

Meta-analytical approaches in systematic reviews of prognosis studies

Trusted evidence.
Informed decisions.
Better health.



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!
- Example: 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+AI)

RESEARCH METHODS AND REPORTING

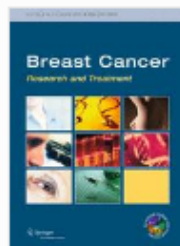


OPEN ACCESS



TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods

Gary S Collins,¹ Karel G M Moons,² Paula Dhiman,¹ Richard D Riley,^{3,4} Andrew L Beam,⁵ Ben Van Calster,^{6,7} Marzyeh Ghassemi,⁸ Xiaoxuan Liu,^{9,10} Johannes B Reitsma,² Maarten van Smeden,² Anne-Laure Boulesteix,¹¹ Jennifer Catherine Camaradou,^{12,13} Leo Anthony Celi,^{14,15,16} Spiros Denaxas,^{17,18} Alastair K Denniston,^{4,9} Ben Glocker,¹⁹ Robert M Golub,²⁰ Hugh Harvey,²¹ Georg Heinze,²² Michael M Hoffman,^{23,24,25,26} André Pascal Kengne,²⁷ Emily Lam,¹² Naomi Lee,²⁸ Elizabeth W Loder,^{29,30} Lena Maier-Hein,³¹ Bilal A Mateen,^{17,32,33} Melissa D McCradden,^{34,35} Lauren Oakden-Rayner,³⁶ Johan Ordish,³⁷ Richard Parnell,¹² Sherri Rose,³⁸ Karandeep Singh,³⁹ Laure Wynants,⁴⁰ Patricia Logullo¹




[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 , Ikhlmaq 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](#)

47

Citations

1.4k

Downloads

[Go to old article view](#)



PDF




Info



[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](https://doi.org/10.1111/jth.13690) [View/save citation](#)

Cited by (CrossRef): 0 articles  [Check for updates](#)  [Citation tools](#) ▼



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](https://doi.org/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.

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

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 Collins⁴  and Karel GM Moons^{1,2}

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCHStatistical Methods in Medical Research
0(0) 1–19

© The Author(s) 2018



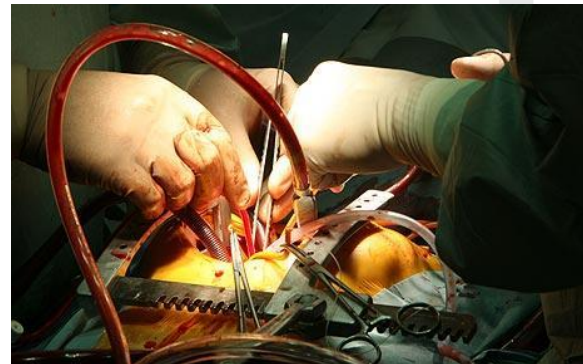
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0962280218785504
journals.sagepub.com/home/smm

 SAGE


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			Cardiac related factors		
Age ¹ (years)	<input type="text" value="0"/>	<input type="text" value="0"/>	NYHA	<input type="text" value="select"/>	<input type="text" value="0"/>
Gender	<input type="text" value="select"/>	<input type="text" value="0"/>	CCS class 4 angina ⁸	<input type="text" value="no"/>	<input type="text" value="0"/>
Renal impairment ² <small>See calculator below for creatinine clearance</small>	<input type="text" value="normal (CC >85ml/min)"/>	<input type="text" value="0"/>	LV function	<input type="text" value="select"/>	<input type="text" value="0"/>
Extracardiac arteriopathy ³	<input type="text" value="no"/>	<input type="text" value="0"/>	Recent MI ⁹	<input type="text" value="no"/>	<input type="text" value="0"/>
Poor mobility ⁴	<input type="text" value="no"/>	<input type="text" value="0"/>	Pulmonary hypertension ¹⁰	<input type="text" value="no"/>	<input type="text" value="0"/>
Previous cardiac surgery	<input type="text" value="no"/>	<input type="text" value="0"/>			
Chronic lung disease ⁵	<input type="text" value="no"/>	<input type="text" value="0"/>			
Active endocarditis ⁶	<input type="text" value="no"/>	<input type="text" value="0"/>			
Critical preoperative state ⁷	<input type="text" value="no"/>	<input type="text" value="0"/>			
Diabetes on insulin	<input type="text" value="no"/>	<input type="text" value="0"/>			
			Operation related factors		
			Urgency ¹¹	<input type="text" value="elective"/>	<input type="text" value="0"/>
			Weight of the intervention ¹²	<input type="text" value="isolated CABG"/>	<input type="text" value="0"/>
			Surgery on thoracic aorta	<input type="text" value="no"/>	<input type="text" value="0"/>
EuroSCORE II <input type="text" value="0"/>					
 Note: This is the 2011 EuroSCORE II			<input type="button" value="Calculate"/> <input type="button" value="Clear"/>		

