UMC Utrecht
Julius Center

Systematic reviews of prediction modeling studies:
planning, critical appraisal and data collection

Karel GM Moons, Lotty Hooft, Hans Reitsma, Thomas Debray
Dutch Cochrane Center
Julius Center for Health Sciences and Primary Care, UMC Utrecht

for the Cochrane Prognosis Methods Group (PMG) *Co-convenors: Doug Altman, Katrina Williams, Jill Hayden, Sue Woolfenden, Richard Riley, Karel Moons*

Conflict of interest

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

Overview Cochrane Prognostic Methods Group (PMG) Workshops

PMG Workshop	Facilitators	When?
W72. Systematic reviews of prediction modelling studies	Karel Moons, Lotty Hooft and Hans Reitsma	**Day 1 - 23 September, Tuesday: 13.30 to 15.00**
W36. Individual Participant Data (IPD) Meta-analysis of prediction modelling studies	Thomas Debray, Hans Reitsma and Karel Moons	**Day 1 - 23 September, Tuesday: 15.30 to 17.00**
W60. PROBAST: Introduction to a new risk of bias tool for prediction modelling studies	Robert Wolff, Penny Whiting and Karel Moons	**Day 3 - 25 September, Thursday: 13.30 to 15.00**
W73. Systematic reviews of prognostic studies: a meta-analytical approach	Thomas Debray and Karel Moons	**Day 4 - 26 September, Friday: 13.30 to 15.00**

Outline of workshop

- Presentations:
  - Introduction to prognostic (prediction) research
  - Review question & objectives
  - Searching
  - Quality assessment / Critical appraisal
- Useful Tools
- Workshop aftercare

Learning objectives

- To know the main types of prognostic studies
- To understand the different aims of systematic reviews of (prognostic) prediction modeling studies
- To describe the similarities and differences between intervention and prediction modeling reviews
- To understand the challenges of reviews of prediction modeling studies
- To present the key ingredients of a protocol for a systematic review of prediction modeling studies
  - except for the MA part – Friday workshop

