UMC Utrecht

Prediction models: Shall we stop developing them?

Karel GM Moons

UMC Utrecht

Julius Center for Health Sciences and Primary Care

Cochrane Netherlands

EiC: BMC Diagnostic & Prognostic Research

www.juliuscenter.nl

Experiences from various consortia:

– BMJ 2009 – series on prognostic modelling

– Plos Med + BMJ 2013 -- PROGRESS series

– HEART 2012 2 papers

– TRIPOD guideline 2015

– PROBAST 2019

PROGNOSIS RESEARCH IN HEALTHCARE
Concepts, Methods, and Impact

Edited by
Richard D. Riley
Danielle van der Windt
Peter Croft
Karel G.M. Moons

Paperback | 9780198796619
January 2019 | 372 pages
~~£44.99~~ **£31.49** | ~~$65.00~~ **$45.50**

Prediction

- Prediction = foreseeing / foretelling

  … (the probability) of something that is yet unknown

- In medicine:

  1. Probability of a future event/outcome = **prognosis**

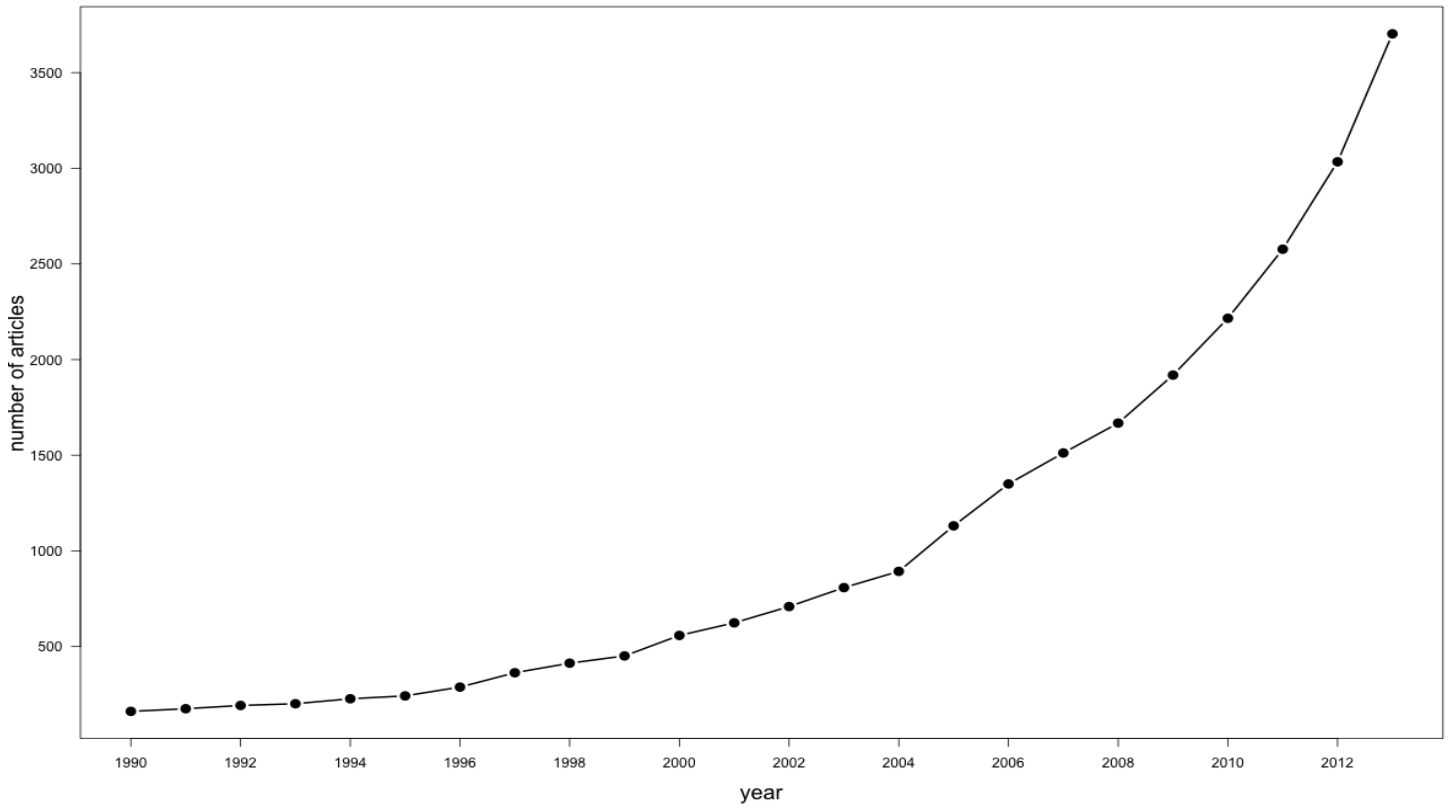
  2. Probability of the result of a more invasive/costly reference (gold) standard that is not yet done = **diagnosis**

What is a prediction model?

