Quantitative data extraction
5. Meta-analysis
6. Investigating heterogeneity across studies
7. Sensitivity analyses
8. Reporting

Illustrative example: EuroSCORE

Predicting 30 day mortality after cardiac surgery

- Cardiac surgery in high-risk population
- Need for risk stratification
- Establish risk profile of cardiac surgical patients using multivariable prediction models
- Establish prediction model performance



Illustrative example: EuroSCORE

Patient related factors		
Age [1] (years)	0	0
Gender	select	0
Renal impairment [2] *See calculator below for creatinine clearance*	normal (CC >85ml/min)	0
Extracardiac arteriopathy [3]	no	0
Poor mobility [4]	no	0
Previous cardiac surgery	no	0
Chronic lung disease [5]	no	0
Active endocarditis [6]	no	0
Critical preoperative state [7]	no	0
Diabetes on insulin	no	0

Cardiac related factors		
NYHA	select	0
CCS class 4 angina [8]	no	0
LV function	select	0
Recent MI [9]	no	0
Pulmonary hypertension [10]	no	0

Operation related factors		
Urgency [11]	elective	0
Weight of the intervention [12]	isolated CABG	0
Surgery on thoracic aorta	no	0

EuroSCORE II ***Euro*SCORE II** 0

Note: This is the 2011 EuroSCORE II

Calculate Clear

Step 1

Formulating the review question and protocol

Formulating the review question and protocol

- Describe rationale, objectives, design, methodology and statistical considerations of the systematic review
- Define the PICOTS

Extensively discussed in workshop 1!

Illustrative example: EuroSCORE

Population	Patients undergoing coronary artery bypass grafting
Intervention	The (additive) EuroSCORE model
Comparator	Not applicable
Outcome(s)	All cause mortality
Timing	30 days, predicted using peri-operative conditions
Setting	risk stratification in the assessment of cardiac surgical results

Step 2

Formulating the search strategy

Formulating the search strategy

- Use information from the PICOTS
- Combine with existing search filters
- Evaluate citations of the development paper

Tools: electronic databases, conference abstracts, hand searching, online registers

Extensively discussed in workshop 1!

Step 3
Critical appraisal

UMC Utrecht

Critical appraisal

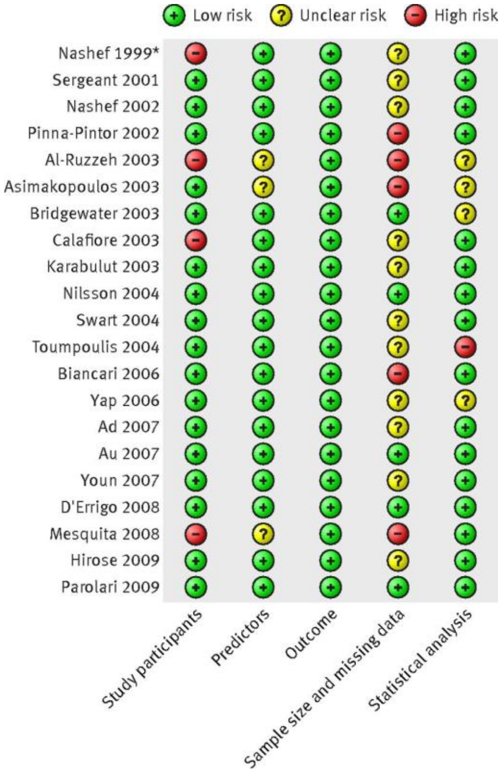
Evaluate **bias and applicability** of each validation study

- CHARMS checklist
- PROBAST (2018)

Decide whether studies should be excluded due to low quality and/or applicability with respect to the current review

Extensively discussed in workshop 2!

Illustrative example: EuroSCORE



Overall judgment for risk of bias of included articles

(21 studies, involving 22 validations)

Step 4

Quantitative data extraction and preparation

Recap: what are validation studies?

- Test a previously developed prediction model into new individuals
  - Same population
  - Different but related population

- Evaluate the predictive accuracy
  - Overall performance
  - Calibration
  - Discrimination

Recap: what are validation studies?

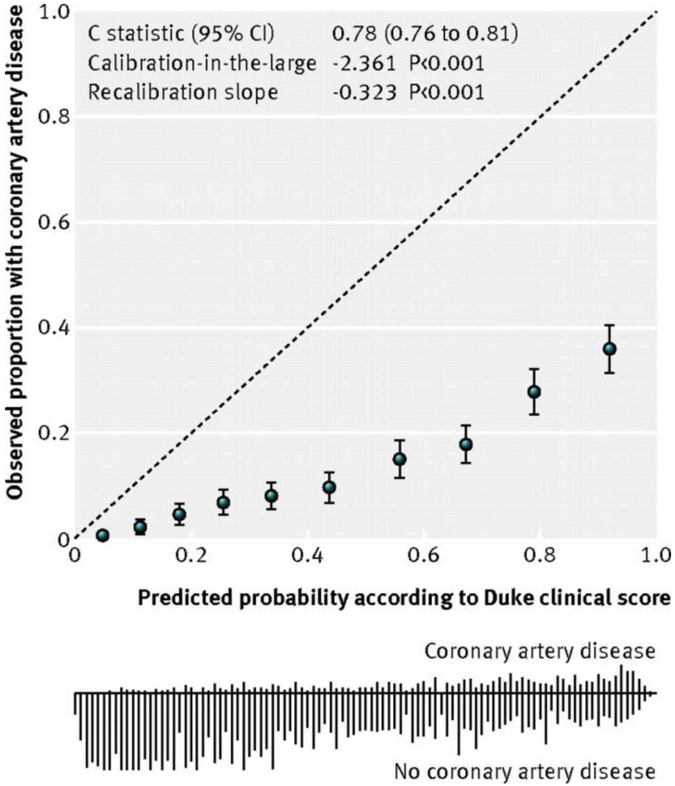
What statistics can we summarize when reviewing external validation studies?