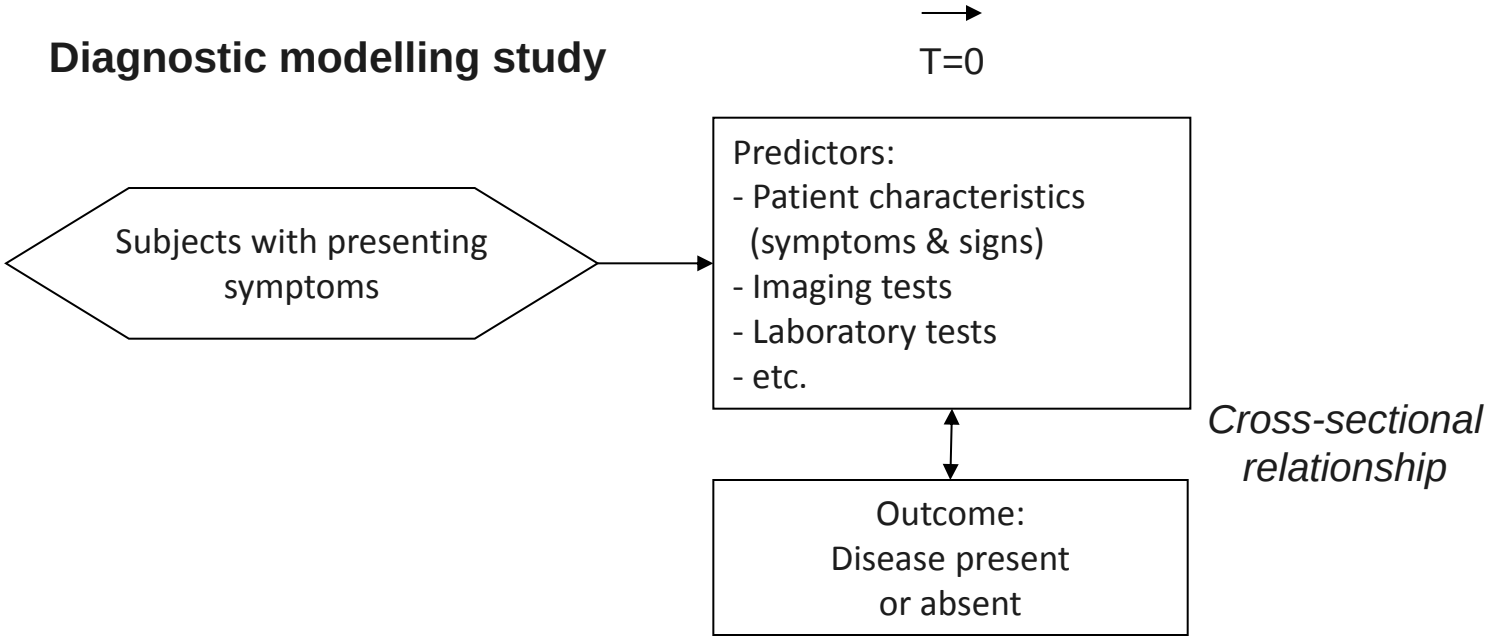
Prediction

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

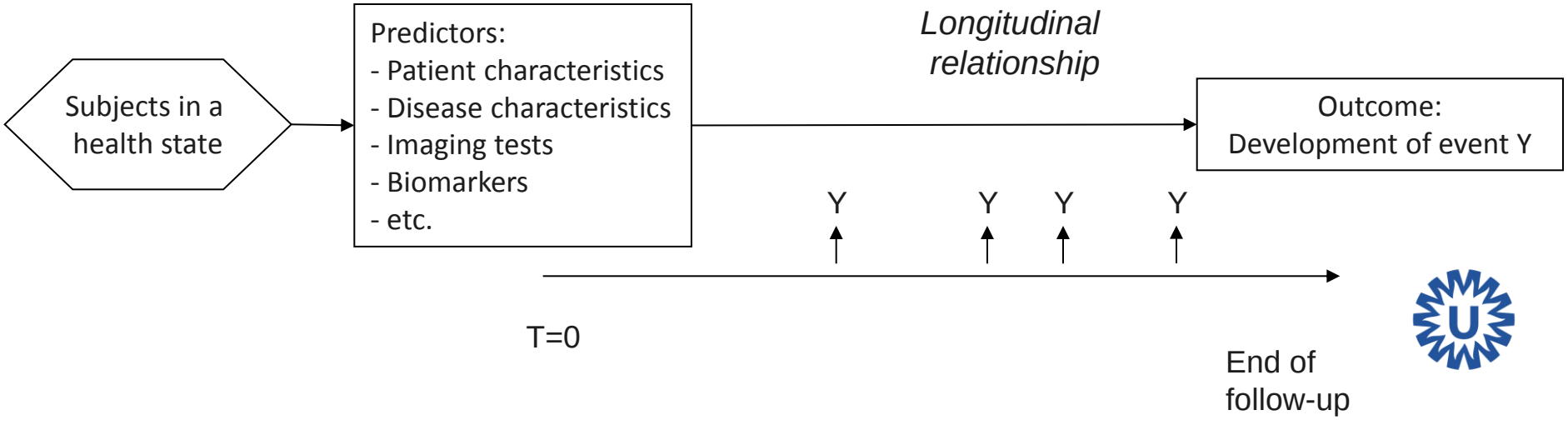
- Turn available information (predictors) into a statement about the probability:
  … diagnosis
  … prognosis

What is the big difference between diagnostic and prognostic ‘prediction’?

Diagnostic modelling study



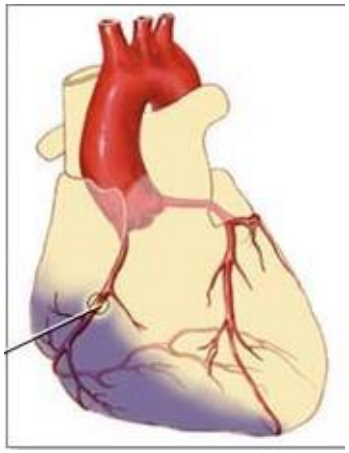
Prognostic modelling study



Prognosis in practice

- Patient:
  52 year old woman with an anterior myocardial infarction. Killip class II and diastolic blood pressure of 60 mmHg

- What would you like to know
  - as a patient? Risk of dying within 30 days?
  - as a physician? Start with treatment?



What is a prognosis?

(BMJ series 2009 (Altman, Moons, Royston, Vergouwe) + Progress series BMJ/Plos Med 2013

In medical context: a **forecast** of the **course** and **outcome** for a **individual** patient in a **certain health state** (given a **specific treatment** management)

- Not necessarily sick people

- What will happen to me?
- What will happen to this specific individual?

Main types of prognosis studies
PROGRESS series 2013: BMJ and Plos Med

Aim of prognostic studies may be:

- Average/overall prognosis: 'What is the most likely course (outcome) of people with this health condition?'
- Prognostic factor(s) finding: 'Which factors are associated with the outcome of interest?
- Prediction modeling studies: 'What is the optimal model or how good is a model in predicting risk?'
- Treatment selection/predictive factors finding: 'Does a factor lead to different treatment effect?'

Focus this workshop: Review of prediction modeling studies → notably prognostic models, but also applies to diagnostic models

Question

What is the relationship/difference between predictive and prognostic tests?

a. The terms are essentially synonymous.
b. Predictive factors are a category of prognostic factors.
c. Prognostic factors are a category of predictive factors.

Phases of Prognostic Modelling

BMJ series 2009 (Altman, Moons, Royston, Vergouwe)

1. Prediction model development studies (D)
   - Without external validation
   - With external validation in independent data

2. External model validation studies (V)
   - Without model updating
   - With model updating

3. Quantify model's impact on doctor's decision making and patient outcome (cost-effectiveness)

What is the difference between 3 versus 1 and 2?