Any combination >= 2 predictors (variables/covariates/ determinants)
→ convert observed values to absolute probability…

- … of having a particular disease/disorder → **diagnosis**
- … of developing particular event/outcome within a certain time (hours, days, weeks, years)→ **prognosis**
  - Not necessarily patients –- subjects at risk of developing outcome

Prediction modelling is hot!



- 100,000s (!) prediction models
- Numerous models for same outcome or target population

Systematic reviews of prediction models

- >110 models for prostate cancer (Shariat 2008)
- >100 models for Traumatic Brain Injury (Perel 2006)
- 83 models for stroke (Counsell 2001)
- 54 models for breast cancer (Altman 2009)
- 43 models for type 2 diabetes (Collins 2011; Dieren 2012)
- 31 models for osteoporotic fracture (Steurer 2011)
- 29 models in reproductive medicine (Leushuis 2009)
- 26 models for hospital readmission (Kansagara 2011)
- >25 models for length of stay in cardiac surgery (Ettema 2010)

- >350 models for prediction of CVD outcomes in general population (Damen, BMJ 2016)

Wilde voorspellingen over uw gezondheid

Er is een overweldigend aanbod aan **apps en sites** die voorspellen of iemand gezondheidsrisico's loopt. Of die hout snijden? Wie zal het zeggen.

Door **Laurens Verhagen** Foto **Marloes Haarmans**

Wat is de kans dat u longkanker krijgt? Of prostaatkanker. Een hartinfarct? Sites als Your Disease Risk geven hier antwoord op. Op de site moeten bezoekers antwoord geven op een aantal vragen. Na wat inleidende vragen over leeftijd, geslacht, gewicht en medische voorgeschiedenis stelt de site een aantal vragen over roken. Rookt u? Bent u gestopt? Wanneer was dat? En hoeveel sigaretten rookte u vroeger dan? Dan zijn de eetgewoonten aan de beurt. Eet u vaak verwerkt vlees? En hoe zit dat met noten, kaas, rood vlees, vers fruit, kant-en-klare pastasaus of vis? O ja, slikt u vaak aspirine of multivitamines? Beweegt u wel genoeg en hoeveel alcohol nuttigt u? Als de rij vragen is doorgeakkerd, geeft Your Disease Risk een prognose. Niet in harde percentages, maar in termen als 'meer dan gemiddeld'.

Carl Moons, hoogleraar epidemiologie aan het UMC Utrecht, waarschuwt voor dit soort sites waarin medische voorspelmodellen worden gebruikt. Niet specifiek voor Your Disease Risk trouwens, maar vooral voor de wildgroei aan vergelijkbare apps en sites. Volgens Moons zijn er wereldwijd meer dan honderdduizend apps en websites die dergelijke modellen exploiteren, vaak gratis beschikbaar. 'We hebben geen idee of al die modellen die zo toegankelijk geëxploiteerd worden accuraat zijn', zegt Moons. Intussen kan iedereen zo'n site uit de grond stampen. De achterliggende predictiemodellen liggen namelijk voor het opscheppen in de medische literatuur.

Volgens Moons zijn er zo'n veertigduizend biomedische tijdschriften en per jaar worden vier miljoen papers gepubliceerd. Daarbinnen is, in het tijdperk van big data, machine learning en kunstmatige intelligentie, het ontwikkelen van predictiemodellen heel hip. 'Maar als je naar de kwaliteit van die modellen kijkt, schrik je je vaak te pletter.'

Vaak gaat het in de basis al mis, stelt Moons: niet de juiste patiënten, of veel te weinig. 'Geen ijkpunten, veel te kleine datasets, geen deugdelijke statistische analyse: je komt echt alles tegen.' Het probleem is de evaluatie van de modellen. Of liever gezegd: de afwezigheid ervan. In tegenstelling tot bij medicijnen is zo'n gedegen validatie namelijk niet verplicht. 'Als consument of zorgverlener moet je zeker weten dat de modellen kloppen. Maar er is geen instantie die zich daarmee bezighoudt.' Met alle gevolgen van dien. Het is soms echt 'een drama', zegt Moons. Zijn onderzoeksteam doet al jaren onderzoek naar de kwaliteit en rapportage van voorspelmodellen.