Discrimination

Quantifies the model's extent to distinguish between events and non-events

- Visual inspection
  - Receiving Operating Characteristics (ROC) curve
- Summary statistics
  - Concordance (c) index
  - Area under the ROC curve (AUC)
  - Discrimination slope

Calibration



Agreement between observed outcomes and predictions

- Total O:E ratio
- Calibration intercept
- Calibration slope

Calibration table – good model?

External validation of EuroSCORE

Expected mortality (%) versus observed in-hospital mortality

Score	N	Expected	Observed
0-2	201	1.4	0.5
3-5	309	4.0	1.0
6-8	181	6.8	2.2
>= 9	66	10.5	3.0

Quantitative data extraction and preparation

Common problems in data extraction

- Selective/inconsistent reporting
- Incomplete assessments (e.g. calibration)
- Missing estimates of precision (e.g. standard error)

Solutions

- C-statistic, O:E ratio and calibration slope can often be derived from reported information
- Several approximations have been proposed to obtain estimates for missing standard errors

Quantitative data extraction and preparation

metamisc: Diagnostic and Prognostic Meta-Analysis

Meta-analysis of diagnostic and prognostic modeling studies. Summarize estimates of prognostic factors, diagnostic test accuracy and prediction model performance. Validate, update and combine published prediction models. Develop new prediction models with data from multiple studies.

Version:	0.1.9
Depends:	R (≥ 3.2.0), stats, graphics
Imports:	metafor (≥ 2.0.0), mvtnorm, ellipse, lme4, plyr, ggplot2
Suggests:	runjags, rjags, testthat (≥ 1.0.2)
Published:	2018-05-13
Author:	Thomas Debray [aut, cre], Valentijn de Jong [aut]
Maintainer:	Thomas Debray <thomas.debray at gmail.com>
License:	GPL-3
URL:	http://r-forge.r-project.org/projects/metamisc/
NeedsCompilation:	no
In views:	MetaAnalysis
CRAN checks:	metamisc results

Downloads:

Reference manual:	metamisc.pdf
Package source:	metamisc_0.1.9.tar.gz
Windows binaries:	r-devel: metamisc_0.1.9.zip, r-release: metamisc_0.1.9.zip, r-oldrel: metamisc_0.1.9.zip
OS X binaries:	r-release: metamisc_0.1.9.tgz, r-oldrel: metamisc_0.1.9.tgz
Old sources:	metamisc archive

Linking:

Please use the canonical form `https://CRAN.R-project.org/package=metamisc` to link to this page.

Quantitative data extraction and preparation

- Information on case-mix variation
  - Mean & standard deviation of key subject characteristics
  - Mean & standard deviation of the linear predictor

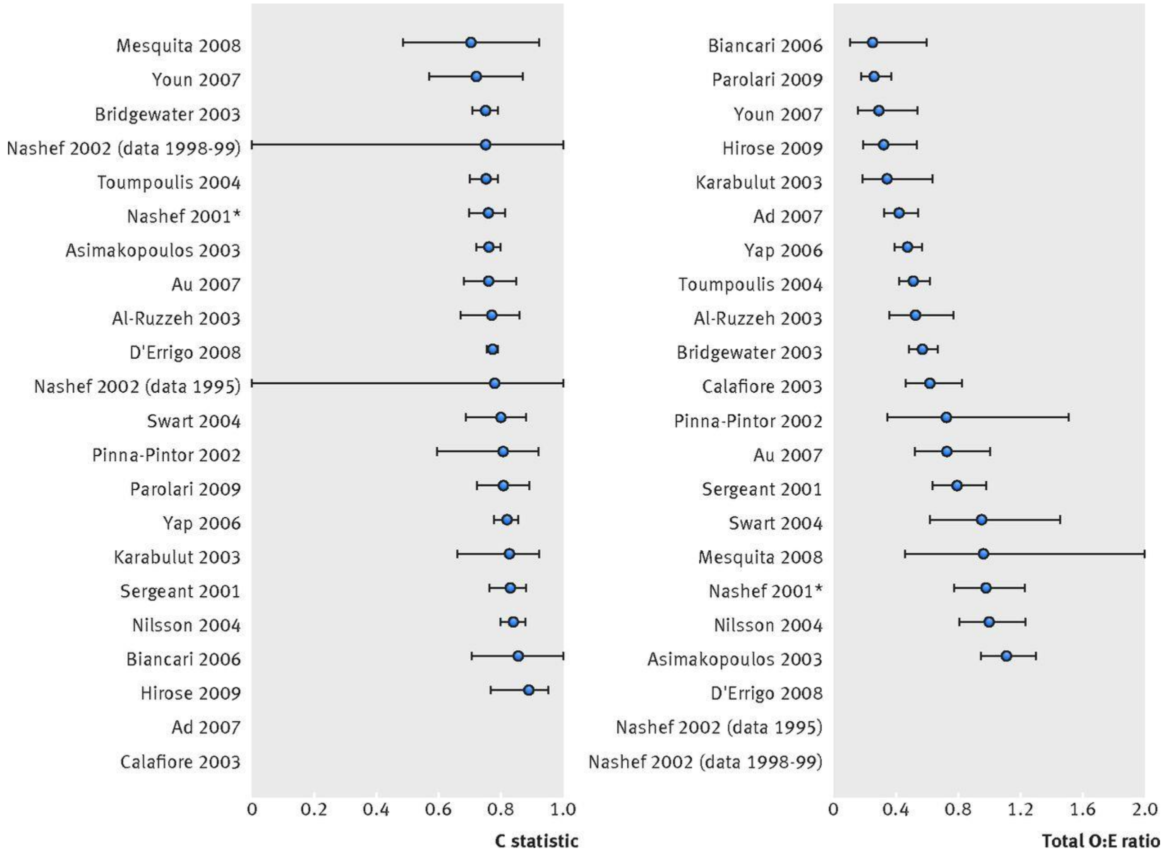
- Information on key study characteristics
  - Location
  - Standards w.r.t. treatments, patient referral, …