Clinical prediction models

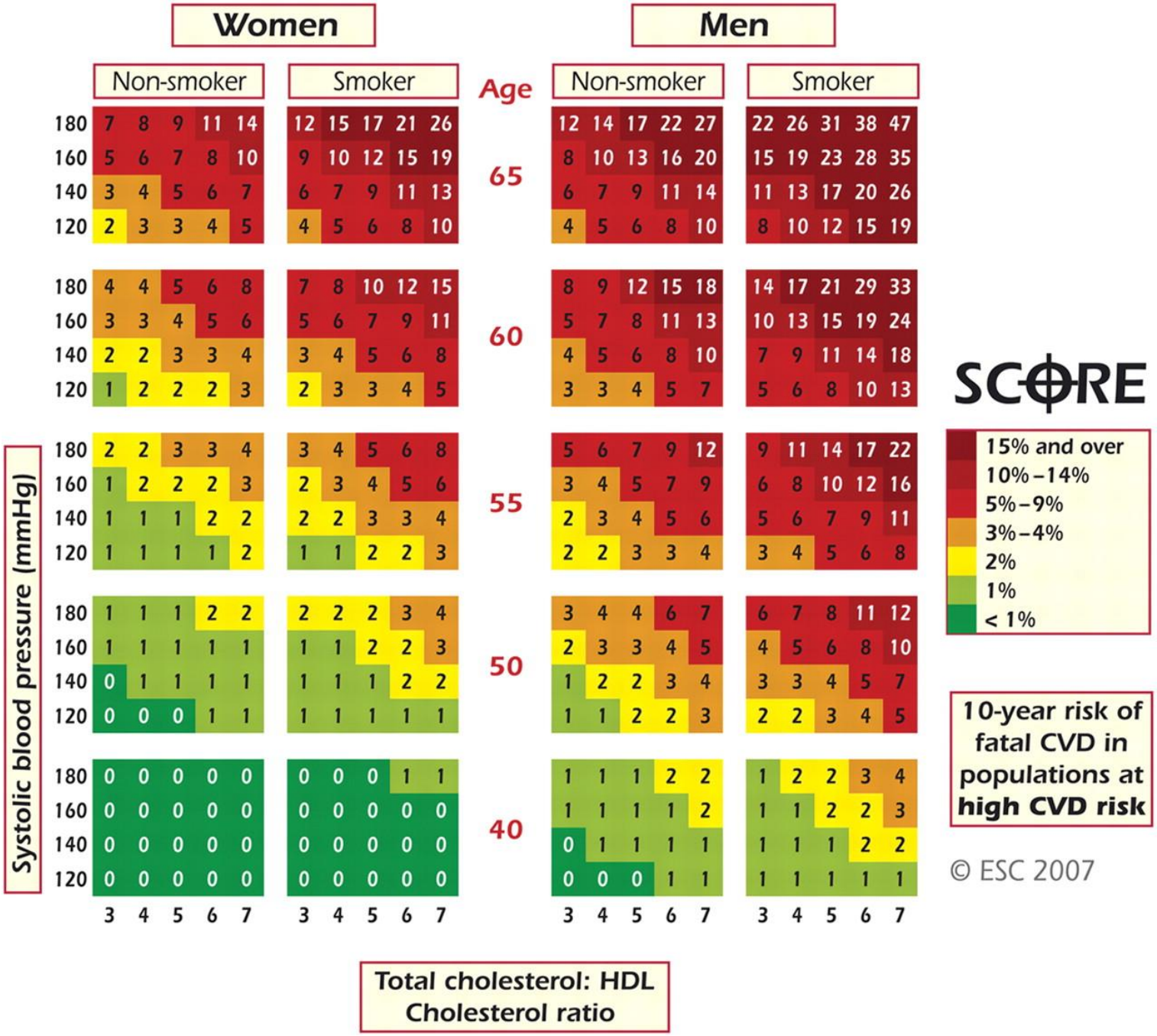
- Presented as:
  - Mathematical formula requiring computer
  - Simple scoring rules (Apgar)
  - Score charts / Nomograms (Framingham)

Apgar score in neonates (JAMA 1958)

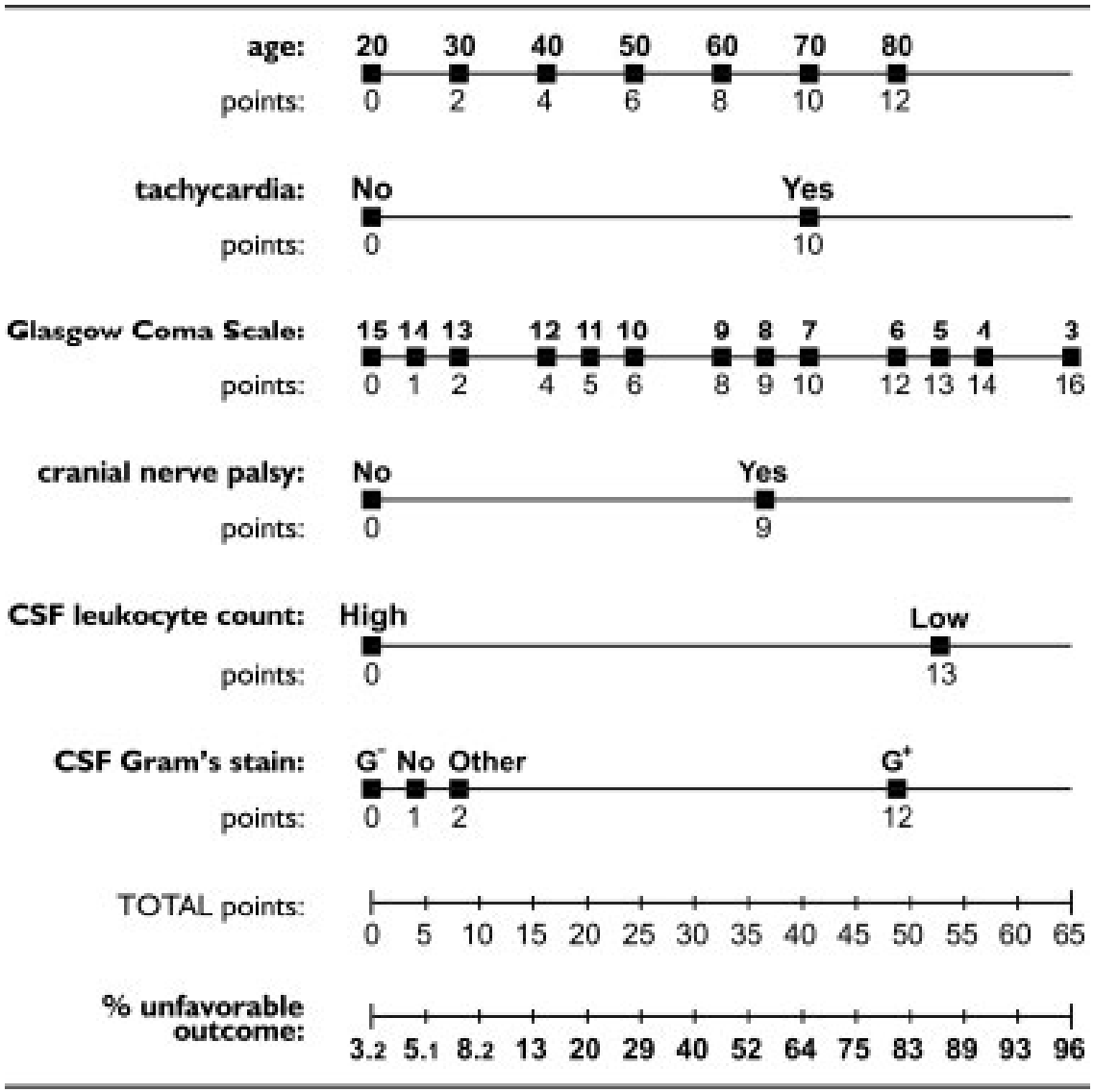
Table 9–1. Apgar scoring.