Illustrative example: EuroSCORE

<u>P</u> opulation	Patients undergoing coronary artery bypass grafting
<u>I</u> ntervention	The (additive) EuroSCORE model
<u>C</u> omparator	Not applicable
<u>O</u> utcome(s)	All cause mortality
<u>T</u> iming	30 days, predicted using peri-operative conditions
<u>S</u> etting	risk stratification in the assessment of cardiac surgical results

Quantitative data extraction and preparation

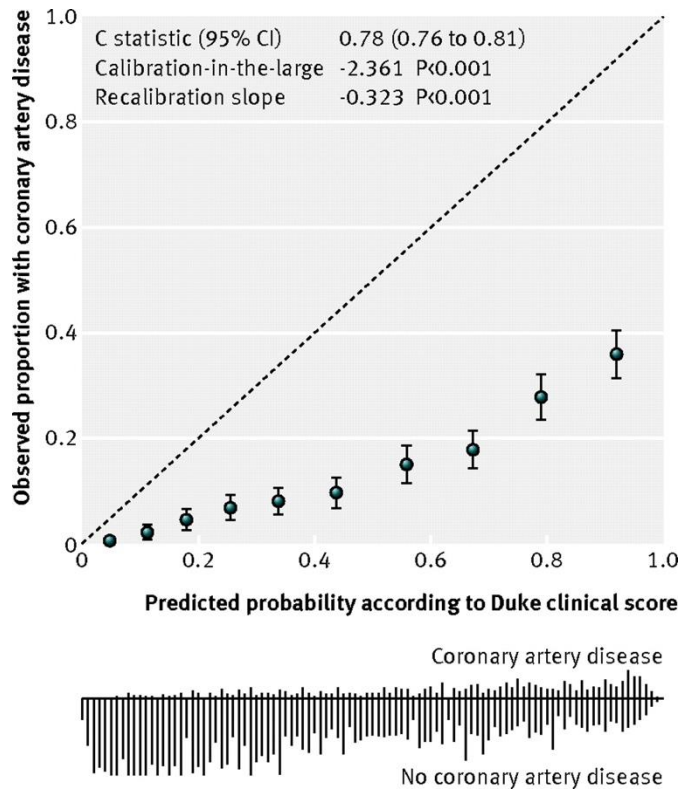
Trusted evidence.
Informed decisions.
Better health.



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





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

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



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 ($\geq 3.2.0$), stats, graphics
Imports: [metafor](#) ($\geq 2.0.0$), [mvtnorm](#), [ellipse](#), [lme4](#), [plyr](#), [ggplot2](#)
Suggests: [runjags](#), [rjags](#), [testthat](#) ($\geq 1.0.2$)
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](#)