Illustrative example: EuroSCORE

Predictive performance of the EuroSCORE

- C-statistic
  - Summary statistic reported in 20 validations
  - SE approximated for 7 studies
- O:E
  - Relevant information obtained for 21 validations

Illustrative example: EuroSCORE



Step 5
Meta-analysis

UMC Utrecht

Meta-analysis

Fixed or random effects?

- Fixed effect meta-analysis
  - The model's *true* predictive accuracy is the same for all validation studies
  - Variation in predictive accuracy only appears due to chance

- Random effects meta-analysis
  - The model's *true* predictive accuracy differs across validation studies
  - Variation in predictive accuracy arises from sampling error and between-study heterogeneity

Meta-analysis

Homogeneous model performance often unrealistic

- Validation studies typically differ in design, execution and case-mix variation
- Ignoring heterogeneity leads to an overly precise summary result
- Summary estimates of predictive accuracy have limited usefulness when there is strong heterogeneity

Meta-analysis

Traditional meta-analysis methods approximate within-study variability with a Normal distribution. This approximation may introduce bias or show other poor statistical properties when

- The c-statistic or O:E ratio is close to 0 or 1
- When sample sizes are relatively small

Need for transformations!

- Meta-analysis of logit c-statistic
- Meta-analysis of log O:E ratio

Meta-analysis

Quantifying heterogeneity

Prediction interval

- Combines the standard error of the summary estimate with the estimate for between-study variability
- Typically based on Student's t distribution
- Provides a range for the potential predictive accuracy in a new validation study
- Ideally calculated from 10 or more validation studies

Meta-analysis

Quantifying heterogeneity

Probability of “good” performance

- Calculate the likelihood of achieving a certain c-statistic and/or total O:E ratio in a new validation study
- Rough indication of model generalizability

Illustrative example: EuroSCORE

Meta-analysis	N	Summary	95% CI	95% PI
C-statistic	18	0.78	0.76 – 0.80	0.73 – 0.83
O:E ratio	19	0.55	0.43 – 0.69	0.20 – 1.53

- Probability of "good" discrimination ($c > 0.75$) = **89%**
- Probability of "good" calibration ($0.8 \leq O:E \leq 1.2$) = **15%**

Step 6

Investigating heterogeneity across studies

Investigating heterogeneity across studies

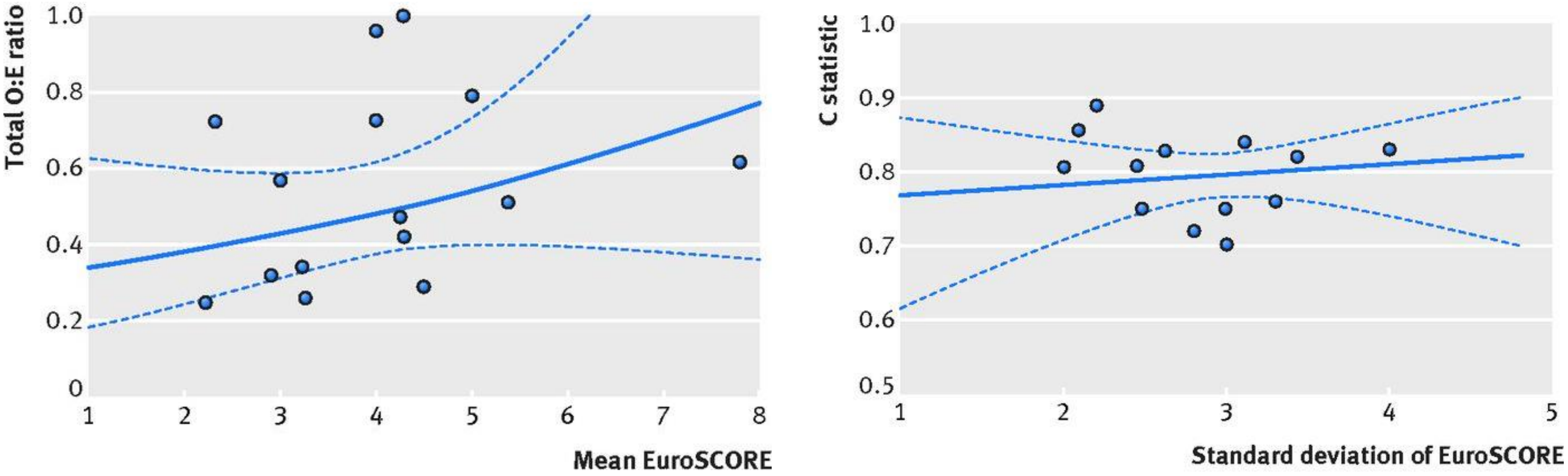
Meta-regression to adjust the meta-analysis for study-level variables