Signs	0	1	2
Heartbeat per minute	Absent	Slow (<100)	Over 100
Respiratory effort	Absent	Slow, irregular	Good, crying
Muscle tone	Limp	Some flexion of extremities	Active motion
Reflex irritability	No response	Grimace	Cry or cough
Color	Blue or pale	Body pink, extremities blue	Completely pink

Σ = Apgar score (0-10)


Women
Men
Non-smoker
Smoker
Non-smoker
Smoker
Age
Systolic blood pressure (mmHg)
180 7 8 9 11 14 | 12 15 17 21 26 | 65 | 12 14 17 22 27 | 22 26 31 38 47
160 5 6 7 8 10 | 9 10 12 15 19 | 8 10 13 16 20 | 15 19 23 28 35
140 3 4 5 6 7 | 6 7 9 11 13 | 6 7 9 11 14 | 11 13 17 20 26
120 2 3 3 4 5 | 4 5 6 8 10 | 4 5 6 8 10 | 8 10 12 15 19
180 4 4 5 6 8 | 7 8 10 12 15 | 60 | 8 9 12 15 18 | 14 17 21 29 33
160 3 3 4 5 6 | 5 6 7 9 11 | 5 7 8 11 13 | 10 13 15 19 24
140 2 2 3 3 4 | 3 4 5 6 8 | 4 5 6 8 10 | 7 9 11 14 18
120 1 2 2 2 3 | 2 3 3 4 5 | 3 3 4 5 7 | 5 6 8 10 13
180 2 2 3 3 4 | 3 4 5 6 8 | 55 | 5 6 7 9 12 | 9 11 14 17 22
160 1 2 2 2 3 | 2 3 4 5 6 | 3 4 5 7 9 | 6 8 10 12 16
140 1 1 1 2 2 | 2 2 3 3 4 | 2 3 4 5 6 | 5 6 7 9 11
120 1 1 1 1 2 | 1 1 2 2 3 | 2 2 3 3 4 | 3 4 5 6 8
180 1 1 1 2 2 | 2 2 2 3 4 | 50 | 3 4 4 6 7 | 6 7 8 11 12
160 1 1 1 1 1 | 1 1 2 2 3 | 2 3 3 4 5 | 4 5 6 8 10
140 0 1 1 1 1 | 1 1 1 2 2 | 1 2 2 3 4 | 3 3 4 5 7
120 0 0 0 1 1 | 1 1 1 1 1 | 1 1 2 2 3 | 2 2 3 4 5
180 0 0 0 0 0 | 0 0 0 1 1 | 40 | 1 1 1 2 2 | 1 2 2 3 4
160 0 0 0 0 0 | 0 0 0 0 0 | 1 1 1 1 2 | 1 1 2 2 3
140 0 0 0 0 0 | 0 0 0 0 0 | 0 1 1 1 1 | 1 1 1 2 2
120 0 0 0 0 0 | 0 0 0 0 0 | 0 0 0 1 1 | 1 1 1 1 1
3 4 5 6 7 | 3 4 5 6 7 | 3 4 5 6 7 | 3 4 5 6 7
Total cholesterol: HDL Cholesterol ratio
SCORE
15% and over
10%–14%
5%–9%
3%–4%
2%
1%
< 1%
10-year risk of fatal CVD in populations at high CVD risk
© ESC 2007

Nomogram for poor outcome in bacterial meningitis



Your Disease Risk

THE SOURCE ON PREVENTION

my results: No Results Yet

Tome el cuestionario en Español

Cancer
Diabetes
Heart disease
Osteoporosis
Stroke

8 ways to prevent disease

What is...?
Prevention
Risk
A Screening Test

How to...
Estimate Risk

Community Action

Disclaimer
Privacy Policy
About This Site

Welcome to *Your Disease Risk*, the source on prevention. Here, you can find out your risk of developing five of the most important diseases in the United States and get personalized tips for preventing them.

Developed over the past ten years by world-renowned experts, *Your Disease Risk* collects the latest scientific evidence on disease risk factors into one easy-to-use tool.

To get started, choose one of the diseases below.

What is your risk?

Cancer: There's much more to it than just smoking and lung cancer. — What's your cancer risk?

Diabetes: Over 18 million in the U.S. suffer from it. Take steps now to lower your risk. — What's your diabetes risk?

Heart disease: The #1 killer in the U.S. is also one of the most preventable. — What's your heart disease risk?

Osteoporosis: Calcium isn't the only way (or even the best way) to protect yourself. — What's your osteoporosis risk?

Stroke: Most cases of this feared disease can be avoided by lifestyle changes. — What's your stroke risk?

Systematic reviews (SRs)

- Applicable to all fields of medical research
  - Therapeutic studies (RCTs): *Cochrane Intervention Reviews*
  - Diagnostic accuracy studies: *Cochrane Diagnostic Test Accuracy Reviews (2008)*
    - Both including meta-analytical approaches
  - Next prognostic studies

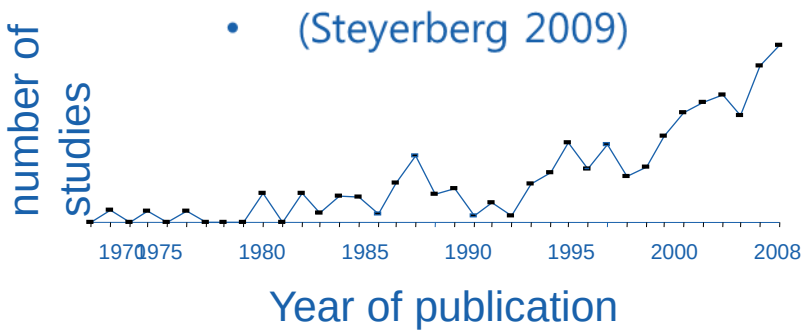
The Cochrane Collaboration
Trusted evidence. Informed decisions. Better health.