KAMPIOEN GOOGLEN OP GEZONDHEID

Nederland heeft van alle landen uit de Europese Unie het hoogste percentage dat informatie opzoekt over gezondheid op internet. In 2018 had 72 procent van de Nederlanders tussen 16 en 74 dat gedaan blijkt uit recente cijfers van het CBS. Het gemiddelde van de Europese Unie was 52 procent.

huisarts. Deze laagdrempeligheid heeft voor- en nadelen, stellen experts. 'Onbehandeld atriumfibrilleren (het onregelmatig kloppen van het hart, *red.*) is een reëel risico voor de gezondheid. Dus als je er hiermee mensen kunt uitvissen die normaal niet doorhebben dat er iets aan de hand is, is dat pure winst', zegt cardioloog Kraaijenhagen. Maar, zegt hij ook: de grote vraag hierbij is hoeveel overdiagnose en extra zorgconsumptie dit oplevert. Kloppen bezorgde burgers straks in groten getale bij de huisarts aan? 'Dat moet de komende tijd gaan uitwijzen.' Ook Joris de Groot, hoogleraar elektrofysiologie van het hart aan het Amsterdam UMC, houdt een slag om de arm: 'Gaan er echt levens gered worden met dit soort nieuwe technologieën? Het antwoord daarop is sterk afhankelijk van welke groep patiënten wordt onderzocht, en hoe hoog het risico in die groep is.' Oftewel: een grote groep jongeren met een Apple Watch om de pols zal niet zo geholpen zijn met een hartfilmpje. 'Als de kans op risico vooraf al heel klein is, dan kun je nog zo'n goed model hebben, maar win je er weinig mee.' De Groot gelooft wel dat zelfdetectie de toekomst heeft, mits toegepast bij de juiste groep. En daarna komt ook bij hem de vraag op: ontstaat er geen overconsumptie?

Moons zoekt de oplossing van het probleem niet bij de appmakers, maar bij de wetenschappelijke studies: 'Je moet zeker weten dat de modellen kloppen. Dan kunnen foute modellen ook niet in apps terechtkomen. We moeten ophouden met wéér een predictiemodel te maken, alleen omdat het kan.' Daarom heeft Moons recentelijk met een grote internationale onderzoeksgroep een eenvoudige checklist ontwikkeld om medische voorspelmodellen op hun kwaliteit en effectiviteit te beoordelen voordat ze gebruikt worden in de praktijk. Deze Probast-checklist bestaat uit twintig items waarop onderzoekers, zorgverleners of appontwikkelaars snel kunnen bepalen hoe bruikbaar een voorspelmodel echt is.

Een ander probleem is het vervolg: 'Welke consequenties trekt de gebruiker uit een getoonde uitkomst? Wat zegt het mij dat ik een kans van 9 procent heb op hart- en vaatziekten? Is dat veel, is dat weinig? Moet ik ermee naar de arts?' Het risico van overdiagnose ligt levensgroot op de loer.

Een van de grote Nederlandse partijen die medische predictiemodellen ontwikkelt is Niped. Dit kennisinstituut voor e-health en preventie werd in 2005 mede opgericht door cardioloog Roderik Kraaijenhagen. Hij herkent de zorgen van Moons: 'Die zijn terecht. Het is precies de reden dat we Niped ooit oprichtten.' De door Niped ontwikkelde site Persoonlijke Gezondheidscheck is 'het wetenschappelijke antwoord op het versnipperde aanbod van losse, vaak onbetrouwbare gezondheidstesten en preventieve onderzoeken'.

Hieraan heeft zich een flink aantal partijen verbonden, waaronder het Nederlands Huisartsen Genootschap. Zo'n driehonderdduizend mensen hebben deze onlinecontrole gedaan. De laatste tijd neemt het een vlucht, zegt Kraaijenhagen, maar het is pas het begin: 'We hebben met het ministerie van VWS afgesproken dat over twee jaar 10 procent van de bevolking de check moet hebben gedaan, dus we hebben nog een lange weg te gaan.' Verder is het volgens Kraaijenhagen van belang na te denken wat er na het invullen van een onlinevragenlijst gebeurt. Een van de ideeën is tussen consument en huisarts een instantie te zetten die een uitslag goed kan interpreteren en ervoor kan zorgen dat de zorg niet onnodig wordt belast.

Why using prediction models?

- … Not meant to replace physicians, but to complement their clinical intuition!!!!!!

- Assumption:
  - They provide accurately + objectively estimated probabilities…
  - …to improve medical decision making …
  - … and thus subject's outcomes
  - … and thus cost-effectiveness of health care

What evidence do we need before using prediction models?