- Study characteristics
  - *Study design, follow-up, …*
  - *Predictor- and outcome definitions*

- Population characteristics
  - *Distribution of linear predictor or individual covariates*
  - *Treatment standards* ***(beware of ecological fallacy)***

Illustrative example: EuroSCORE

Adjustment for case-mix variation



Step 7

Sensitivity analyses

UMC Utrecht

Sensitivity analyses

Evaluate the robustness of drawn conclusions

- Influence of low(er) quality validation studies
- Influence of key modelling assumptions
  - Use of "exact" likelihood models
  - Joint pooling of discrimination and calibration
- …

Illustrative example: EuroSCORE

Meta-analysis	ROB	M	Summary	95% CI	95% PI
C-statistic	All	18	0.78	0.76 – 0.80	0.73 – 0.83
	Low	4	0.80	0.73 – 0.85	0.66 – 0.89
O:E ratio	All	19	0.55	0.43 – 0.69	0.20 – 1.53
	Low	3	0.57	0.10 – 3.33	0.02 – 19.15

Step 8

Reporting

Reporting

Relevant guidelines

- PRISMA
- TRIPOD
- GRADE

Case study

Performance of the Pooled Cohort Equations prognostic model

Step 1

Formulating the review question and protocol

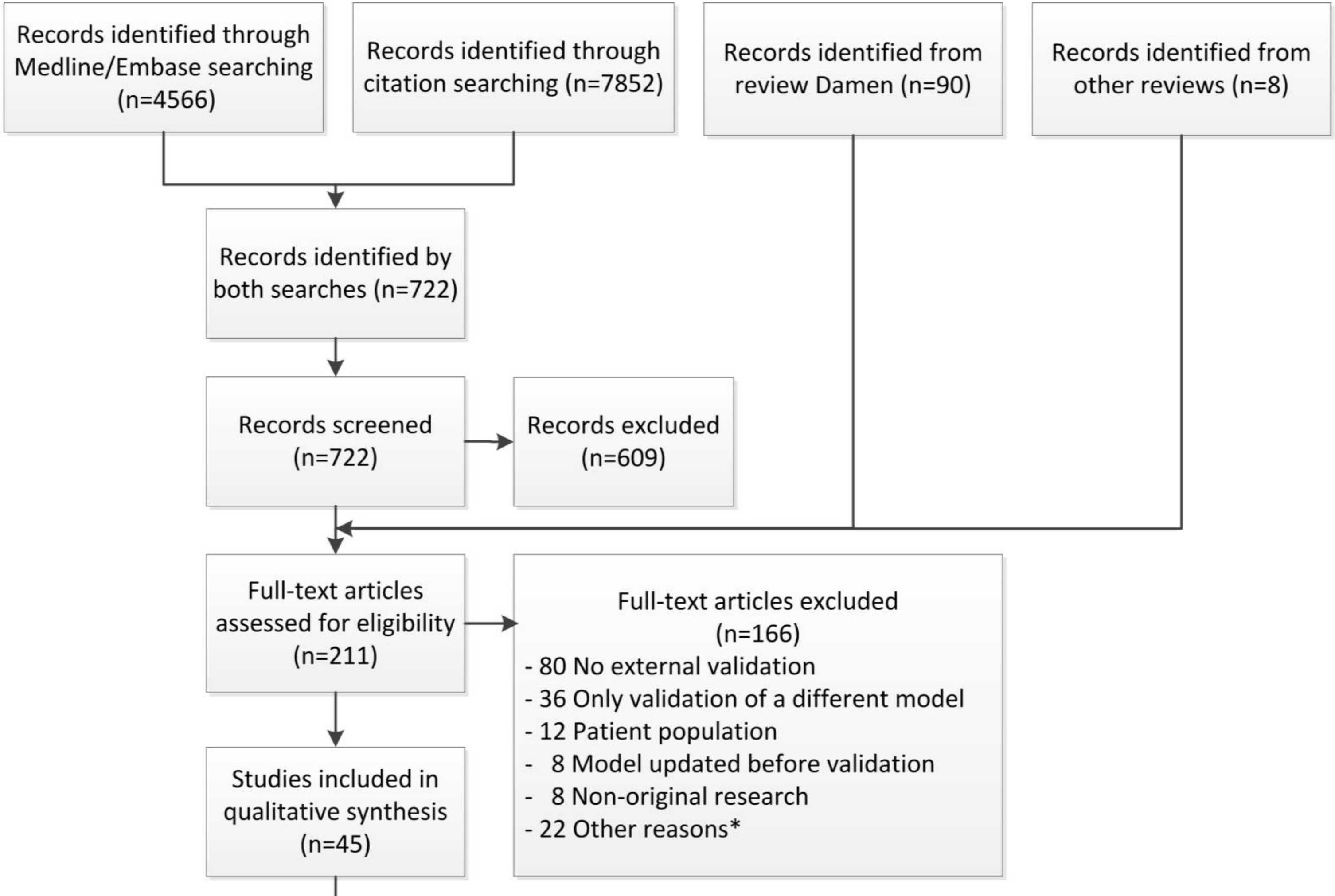
Predictive performance of PCE

Population	General population
Intervention	PCE
Comparator	Framingham Wilson and ATP III
Outcome(s)	Cardiovascular Disease (CVD)
Timing	10 year
Setting	Primary care and public health