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

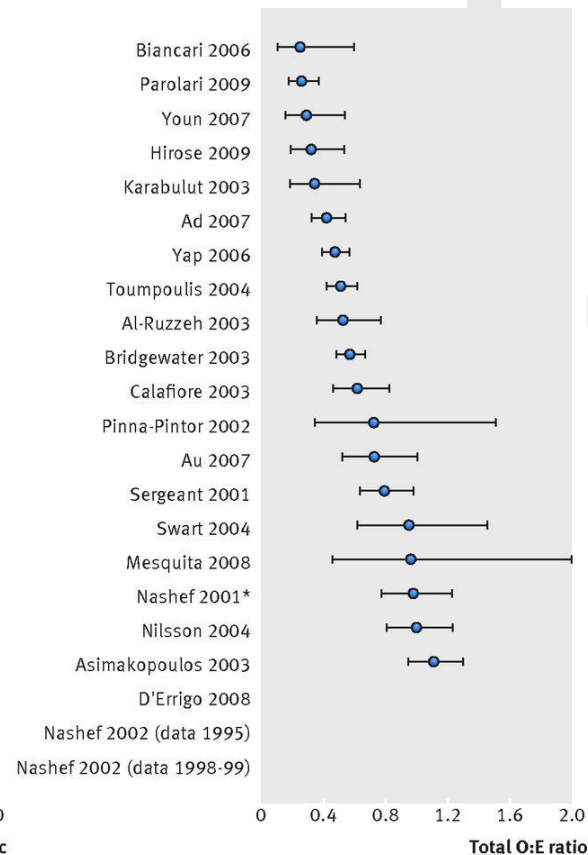
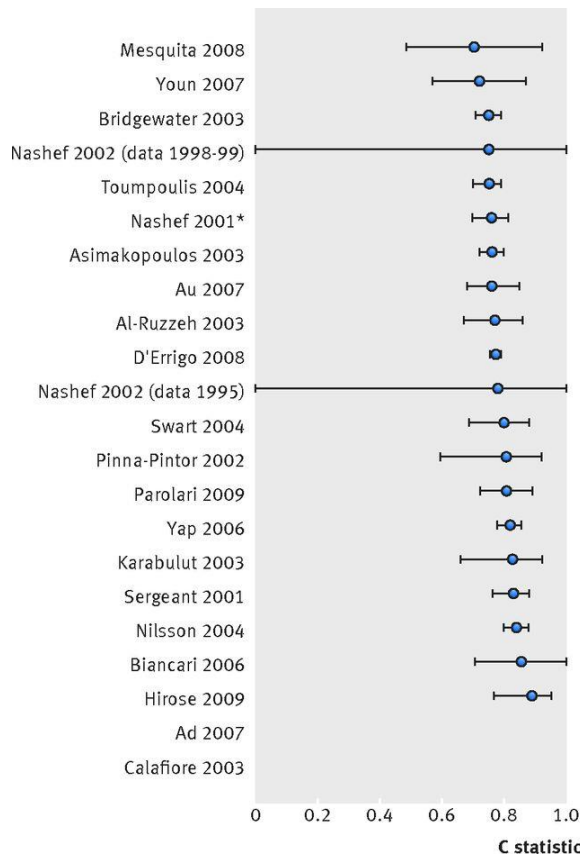
- Summary statistic reported in 20 validations
- SE approximated for 7 studies

O:E

- Relevant information obtained for 21 validations



Illustrative example: EuroSCORE



Meta-analysis

Trusted evidence.
Informed decisions.
Better health.



Fixed or random effects meta-analysis?

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
- **Need for transformations!**
 - Meta-analysis of **logit** c-statistic
 - Meta-analysis of **log** O:E ratio



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



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

Investigating heterogeneity across studies

Trusted evidence.
Informed decisions.
Better health.