4 steps in prediction modelling

BMJ series 2009; HEART series 2012; PROGRESS series BMJ + PLOS MED 2013, TRIPOD Ann Intern Med 2015

- 1. Developing prediction model from a particular (your) dataset
- 2. Validate/test the predictive accuracy of previously developed model in (data of) other subjects
- 3. Adjust/tailor model to local situation/care setting using the validation dataset
- 4. Quantify impact of using a model on decision making and patient outcomes

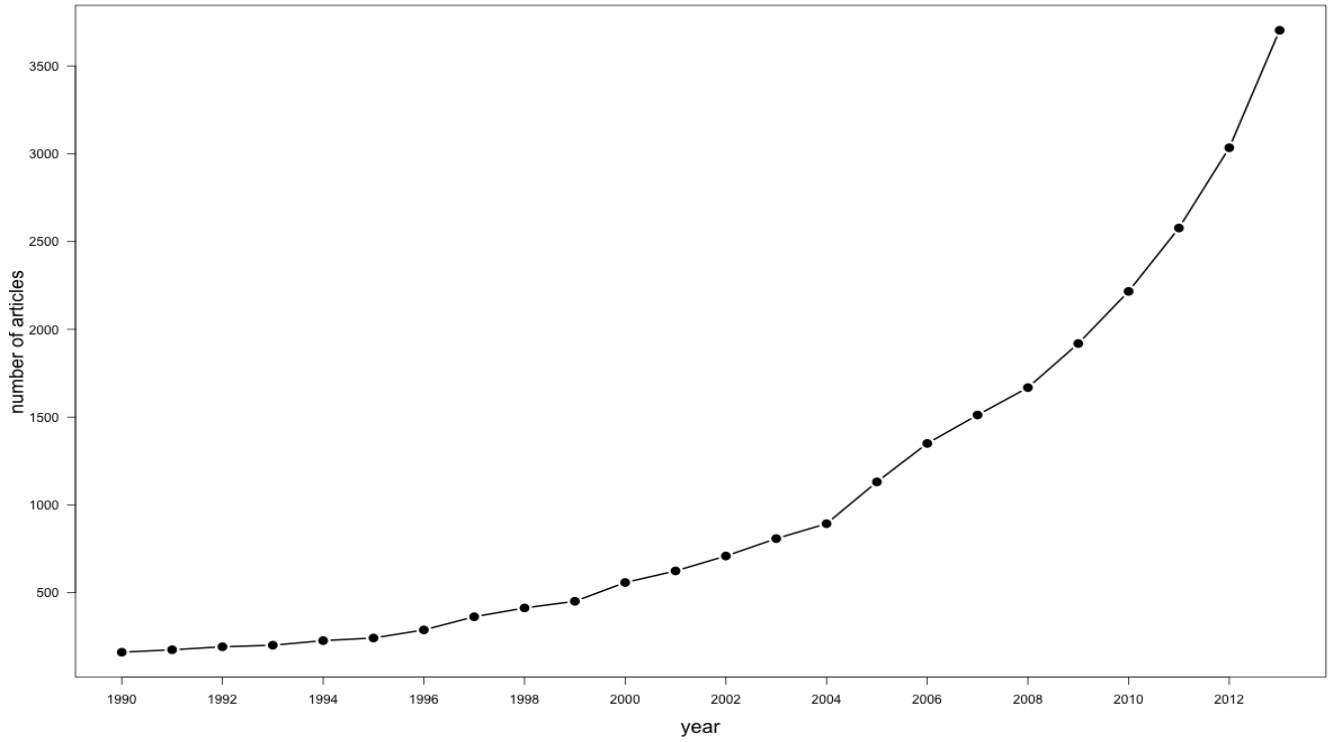
1. Developing a prediction model from your dataset

Don't

Don't develop a model from your data – skip this phase

1. **Suppress your reflex**
   - Hard: we finally learned 'trics' to develop models (standard software)
   - 'Own' model makes us famous (Apgar; Goldman; Gail; Wells)
     - Validation of somebody else's model is only to support citation index of others

Prediction modelling is hot!



Majority is newly developed models – few validation studies

Numerous systematic Reviews

- Regardless clinical domain: numerous models developed → few validated
- Too much focus on developing → hardly on validation
- Like biomarker world: discovery driven → validation uninteresting ('losers'/non-innovative)
- But: with all these models for same outcome or target population: we/professionals have 'no clue' which model to use in which situation
  - Is our healthcare better of with yet another developed model?