Why SRs of prognostic studies? (… and preferably a MA)

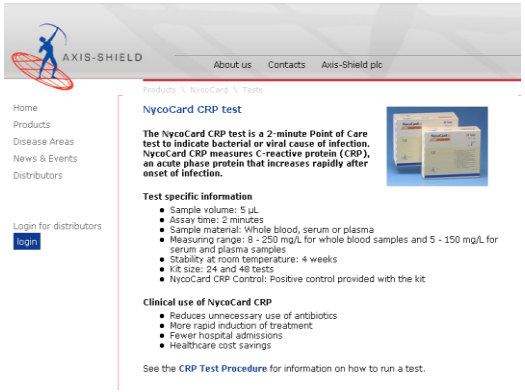
1. Summaries for evidence-based practice as the #number of prognostic studies increases per day

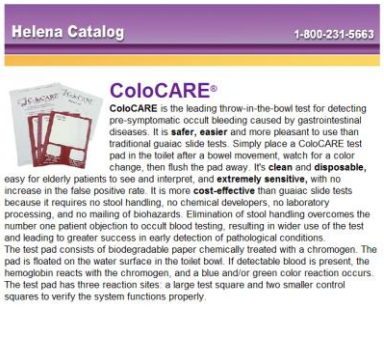
   a. Prediction models

   – Focus of today



   b. Also biomarkers (all types → high throughput studies)





Why SRs Prognostic studies? (… and preferably a MA)

2. Most studies have varying or even conflicting results

   – more prominent than in therapeutic trials

3. Relatively small studies (compared to therapeutic trials)

   – Kyzas Eur J Canc 2007; > 1500 studies cancer prognostic markers in 2005 → largest just over 1000 pts.

   – DM2 Models (Collins 2007) ; Cancer Models (Mallet 2010) ; CVD prediction models (100's)

Challenges for Prognosis Reviews

- Inconsistent terminology
  - Poor indexing in electronic databases
- Difficult searching
- Observational study designs
  - Many difference s in designs & modeling approaches
- Critical appraisal and data collection issues
- Meta Analysis

But: SR methods now available → 4 PMG workshops

Conducting a systematic review: 7 steps

1. Well-formulated review question (PICO)
2. Extensive search for studies
3. Objective selection of studies
4. Objective extraction of data
5. Critical appraisal of methodological quality
6. Synthesis of data (meta-analysis)
7. Conclusions and recommendations

Conducting a systematic reviews: Step 1

SR elements	Intervention	Prognosis
Well formulated review question	PICO Outcomes: effectiveness and harm of specific treatment	PICOTS Outcomes: factors associated with certain health condition Single factor or prediction model (different phases)
Extensive search		
Selection & Data extraction		
Critical appraisal of methodological quality		
Data syntheses (meta-analysis)		
Interpretation of results		

Well-formulated review question (PICO)

Different clinical questions possible → different aims of SR of prediction models?

Can you give some clinically relevant prognostic model review questions?

Well-formulated review question (PICO)

– How good is predictive performance of specific model for specific target population

- Predictive performance Framingham risk model / GAIL model

- Review all existing models for specific target population
  - all models for developing D2M ; for CVD consequences of DM2; for consequences after breast cancer; after cardiac surgery

– Review on added predictive value of specific predictor/biomarker/test to a specific model

- Adding CRP to Framingham risk score; D-dimer to Wells Rule
- Adding imaging results to 'basic risk scores' (cancer models)

Well-formulated review question (PICO)

– Review all models for specific outcome in specific target population

- Models predicting fatal/non-fatal CHD in general population; models predicting stroke in general population;
- Models predicting survival after cardiac surgery ; predicting Length of stay after cardiac surgery ; predicting QoL after surgery

– Review all existing models in a particular clinical field

- e.g. all models for any CVD outcome in general population ; all developed models in obstetrics ; cancer in past 2 years

Collins et al. BMC Medicine 2011, 9:103
http://www.biomedcentral.com/1741-7015/9/103

BMC Medicine

RESEARCH ARTICLE Open Access

Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting

Gary S Collins*, Susan Mallett, Omar Omar and Ly-Mee Yu

Abstract

Background: The World Health Organisation estimates that by 2030 there will be approximately 350 million people with type 2 diabetes. Associated with renal complications, heart disease, stroke and peripheral vascular disease, early identification of patients with undiagnosed type 2 diabetes or those at an increased risk of developing type 2 diabetes is an important challenge. We sought to systematically review and critically assess the conduct and reporting of methods used to develop risk prediction models for predicting the risk of having undiagnosed (prevalent) or future risk of developing (incident) type 2 diabetes in adults.

Methods: We conducted a systematic search of PubMed and EMBASE databases to identify studies published before May 2011 that describe the development of models combining two or more variables to predict the risk of prevalent or incident type 2 diabetes. We extracted key information that describes aspects of developing a prediction model including study design, sample size and number of events, outcome definition, risk predictor selection and coding, missing data, model-building strategies and aspects of performance.

Results: Thirty-nine studies comprising 43 risk prediction models were included. Seventeen studies (44%) reported the development of models to predict incident type 2 diabetes, whilst 15 studies (38%) described the derivation of models to predict prevalent type 2 diabetes. In nine studies (23%), the number of events per variable was less than

Online First, published on December 18, 2011 as 10.1136/heartjnl-2011-300734

Prediction models for the risk of cardiovascular disease in patients with type 2 diabetes: a systematic review

S van Dieren,[1] J W J Beulens,[1] A P Kengne,[1,2,3] L M Peelen,[1] G E H M Rutten,[1] M Woodward,[3] Y T van der Schouw,[1] K G M Moons[1]

ABSTRACT

Context A recent overview of all CVD models applicable to diabetes patients is not available.

Objective To review the primary prevention studies that focused on the development, validation and impact assessment of a cardiovascular risk model, scores or rules that can be applied to patients with type 2 diabetes.

Design Systematic review.

Data sources Medline was searched from 1966 to 1 April 2011.