Investigating heterogeneity across studies

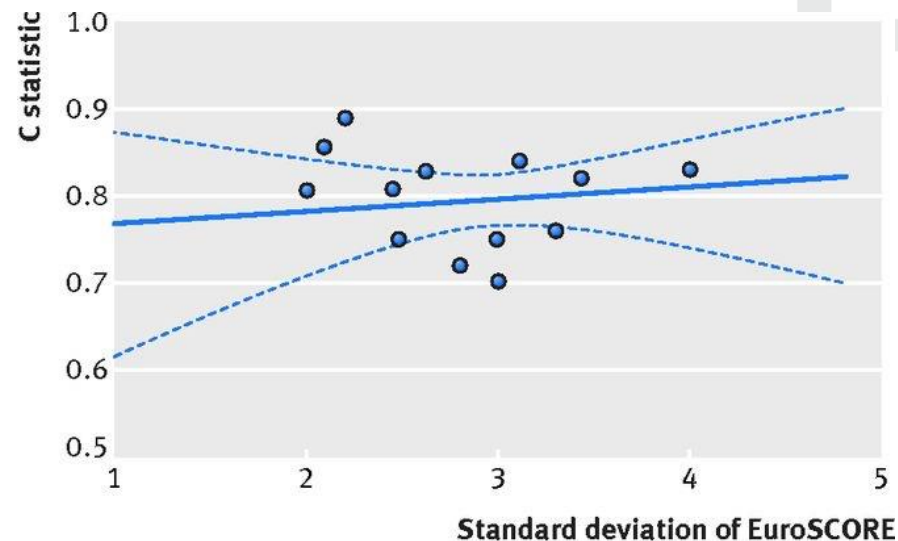
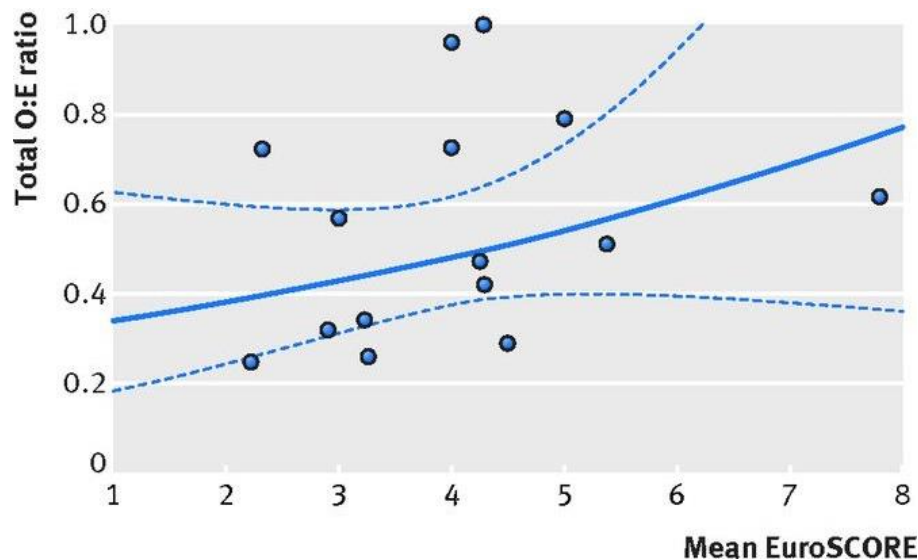
Meta-regression to adjust the meta-analysis for study-level variables

- 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



Sensitivity analyses

Trusted evidence.
Informed decisions.
Better health.



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

Apply your knowledge!

Damen et al. *BMC Medicine* (2019) 17:109
<https://doi.org/10.1186/s12916-019-1340-7>


BMC Medicine

RESEARCH ARTICLE

Open Access

Performance of the Framingham risk models and pooled cohort equations for predicting 10-year risk of cardiovascular disease: a systematic review and meta-analysis



Johanna A. Damen^{1,2*} , Romin Pajouheshnia², Pauline Heus^{1,2}, Karel G. M. Moons^{1,2}, Johannes B. Reitsma^{1,2}, Rob J. P. M. Scholten^{1,2}, Lotty Hooft^{1,2} and Thomas P. A. Debray^{1,2}