So when we are behind our dataset and aimed to develop a prediction model

… Starts with …

…NOT developing a model…

… First search, review and validate existing models for your domain, target population or outcome at interest

When behind our dataset and aimed to develop a prediction model

- There are (almost) always existing models that apply to your patient population/outcome
  - We hardly search for existing models to first test on our datasets
  - We rather pursue to develop yet another (own) model

- Test and directly compare (!) the predictive performance of these models on your data set = **comparative (external) validation**

BMJ 2012;344:e3186 doi: 10.1136/bmj.e3186 (Published 24 May 2012)

EDITORIALS

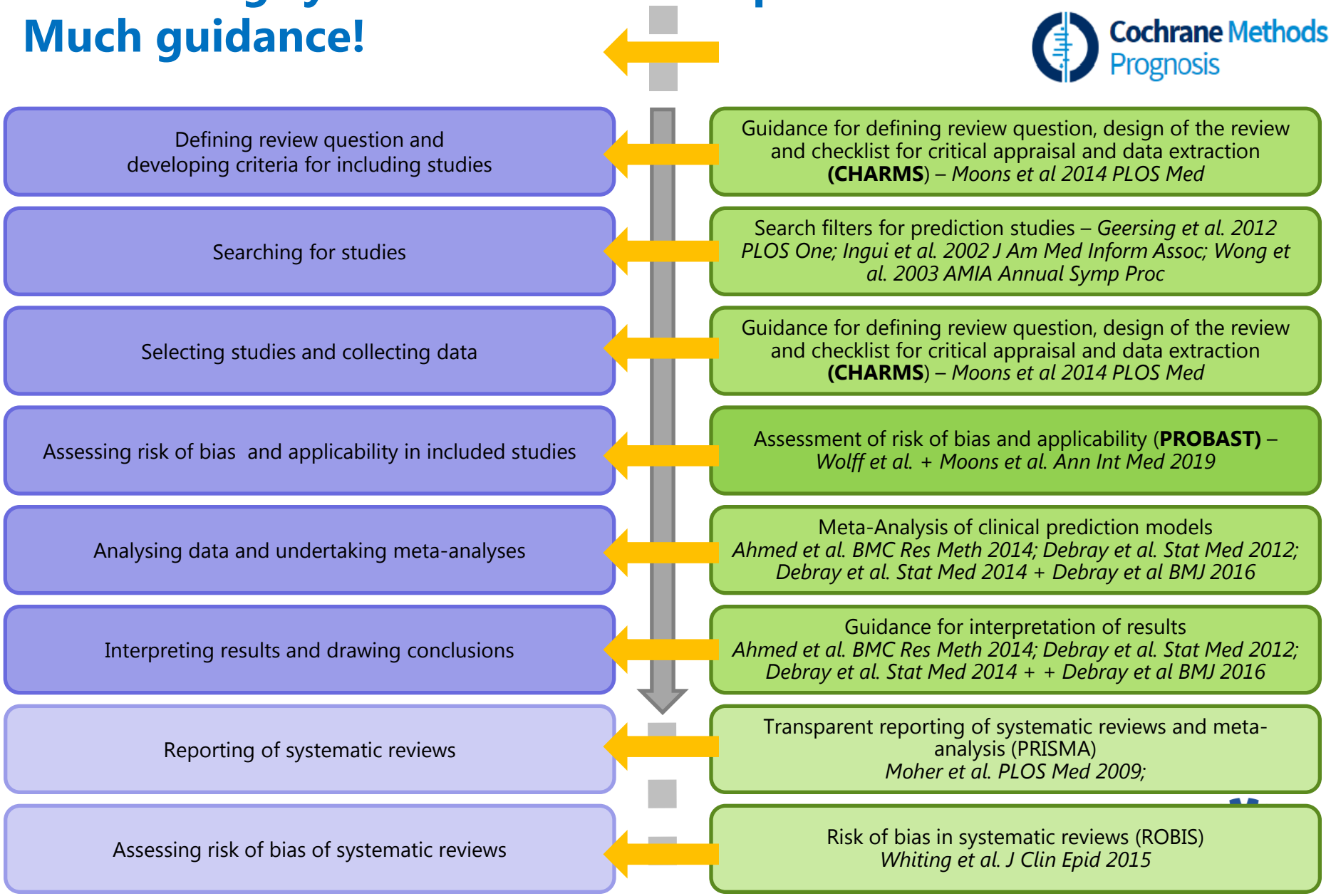
Gary S Collins *senior medical statistician*[1],

Comparing risk prediction models

Should be routine when deriving a new model for the same purpose

Conducting systematic reviews of prediction model studies
Much guidance!

Cochrane Methods Prognosis



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

Guidelines and Guidance

Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies: The CHARMS Checklist

Karel G. M. Moons[1¶*], Joris A. H. de Groot[1¶], Walter Bouwmeester[1], Yvonne Vergouwe[1], Susan Mallett[2], Douglas G. Altman[3], Johannes B. Reitsma[1], Gary S. Collins[3]

OPEN ACCESS Freely available online

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],

Annals of Internal Medicine RESEARCH AND REPORTING METHODS

PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration

www.probast.org

Karel G.M. Moons, PhD*; Robert F. Wolff, MD*; Richard D. Riley, PhD; Penny F. Whiting, PhD; Marie Westwood, PhD; Gary S. Collins, PhD; Johannes B. Reitsma, MD, PhD; Jos Kleijnen, MD, PhD; and Sue Mallett, DPhil