Step 2
Formulating the search strategy

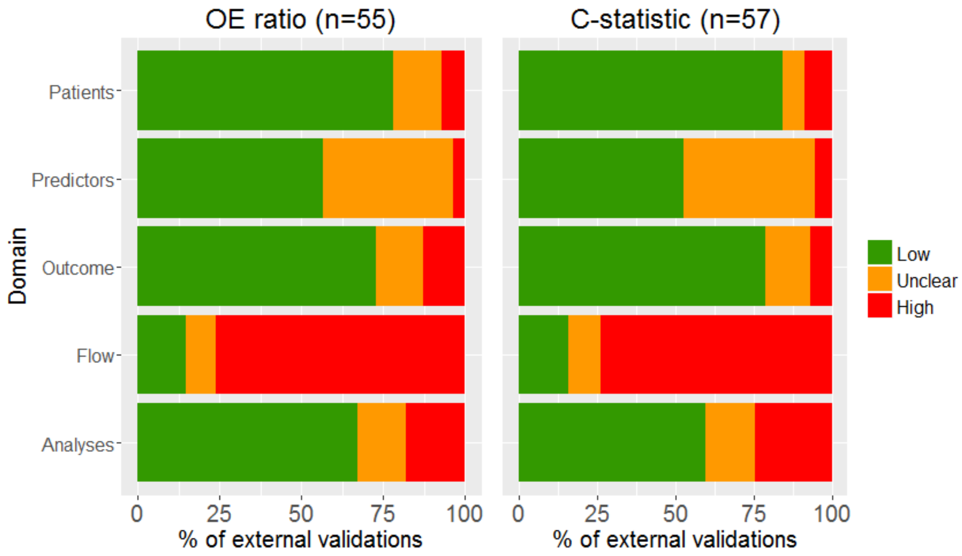
- Articles published before June 2013 selected from a previous review[1]
- Update using citation search

[1]BMJ 2016;353:i2416


Records identified through Medline/Embase searching (n=4566)
Records identified through citation searching (n=7852)
Records identified from review Damen (n=90)
Records identified from other reviews (n=8)
Records identified by both searches (n=722)
Records screened (n=722)
Records excluded (n=609)
Full-text articles assessed for eligibility (n=211)
Full-text articles excluded (n=166)
- 80 No external validation
- 36 Only validation of a different model
- 12 Patient population
- 8 Model updated before validation
- 8 Non-original research
- 22 Other reasons*
Studies included in qualitative synthesis (n=45)

Step 3
Critical appraisal

Risk of bias assessed using a preliminary version of PROBAST

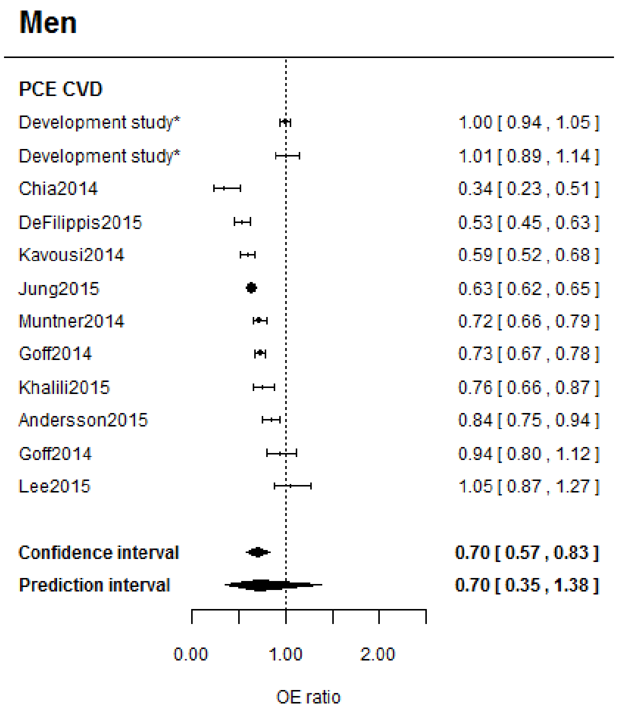


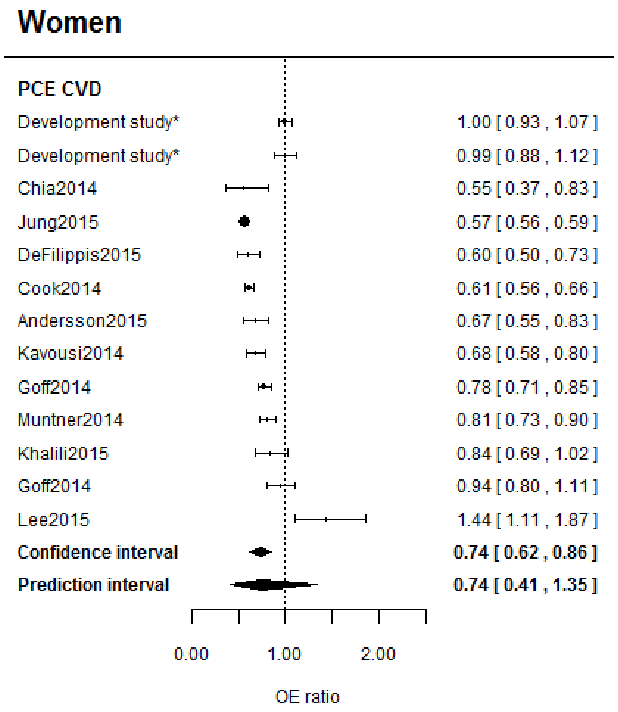
Step 4
Quantitative data extraction and preparation