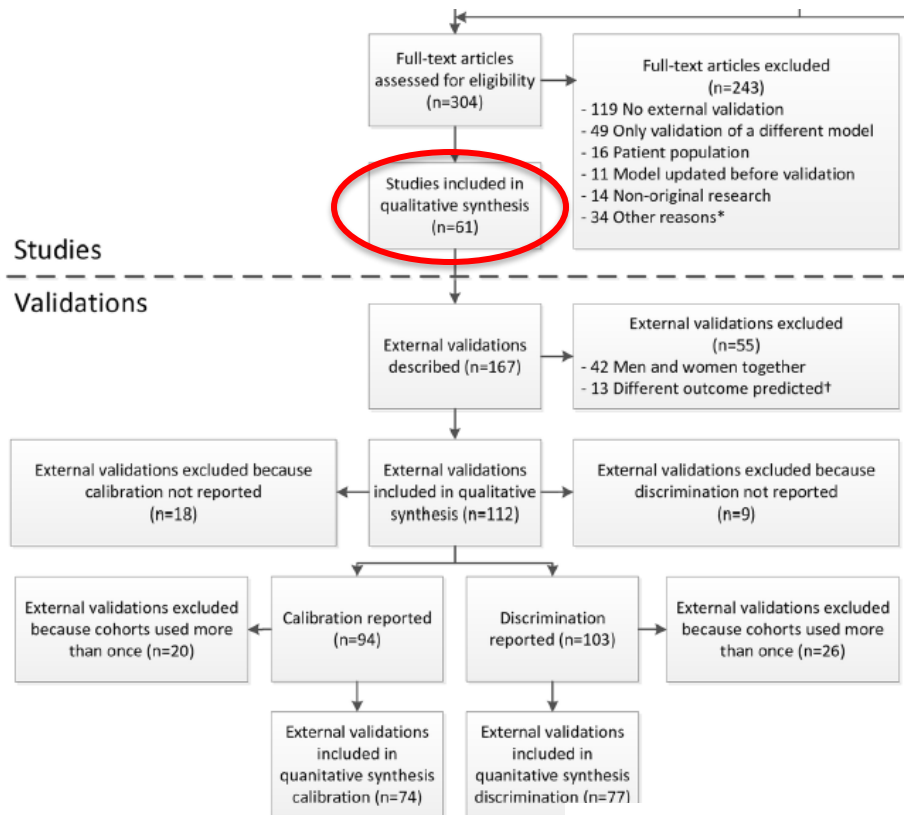
Questions (*focus on men only!*)

1. Formulate the PICOTS of the review
2. How many studies were included in the systematic review? How many validations were included in the meta-analysis of the OE ratio of the PCE? Explain the difference.
3. What are the pooled estimates of the c-statistic (including CI and PI) of the 3 models? How would you interpret these values?
4. What are the pooled estimates of the OE ratio (including CI and PI) of the 3 models? How would you interpret these values?
5. Is there a lot of heterogeneity (i.e. variation) in the meta-analysis of the c-statistic? And the OE ratio? What could be causing this heterogeneity? Explain your answers.
6. What would be your advise for using these models in clinical practice?

PICOTS

<u>P</u> opulation	General population
<u>I</u> ndex model	PCE
<u>C</u> omparator model	Framingham Wilson and ATP III
<u>O</u> utcome(s)	Coronary heart disease (CHD) or cardiovascular disease (CVD)
<u>T</u> iming	10 year
<u>S</u> etting	Primary care and public health

Number of included studies



PCE CVD

Development study*

Development study*

Chia 2014

Emdin 2017

De Filippis 2017

Kavousi 2014

Pike 2016

De Las Heras Gala 2016

Sussman 2017

Jung 2015

Yang 2016

Yang 2016

Yang 2016

Mortensen 2017**

De Las Heras Gala 2016

Muntner 2014

Goff 2014

Khalili 2015

Andersson 2015

Goff 2014

Lee 2015

1.00 [0.94 , 1.05]

1.01 [0.89 , 1.14]

0.34 [0.23 , 0.51]

0.41 [0.30 , 0.55]

0.52 [0.46 , 0.59]

0.59 [0.52 , 0.68]

0.61 [0.54 , 0.69]

0.62 [0.54 , 0.71]

0.63 [0.62 , 0.63]

0.63 [0.62 , 0.65]

0.64 [0.59 , 0.68]

0.65 [0.57 , 0.74]

0.66 [0.60 , 0.72]

0.66 [0.01 , 100.00]

0.70 [0.62 , 0.79]

0.72 [0.66 , 0.79]

0.73 [0.67 , 0.78]

0.76 [0.66 , 0.87]

0.84 [0.75 , 0.94]