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

BMJ 2017

Thomas P.A. Debray†, Johanna A. A. G. Damen†, Kym I. E. Snell, Joie Ensor, Lotty Hooft, Johannes Reitsma, Richard D. Riley†, Karel G. M. Moons†

Meta-analysis and aggregation of multiple published prediction models

Thomas P. A. Debray,[a*†] Hendrik Koffijberg,[a] Daan Nieboer,[b] Yvonne Vergouwe,[b] Ewout W. Steyerberg[b] and Karel G. M. Moons[a]

Cochrane Netherlands

Cochrane Methods Prognosis

Basic & Advanced Courses

MSC EPIDEMIOLOGY

Systematic Reviews, Meta Analysis

Face to Face & Online

Accessible from all over the world
More than 50 courses…

- Systematic Reviews of Randomised Intervention Studies
- Systematic Reviews of Diagnostic Studies
- Systematic Reviews of Prognostic Studies
- Meta Analysis with Individual Participants Data
- ….and many more

www.msc-epidemiology.eu **www.msc-epidemiology.online**

UMC Utrecht Julius Center

Universiteit Utrecht

elevate

Hence …

… prognostic/prediction studies are hot

… SR's and notably MA of prognostic studies as well

- highly desired and well received by journals/policy makers →

- it is time to first systematically summarise existing evidence in your field before developing 'your own model'

You are still behind your dataset and aimed to develop a prediction model

You have done your review

Selected the (most) relevant models for your interest

Published your review in a MAJOR journal

(Most prediction model papers do not appear in such journals!)

And then…..

What evidence do we need before using prediction models?

4 Steps in prediction modelling

BMJ series 2009; HEART series 2012; PROGRESS series BMJ + PLOS MED 2013, TRIPOD Ann Intern Med 2015

- 1. Developing prediction model from a particular dataset

- 2. Validate/test predictive accuracy of previously developed model in your data

- 3. Adjust/tailor model to local situation/care setting using the validation dataset

- 4. Quantify impact of using a model on decision making and patient outcomes

Validating

- Test and directly compare (!) the predictive performance of the selected models, on your data set = (external) validation

BMJ 2012;344:e3186 doi: 10.1136/bmj.e3186 (Published 24 May 2012)

EDITORIALS

Gary S Collins *senior medical statistician*[1],

Comparing risk prediction models

Should be routine when deriving a new model for the same purpose

2. Model validation studies: Don'ts

BMJ series 2009; HEART series 2012; PROGRES series BMJ + PLOS MED 2013

- Aim: to demonstrate predictive performance of competing models in (data of) subjects that were not used to develop model – direct comparison!
  - Calibration, discrimination, (re)classification

- Validating model(s) is not …
  - …Repeat one's analysis in your data → to check whether you find same predictors, regression coefficients, predictive performance

  or

  - …Fit the previously found predictors and compare your performance with performance in development set

2. Model validation studies: Do's

BMJ series 2009; HEART series 2012; PROGRES series BMJ + PLOS MED 2013

- Use original developed model → apply 'as is' to your data → compare predicted with observed outcomes
  - Discrimination, calibration and (re)classification

- Validation studies require that developed prediction models properly reported
  - Original beta's – plus intercept / baseline hazard
    - Not just simplified score (too often done)
  - Clear definition and measurement method of predictors + outcome
  - Someone can indeed validate and use the model

Annals of Internal Medicine RESEARCH AND REPORTING METHODS

Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement

Gary S. Collins, PhD; Johannes B. Reitsma, MD, PhD; Douglas G. Altman, DSc; and Karel G.M. Moons, PhD *Ann Intern Med.* 2015;162:55-63. doi:10.7326/M14-0

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

www.tripod-statement.org

Types of Validation Studies

BMJ series 2009; HEART series 2012; PROGRES series BMJ + PLOS MED 2013

1. Temporal validation

- Often same setting, measurement methods, investigators only later in time
  - Many similarities → very 'high' chance of good performance
- If large dataset: Split over time
- Don't randomly split – no difference but chance

Types of Validation Studies

BMJ series 2009; HEART series 2012; PROGRES series BMJ + PLOS MED 2013

2. Geographic validation

- Validation in other centers/region; often other investigators
- Often other measurement methods
- If multicenter or combination of datasets (= IPD meta analysis)
  - split sample by center/region