Items extracted	
Study design	Outcome definition
Study population, location	Sample size
Study dates	Model discrimination (c-statistic)
Case-mix	Model calibration (O:E ratio)
Predictors	Model calibration (slope)

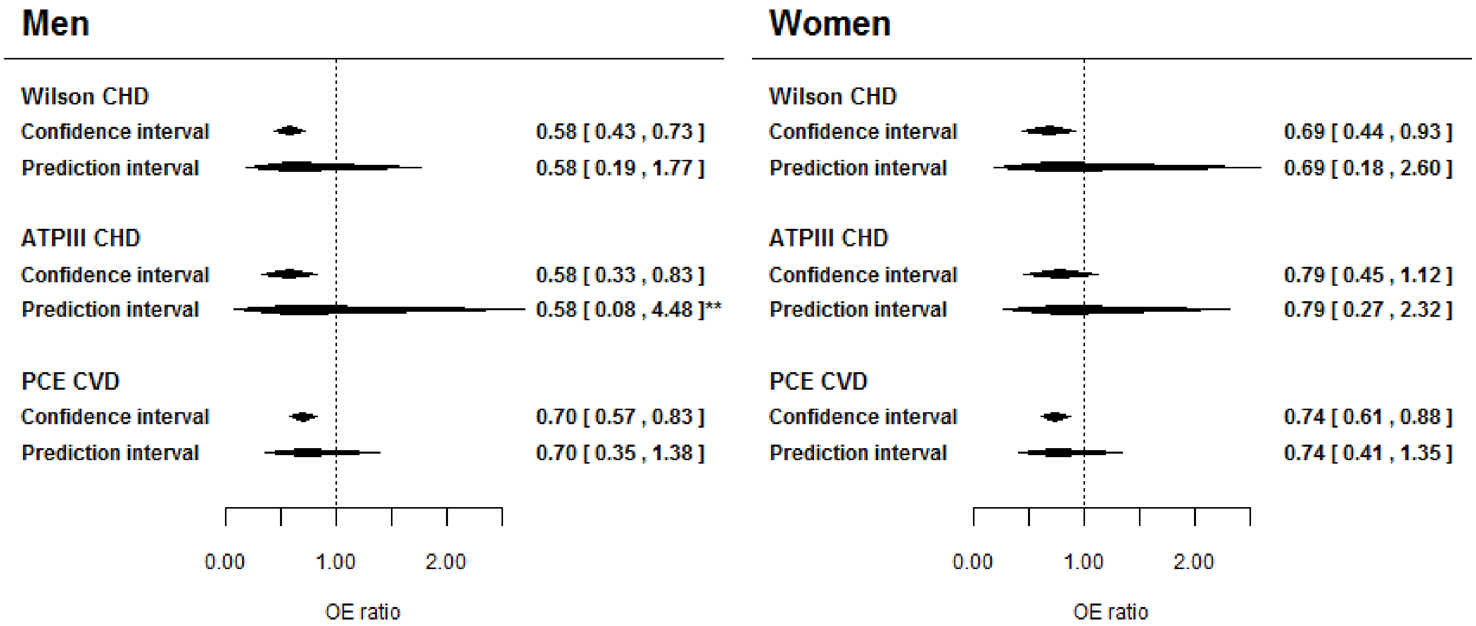
Step 5

Meta-analysis

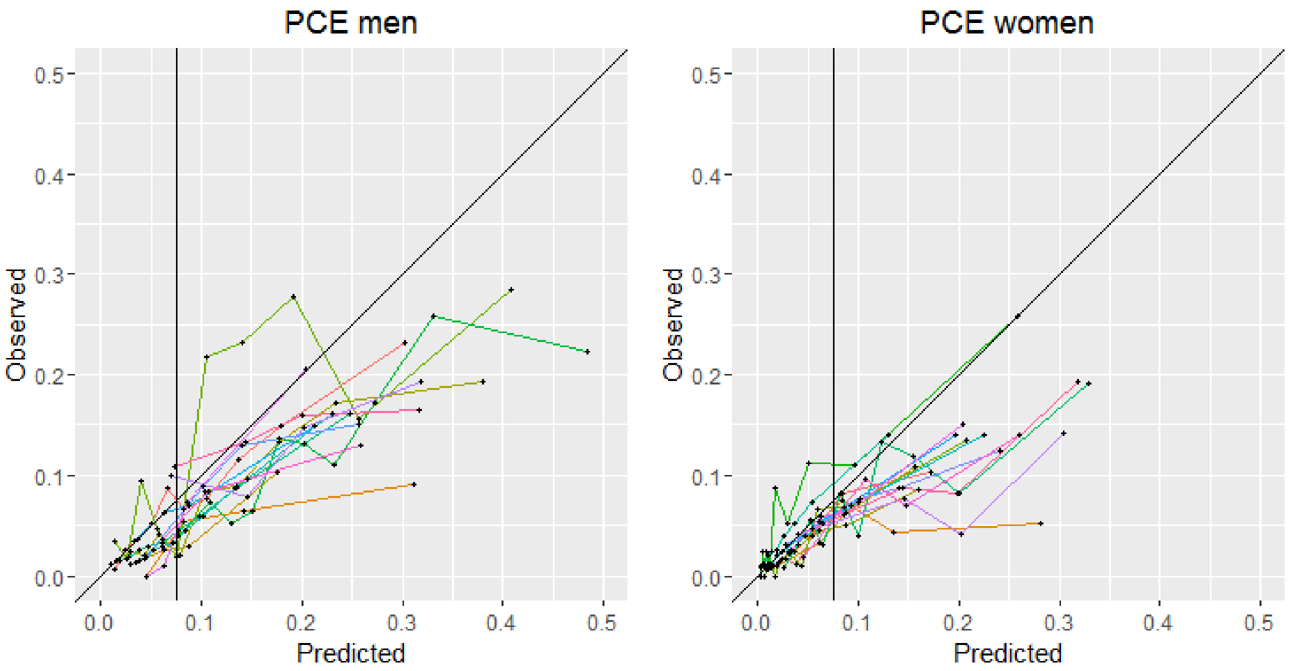




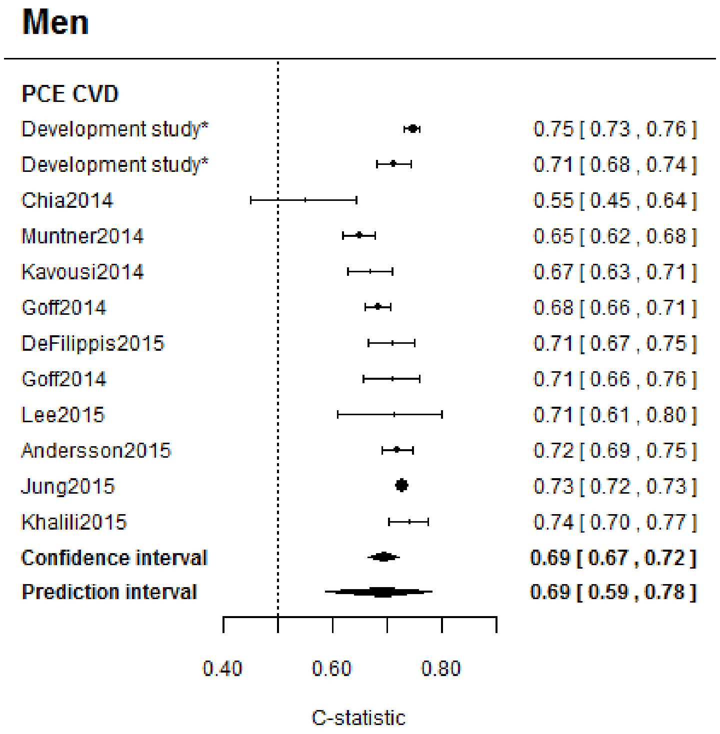
Step 5
Meta-analysis

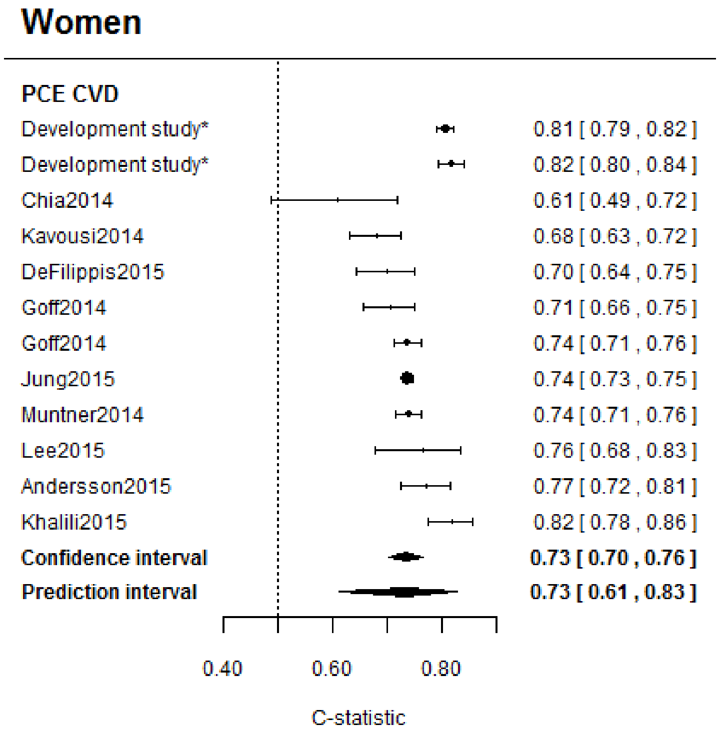


Step 5
Meta-analysis

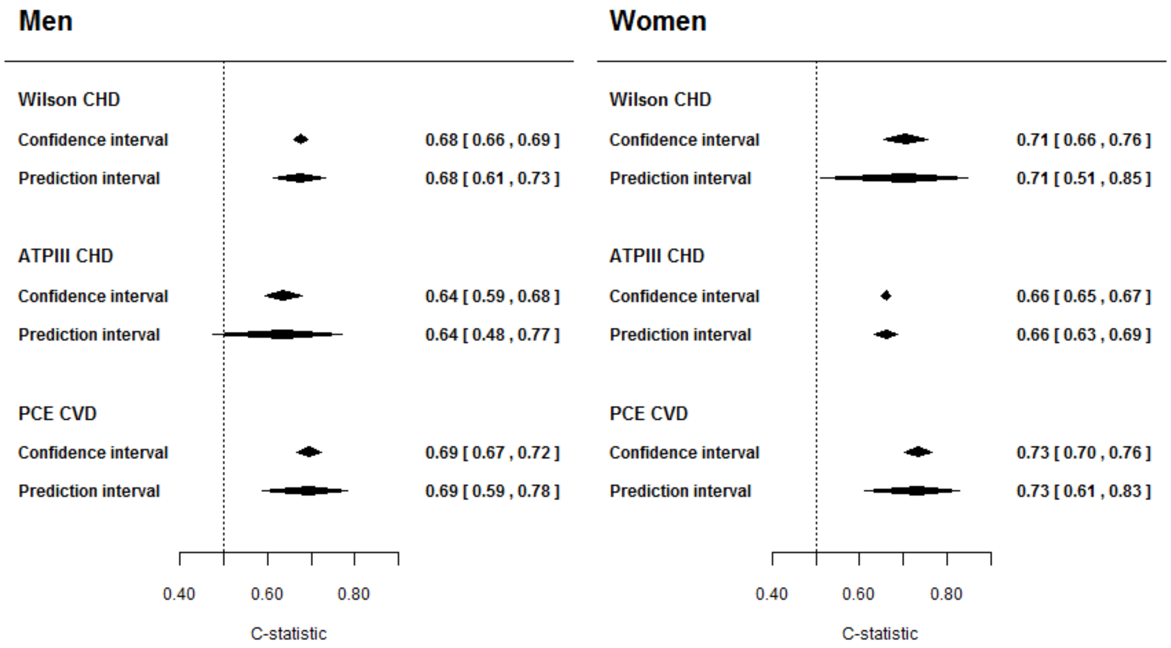


Step 5
Meta-analysis





Step 5
Meta-analysis



Step 6

Investigating heterogeneity across studies

<u>OE ratio</u>

- Closer to 1 in US compared to other continents
- No association found for other variables (e.g. elgibility criteria, patient characteristics, year)

<u>C-statistic</u>

- Decrease with higher mean age, mean SBP and lower sd age
- No association found for other variables

Step 7
Sensitivity analyses

		PCE men		PCE women
OE ratio	N	OE (95%CI)	N	OE (95%CI)
All validations	10	0.698 (0.565-0.862)	11	0.742 (0.62-0.888)
Low risk of bias for all domains	2	-	3	-
Weighted by number of events	10	0.698 (0.567-0.86)	11	0.739 (0.619-0.881)
Bivariate analyses	10	0.693 (0.58-0.828)	11	0.739 (0.633-0.863)
Not extrapolated to 10 year	10	0.698 (0.565-0.862)	11	0.742 (0.62-0.888)
C-statistic	N	C (95%CI)	N	C (95%CI)
All validations	10	0.694 (0.660-0.726)	10	0.733 (0.695-0.768)
Low risk of bias for all domains	2	-	2	-
Weighted by number of events	10	0.696 (0.664-0.726)	10	0.733 (0.694-0.769)