Study selection A study was eligible when it described the development, validation or impact assessment of a model that was constructed to predict the occurrence of cardiovascular disease in people with type 2 diabetes,

nation (ability to discriminate between patients who will get the disease and those who will not) and calibration (ability to correctly quantify the absolute risk), but the outcomes have varied widely.[7–9]

A systematic review by Chamnan *et al*[6] provides an overview of CVD prediction models that have been developed in diabetes populations, and prediction models for the general population that have been validated in a diabetes population. However, new prediction models for the diabetes population have been developed since this review, and many more prediction models exist that can be applied to people with diabetes. Moreover, it is unknown whether applying a certain prediction model in clinical prac-

Circulation
JOURNAL OF THE AMERICAN HEART ASSOCIATION

American Heart Association®
Learn and Live℠

Prediction Models for Prolonged Intensive Care Unit Stay After Cardiac Surgery : Systematic Review and Validation Study

Roelof G.A. Ettema, Linda M. Peelen, Marieke J. Schuurmans, Arno P. Nierich, Cor J. Kalkman and Karel G.M. Moons

Circulation 2010, 122:682-689: originally published online August 2, 2010
doi: 10.1161/CIRCULATIONAHA.109.926808
Circulation is published by the American Heart Association. 7272 Greenville Avenue, Dallas, TX 72514
 Print ISSN: 0009-7322. Online ISSN: 1524-4539

BMC Medical Informatics and Decision Making

Research article

Systematic review of prognostic models in traumatic brain injury

Pablo Perel*, Phil Edwards, Reinhard Wentz and Ian Roberts

Address: Nutrition and Public Health Intervention Research Unit, Epidemiology and Population Health Department, London School of Hygi & Tropical Medicine, Keppel Street, London WC1E 7HT, UK

Email: Pablo Perel* - pablo.perel@lshtm.ac.uk; Phil Edwards - phil.edwards@lshtm.ac.uk; Reinhard Wentz - reinhard.wentz@lshtm.ac.uk; Ian Roberts - ian.roberts@lshtm.ac.uk

* Corresponding author

Published: 14 November 2006

BMC Medical Informatics and Decision Making 2006, **6**:38 doi:10.1186/1472-6947-6-38

Received: 03 August 2006
Accepted: 14 November 2006

This article is available from: http://www.biomedcentral.com/1472-6947/6/38

Abstract

Background: Traumatic brain injury (TBI) is a leading cause of death and disability world-wide.

Considerations for framing the review question: CHARMS checklist Moons et al (Plos Med 2014)

Item	Comments and examples
1. Prognostic versus diagnostic prediction model	Define whether the aim is to review models to predict: •Future events: prognostic prediction models •Current (disease) status: diagnostic prediction models
2. Intended scope of the review	Define intended scope of the review and intended purpose of the models reviewed in it. Examples: •Models to inform physicians' therapeutic decision making •Models to inform referral to or withholding from invasive diagnostic testing
3. Type of prediction modelling studies	Define the type of prediction modelling studies to include. Examples of study types (Box 1): •Prediction model development without external validation in independent data •Prediction model development with external validation in independent data •External model validation, possibly with model updating
4. Target population to whom the prediction model applies	Define the target population relevant to the review scope. Examples: •Women with diagnosed breast cancer •Healthy adult men in the general population
5. Outcome to be predicted	Define the outcome of interest to be predicted: •Specific future event, such as a fatal or non-fatal coronary heart disease •Specific diagnostic target disease, such as presence of lung embolism
6. Time span of prediction	Define over what specific time period the outcome is predicted (prognostic models only). Example: •Event within a specific time interval, such as event within 3 months, 1 year or 10 years
7. Intended moment of using the model	The systematic review may focus on models to be used at a specific moment in time. Examples: •Models to be used at the moment of diagnosis of a particular disease •Models to be used preoperatively to predict the risk of postoperative complications •Models to be used in asymptomatic adults to detect undiagnosed diabetes mellitus type 2

Well-formulated review question (PICO)

- Considerations for framing review aim (based on CHARMS checklist)

PICOTS

Population	Define target population to whom the prediction model applies
Intervention	Define type of prediction modelling studies to include *(note different modeling phases)* Define intended scope of SR and intended purpose of the models reviewed *(see different aims)*
Comparator	Different models for predicting the same outcome (or non-issue)
Outcomes	Define outcome to be predicted *(e.g. single or composite outcome)*
Timing	Define over what specific time period the outcome is predicted
Setting	Intended moment of using the prediction model; Who will use the model; how will it be used

BMC Medical Informatics and Decision Making

BioMed C

Research article

Open Ac

Systematic review of prognostic models in traumatic brain injury

Pablo Perel*, Phil Edwards, Reinhard Wentz and Ian Roberts

Exercise: define review question + PICOTS (see nex slide)

Recall: Well-formulated review question (PICO)

Population	Define target population to whom the prediction model applies
Intervention	Define type of prediction modelling studies to include *(note different modeling phases)* Define intended scope of SR and intended purpose of the models reviewed
Comparator	Different models for predicting same outcome (or non-issue)
Outcomes	Define outcome to be predicted *(e.g. single or composite outcome)*
Timing	Define over what specific time period the outcome is predicted
Setting	Intended moment of using the prediction model; Who will use the model; how will it be used

Practice question: Examples of PICOTS

Population	*Examples* • Women with diagnosed breast cancer • Healthy adult men in the general population
Intervention	*Examples* • Prediction model development without external validation • Prediction model development with external validation in independent data • External model validation, possibly with model updating *Examples* • Models to inform physicians' therapeutic decision making • Models to inform referral to or withholding from invasive diagnostic testing
Outcomes	*Examples* • Specific future event, such as a fatal or non-fatal coronary heart disease
Timing	*Example* • Event within a specific time interval, such as event within 3 months, 1 year or 10 years
Setting	*Examples* • Models used preoperatively to predict risk of postoperative complications • Models used in asymptomatic adults to detect undiagnosed DM2