0.94 [0.80 , 1.12]

1.05 [0.87 , 1.27]

Confidence interval
Prediction interval

0.66 [0.59 , 0.73]

0.66 [0.41 , 1.06]

0.00 1.00 2.00

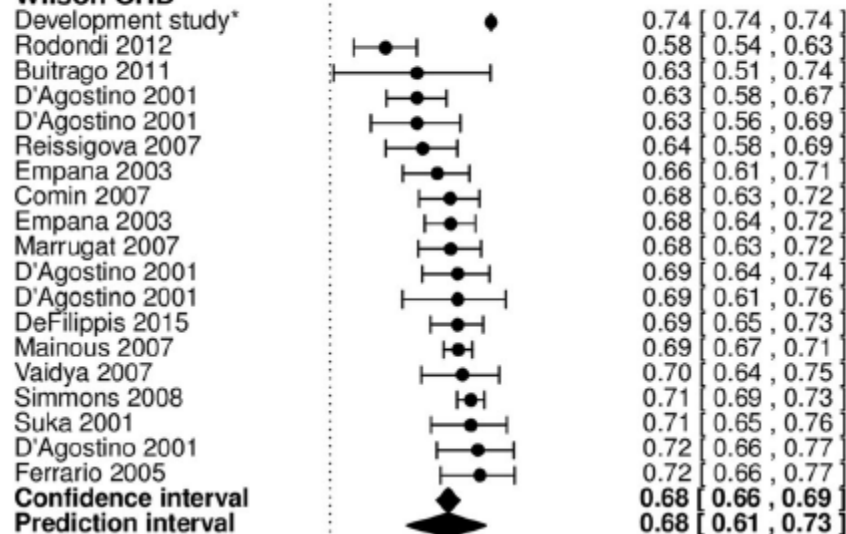
tio

N=19 validations from N=15 studies

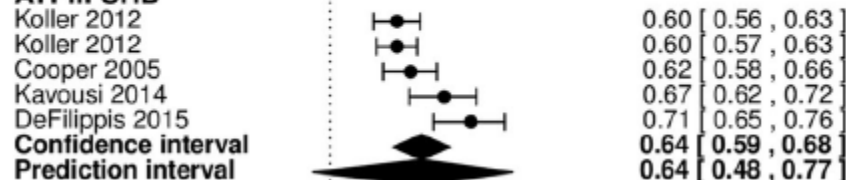
Results: c-statistic

Men

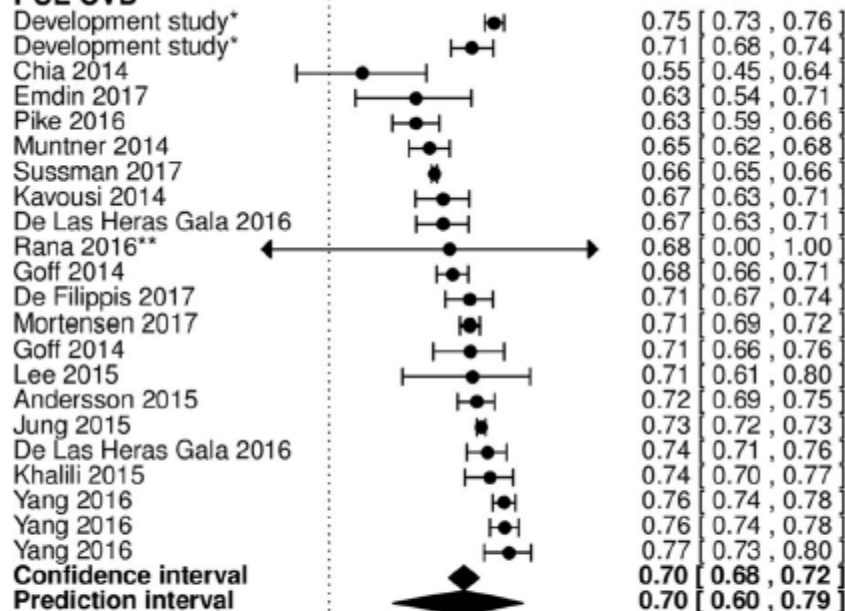
Wilson CHD



ATPIII CHD



PCE CVD



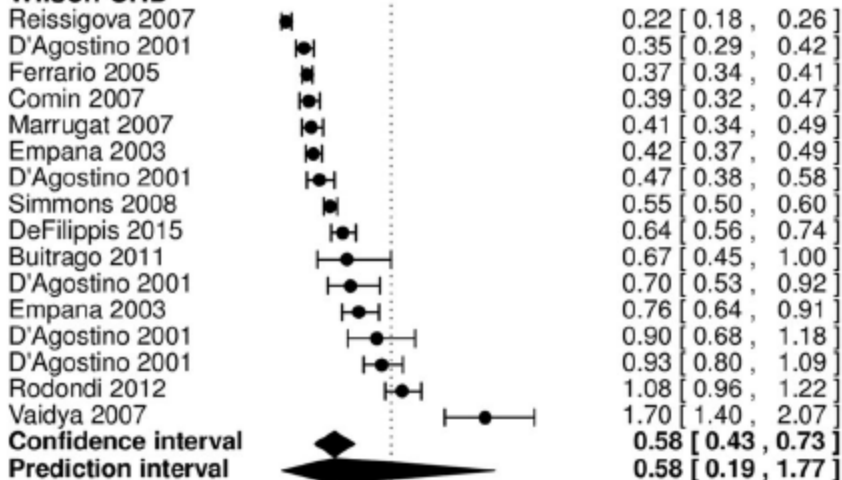
0.40 0.60 0.80

C-statistic

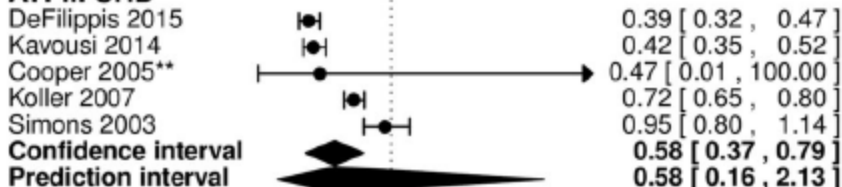