Bivariate analyses	10	0.695 (0.665-0.724)	11	0.734 (0.703-0.762)

Closing remarks

UMC Utrecht

Closing remarks

- Many similarities to other types of meta-analysis, however,
  - Data extraction more difficult
  - Heterogeneity more common
  - Summary estimates less meaningful
- Need to focus more on
  - Quantifying between-study heterogeneity
  - Assessing sources of variability in model performance

Conducting systematic reviews of prediction model studies

Reporting of primary study	Transparent reporting of prediction models for prognosis and diagnosis (TRIPOD) – *Collins et al. 2015 Ann Intern Med; Moons et al. 2015 Ann Intern Med*
Defining review question and developing criteria for including studies	Guidance for defining review question, design of the review and checklist for critical appraisal and data extraction (CHARMS) – *Moons et al 2014 PLOS Med*
Searching for studies	Search filters for prediction studies – *Geersing et al. 2012 PLOS One; Ingui et al. 2002 J Am Med Inform Assoc; Wong et al. 2003 AMIA Annual Symp Proc*
Selecting studies and collecting data	Guidance for defining review question, design of the review and checklist for critical appraisal and data extraction (CHARMS) – *Moons et al 2014 PLOS Med*
Assessing risk of bias and applicability in included studies	Assessment of risk of bias and applicability (PROBAST) – *Wolff et al. Publication in 2017, Moons et al. Publication in 2017*
Analysing data and undertaking meta-analyses	Meta-Analysis of clinical prediction models *Ahmed et al. BMC Res Meth 2014; Debray et al. Stat Med 2012; Debray et al. Stat Med 2014 + Debray et al BMJ 2016*