Conducting a systematic reviews: Step 2

SR elements	Intervention	Prognosis
Well formulated question		
Extensive search	Use of methodological filter Study design: RCT, observational studies (cohort, case control)	Filters of limited use Study design: RCT, observational (cohort, case control)
Selection & Data extraction		
Critical appraisal of methodological quality		
Data syntheses (meta-analysis)		

Extensive search

- A single study can address various aims
  - Development one or more model
  - Validation/comparison different models
  - Incremental predictive value of a specific factor/marker
  - Models for various outcomes

- Addressed by RCTs, large cohort studies, nested case-control studies, registry studies

Identifying relevant publications

- No optimal, reliable methods for searching the literature for prognostic information
  - As for RCTs and Diagnostic Test Accuracy Studies

- A few published
  - Altman DG (2001): single prognostic factors
  - Wong SS (2003): very generic
  - Ingui BJ (2001): prediction models
  - Geersing (2012): validation Ingui (2001) and updated (new) search strategy

Search Filters for Finding Prognostic and Diagnostic Prediction Studies in Medline to Enhance Systematic Reviews

Geert-Jan Geersing[1*☯], Walter Bouwmeester[1☯], Peter Zuithoff[1], Rene Spijker[2,4], Mariska Leeflang[3,4], Karel Moons[1]

1 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands, **2** Medical Library Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands, **3** Department of Clinical Epidemiology and Bio-Informatics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands, **4** Dutch Cochrane Center, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

Table 1. Search strategies for finding prediction research in Medline.

Filter	Search terms included in the filter*	Sensitivity# (95% CI)	Specificity# (95% CI)
Ingui filter	(Validat$ OR Predict$.ti. OR Rule$) OR (Predict$ AND (Outcome$ OR Risk$ OR Model$)) OR ((History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$) AND (Predict$ OR Model$ OR Decision$ OR Identif$ OR Prognos$)) OR (Decision$ AND (Model$ OR Clinical$ OR Logistic Models/)) OR (Prognostic AND (History OR Variable$ OR Criteria OR Scor$ OR Characteristic$ OR Finding$ OR Factor$ OR Model$))	0.98 (0.92–1.0)	0.86 (0.85–0.87)
Haynes broad filter	(Predict*[tiab] OR Predictive value of tests[mh] OR Scor*[tiab] OR Observ*[tiab] OR Observer variation[mh])	0.96	0.79

*Using the Pubmed interface for MEDLINE.

#Sensitivity and specificity as reported by Ingui and Haynes in their original publication; CI = confidence interval, for the Haynes broad filter no confidence intervals were given in the original publication.

doi:10.1371/journal.pone.0032844.t001

Geersing et al 2012
Conclusions

- Available search strategies for prediction research good in retrieving "Prediction modelling studies (Se 0.78 to 0.89)

- Less value in retrieving "Predictor Finding" and "Prediction Model Impact Studies"

Table 4. Updated search string for finding prediction research.

"Stratification" OR "ROC Curve"[Mesh] OR "Discrimination" OR "Discriminate" OR "c-statistic" OR "c statistic" OR "Area under the curve" OR "AUC" OR "Calibration" OR "Indices" OR "Algorithm" OR "Multivariable"

doi:10.1371/journal.pone.0032844.t004

- Strategy for "Predictor Finding" studies still sub-optimal

Recommendations for SRs (Geersing 2012)

0. define aim of SR (model development, validation, updating, predictor finding, model impact studies):

1. perform Medline based search using a search strategy, e.g. Geersing or Ingui strategy, combined with subject matter queries;
2. search other databases, such as EMBASE, Cochrane, or even a "Google (Scholar) search";
3. perform cross-reference checking search on all retrieved articles;
4. contact experts in the field for additional articles;
5. be transparent in flow-chart figure (# papers found by search versus hand versus experts etc.)

Ad. Reporting biases: trials

- statistically significant, 'positive' results more likely to be published…

- …therefore more likely to be included in your review
  - leads to exaggerated effects
  - large studies likely to be published anyway, so small studies most likely to be affected towards positive results
  - non-significant results are as important to your review as significant results

Question: What are possible effects of reporting bias in SR of prognosis studies?

Answer

- Many reports support only when a predictor was indeed associated with the outcome
- It is difficult to find reports where the predictor was not significant in the multivariable model
- Seldom see a model published with poor predictive accuracy
  - Perhaps validation studies of existing (well known) model

Conducting a systematic reviews: Step 3

SR elements	Intervention	Prognosis
Well formulated question		
Extensive search		
Selection & Data extraction	Two persons, independently; clear criteria; documentation	
Critical appraisal of methodological quality		
Data syntheses (meta-analysis)		
Interpretation of results		

Study selection

- Selecting studies involves judgement, and is highly influential on the outcomes of the review

- Two (or more…) reviewers, independently
  - minimizing bias
  - pilot selection on a few papers first: substantial variation
  - how will disagreements be managed

- Examine titles and abstracts
  - often many hits >10.000 hits

- Retrieve and examine full text reports
  - time consuming

Data extraction: information required about:

- Flow of included/excluded studies
  - References
  - Author contact details

- Extraction of characteristics/data of included studies + Critical appraisal
  - PICOTS elements → **CHARMS** (Plos Med 2014)

Conducting a systematic reviews: Step 5

SR elements	Intervention	Prognosis
Well formulated question		
Extensive search		
Selection & Data extraction		
Critical appraisal quality + Risk of Bias assessment	Main items: randomization; allocation concealment, outcome assessment, dropouts/withdrawals	CHARMS (2014) PROBAST (2015) Main items: Same + prediction modeling specific issues
Data syntheses (meta-analysis)		

Critical Appraisal + RoB: instruments

- Reviewers usually developed their own (or do not appraise at all)
  - Hayden (Ann Int Med 2006+2013)→ 163 SRs about prognosis
  - Less than 50% adequately addressed/checked relevant sources of bias
- Prognostic factor/predictor finding studies
  - RoB tool : **QUIPS** → J Haydn, Ann Int Med 2006 + 2013
- Prediction modelling (development and validation)
  - Critiical Appraisal: **CHARMS** -- Plos Med 2014 – Review design formulation + data extraction
  - RoB: **PROBAST** – 2015/2016 → **other Workshop**