3. Setting/domain/subgroup validation

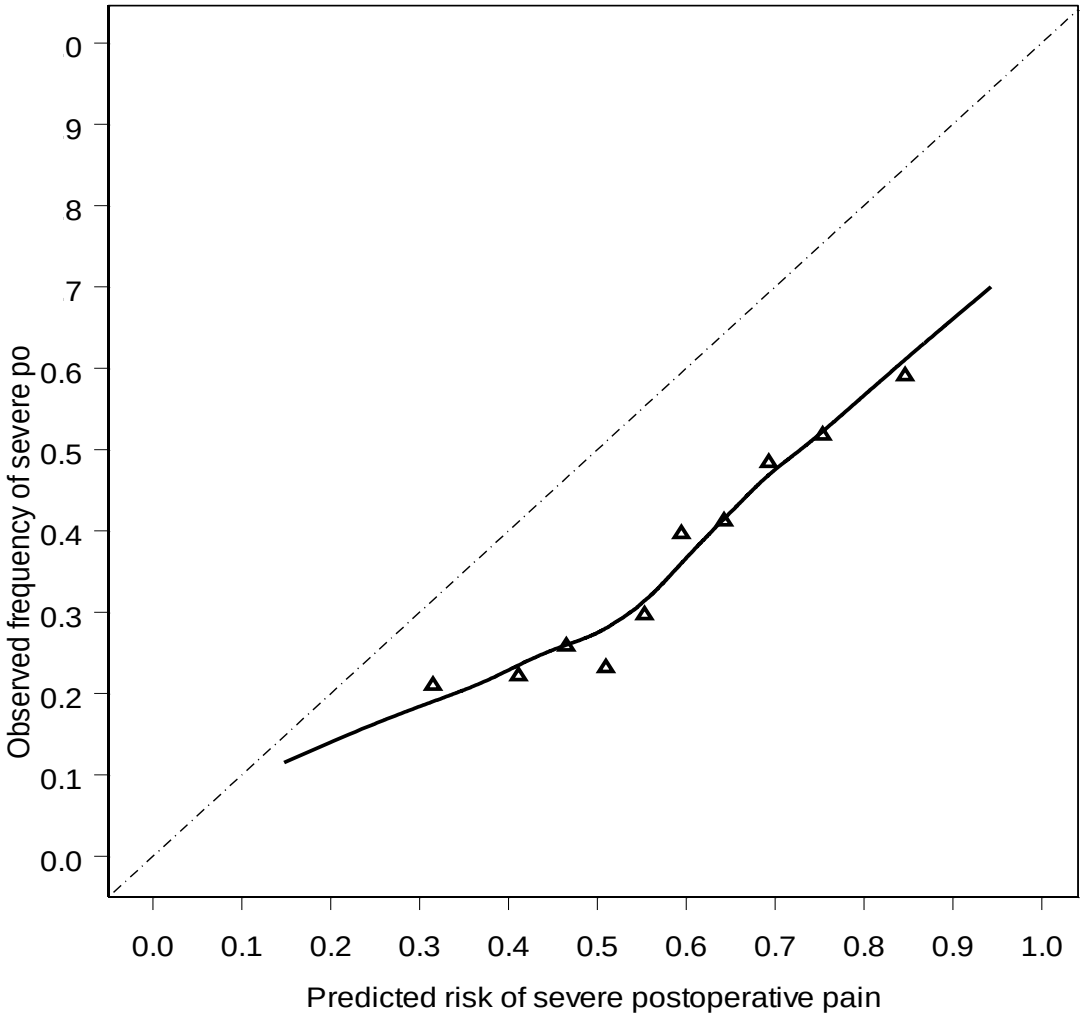
- Secondary → primary care
- Adults → children
- Men → women
- first VT → recurrent VT