Interpreting results and drawing conclusions	Guidance for interpretation of results *Ahmed et al. BMC Res Meth 2014; Debray et al. Stat Med 2012; Debray et al. Stat Med 2014; PROBAST*
Reporting of systematic reviews	Transparent reporting of systematic reviews and meta-analysis (PRISMA) *Moher et al. PLOS Med 2009*
Assessing risk of bias of systematic reviews	Risk of bias in systematic reviews (ROBIS) *Whiting et al. J Clin Epid 2015*

Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 - http://handbook.cochrane.org/

Handy tools/papers

- Debray TPA et al. A new framework to enhance the interpretation of external validation studies of clinical prediction models. J Clin Epidemiol 2015.
- Debray TPA et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ 2017.
- Debray TPA et al. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. Stat Methods Med Res 2018.
- Snell KIE et al. Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model. J Clin Epidemiol 2015.
- Snell KIE et al. Prediction model performance across multiple studies: which scale to use for the c-statistic and calibration measures? Stat Met Meth Res 2017.

Workshop aftercare

- Questions about workshop?
- Assistant needed with review of studies of prognosis studies?
- Visit our website: https://methods.cochrane.org/prognosis/
- Please contact:
  - PMG Coordinator: Anneke Damen (CochranePMG@umcutrecht.nl)
  - PMG Co-convenor: Karel Moons (K.G.M.Moons@umcutrecht.nl)