Critical Appraisal

Domains in CHARMS checklist

Domains to extract in SRs of Prediction Modeling studies

- Source of data
- Participants
- Outcome(s) to be predicted
- Candidate predictors (or index tests)
- Sample size
- Missing data
- Model development
- Model performance
- Model evaluation
- Results

Critical Appraisal

Key issues in CHARMS checklist

- Source of data
  - e.g. cohort, case-control, randomized trial participants, or registry data

- Participants
  - Definition and method for measurement of outcome
  - Was the same outcome definition (and method for measurement) used in all patients?
  - Type of outcome (e.g. single or combined endpoints)
  - Was the outcome assessed without knowledge of the candidate predictors (i.e. blinded)?
  - Were candidate predictors part of the outcome (e.g. in panel or consensus diagnosis)?
  - Time of outcome occurrence or summary of duration of follow-up

Critical Appraisal

Key issues in CHARMS checklist

- Outcome(s) to be predicted
  - Number and type of predictors (e.g. demographics, patient history, physical examination, additional testing, disease characteristics)
  - Definition and method for measurement of candidate predictors
  - Timing of predictor measurement (e.g. at patient presentation, at diagnosis, at treatment initiation)
  - Were predictors assessed blinded for outcome, and for each other (if relevant)?
  - Handling of predictors in the modelling (e.g. continuous, linear, non-linear transformations or categorised)

- Candidate predictors (or index tests)
  - Number of participants and number of events
  - Number of events in relation to the number of candidate predictors (Events Per Variable)

Critical Appraisal

Key issues in CHARMS checklist

- Missing data
  - Number of participants with any missing value (include predictors and outcomes)
  - Number of participants with missing data for each predictor
  - Handling of missing data (e.g. complete-case analysis, imputation, or other methods)

- Model development
  - Modelling method (e.g. logistic, survival, neural networks, or machine learning techniques)
  - Modelling assumptions satisfied
  - Method for selection of predictors for inclusion in multivariable modelling (e.g. all candidate predictors, pre-selection based on unadjusted association with the outcome)
  - Method for selection of predictors during multivariable modelling (e.g. full model approach, backward or forward selection) and criteria used (e.g. p-value, Akaike Information Criterion)
  - Shrinkage of predictor weights or regression coefficients (e.g. no shrinkage, uniform shrinkage, penalized estimation)

Critical Appraisal

Key issues in CHARMS checklist

- Model performance
  - Calibration (calibration plot, calibration slope, Hosmer-Lemeshow test) and Discrimination (C-statistic, D-statistic, log-rank) measures with confidence intervals
  - Classification measures (e.g. sensitivity, specificity, predictive values, net reclassification improvement) and whether a-priori cut points were used

- Model evaluation
  - Method used for testing model performance: development dataset only (random split of data, resampling methods e.g. bootstrap or cross-validation, none) or separate external validation (e.g. temporal, geographical, different setting, different investigators)
  - In case of poor validation, whether model was adjusted or updated (e.g. intercept recalibrated, predictor effects adjusted, or new predictors added)

Critical Appraisal

Key issues in CHARMS checklist

- Results
  - Final and other multivariable models (e.g. basic, extended, simplified) presented, including predictor weights or regression coefficients, intercept, baseline survival, model performance measures (with standard errors or confidence intervals)
  - Any alternative presentation of the final prediction models, e.g. sum score, nomogram, score chart, predictions for specific risk subgroups with performance
  - Comparison of the distribution of predictors (including missing data) for development and validation datasets

Conducting a systematic reviews: Step 6

SR elements	Intervention	Prognosis
Well formulated question		
Extensive search		
Selection & Data extraction		
Critical appraisal of methodological quality		
Data syntheses (meta-analysis)	Effect measures: RD, RR, mean difference	Various effect measures: C-index ; calibration statistics; NRI
Interpretation of results		

Meta-analysis of prediction models

Two types

1. **In case no own (validation) IPD set – aggregate data only: 2 cases**
    1. MA of a specific prediction model across multiple 'model-validation-studies'
    2. MA of a specific predictor when added to a specific model across multiple 'added-value-studies'

2. **In case own (validation) IPD set – combination of aggregate and IPD**

Other PMG Workshops

Conducting a systematic reviews: Step 7

SR elements	Intervention	Prognosis
Well formulated question		
Extensive search		
Selection & Data extraction		
Critical appraisal quality + Risk of Bias		
Data syntheses (meta-analysis)		
Interpretation of results	Effective and relevant?	Average Model Performance; Average added value; Which model predicts best?

Interpretation of results

CHARMS checklist

- Interpretation of presented models
  - confirmatory, i.e. model useful for practice or exploratory, i.e. more research needed
- Comparison with other studies, discussion of generalizability of model; strengths and limitations.

Take home message

- To know the main types of ‘prediction studies’
- To understand the different aims of systematic reviews of prediction modeling studies
- To describe the similarities and differences between intervention and prediction modeling reviews
- To understand the challenges of reviews of prediction modeling studies
- To develop structure of a protocol for a systematic review of prediction modeling studies (except for the MA part – next workshop)

Handy Tools/Papers

- CHARMS paper
- TRIPOD paper
- PROBAST –Robert Wolff
- Cochrane PMG protocol template
  - PMG Coordinator: Alexandra Hendry (Alexandra.Hendry@sswahs.nsw.gov.au)

Workshop aftercare

- Questions about workshop?
- Assistant needed with review of studies of prognosis studies?
- Please contact:
  - PMG Coordinator: Alexandra Hendry (Alexandra.Hendry@sswahs.nsw.gov.au)
  - PMG Co-convenor: Karel Moons (K.G.M.Moons@umcutrecht.nl)