Results: OE ratio

Men

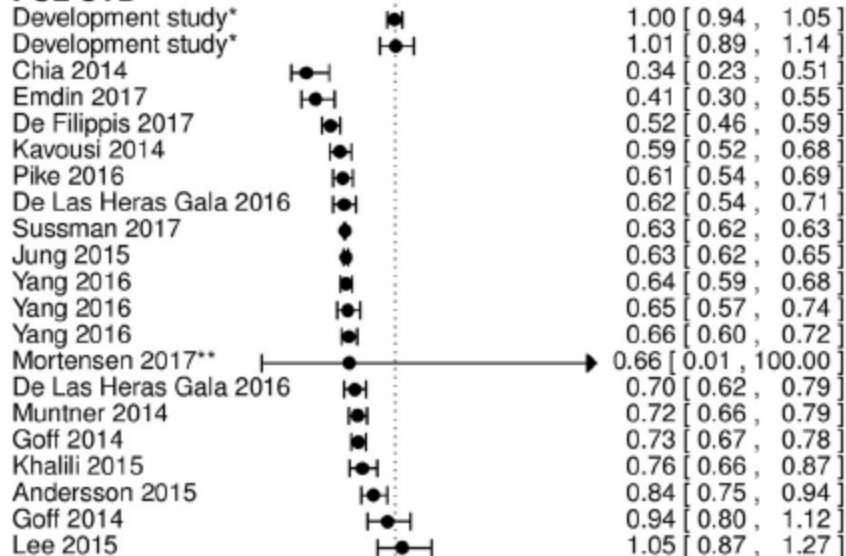
Wilson CHD



ATPIII CHD



PCE CVD



0.00 1.00 2.00

OE ratio

Heterogeneity?

Huge differences in performance between studies (especially in OE ratio)

Clinical practice?

Broad prediction intervals



Closing remarks

Trusted evidence.
Informed decisions.
Better health.



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



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.