Types of Validation Studies

BMJ series 2009; HEART series 2012; PROGRES series BMJ + PLOS MED 2013

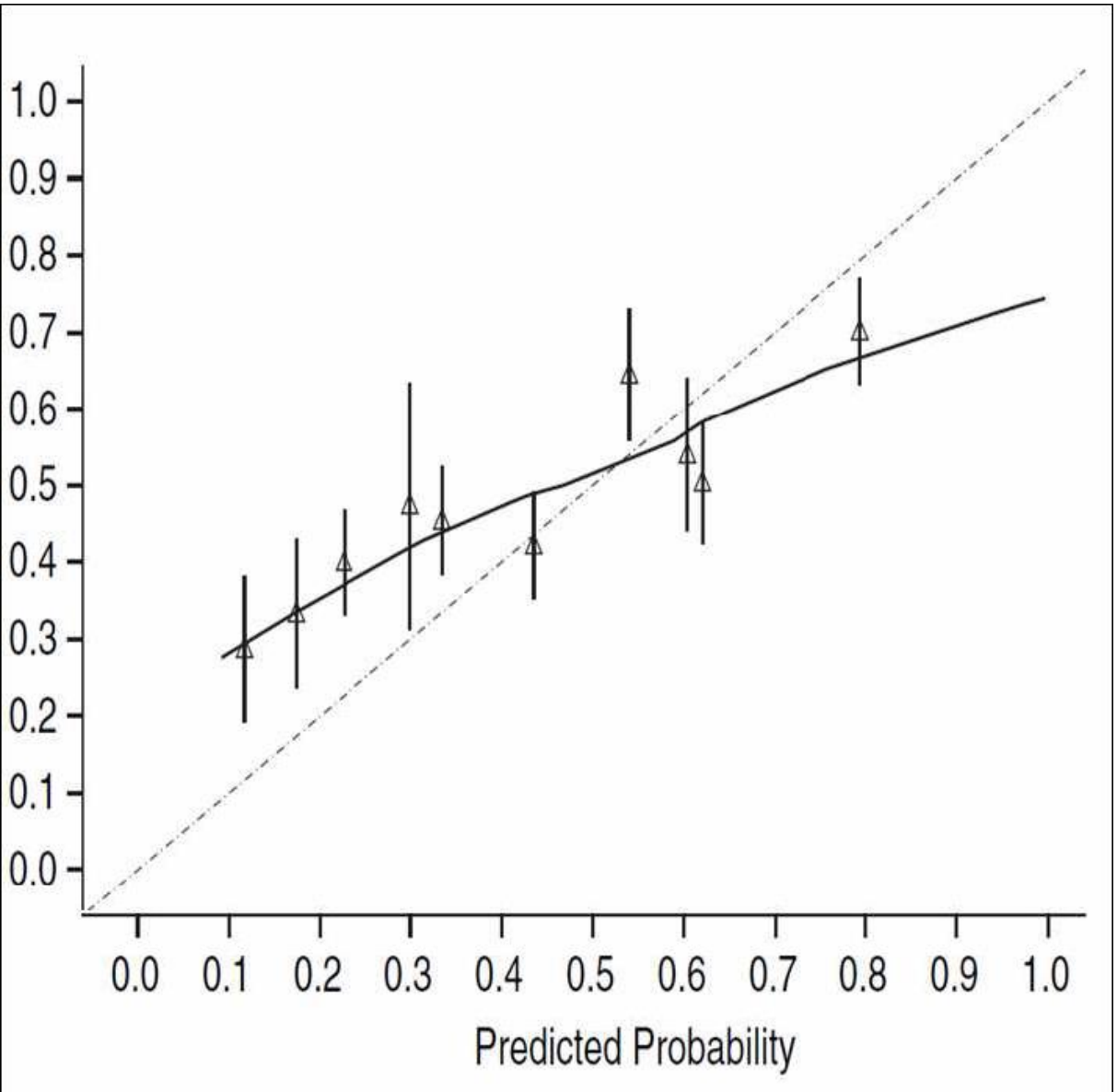
- **Aim of validation studies is not to find similar predictive accuracy as in development set…**
  - But to find satisfactory performance in validation set
  - AUC of 0.60 is not per se bad
    - Depends on accepted consequences of false predictions/decisions
    - You can always find low or high risk group –- despite small
- **YES: commonly find poorer performance when validating existing model in your data**
  - Still suppress reflex to develop a new model – be patient!

Typical Model Validation Result



- Systematically too high predictions
  - **Higher outcome frequency in development set**
    - **Intercept/baseline hazard too high for new subjects**

Typical Model Validation Result



Slope plot < 1.0

- Low prob too low
- High prob too high
  - **Typical overfitted model in development set**
  - **Too extreme regression coefficients (OR/HR)**

Poor validation = expected

Reilly Ann Int Med 2009; Moons BMJ 2009 + Heart 2012;Steyerberg Plos Med 2013

- Different outcome occurrence (usually lower)
  - Due to improvements in care
- Different patients (case mix)
- Different definition of predictors
- Improvement in measurements: e.g. imaging tests
  - Previous CTs less accurate than spiral CT
- Original model missed important predictor

BMJ Open

Empirical evidence of the impact of study characteristics on the performance of prediction models: a meta-epidemiological study

Johanna A A G Damen,[1,2] Thomas P A Debray,[1,2] Romin Pajouheshnia,[2] Johannes B Reitsma,[1,2] Rob J P M Scholten,[1,2] Karel G M Moons,[1,2] Lotty Hooft[1,2]

ELSEVIER

Journal of Clinical Epidemiology ■ (2014) ■

ORIGINAL ARTICLE

A new framework to enhance the interpretation of external validation studies of clinical prediction models

Thomas P.A. Debray[a,*], Yvonne Vergouwe[b], Hendrik Koffijberg[a], Daan Nieboer[b], Ewout W. Steyerberg[b,1], Karel G.M. Moons[a,1]

External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges

Richard D Riley,[1] Joie Ensor,[1] Kym I E Snell,[2] Thomas P A Debray,[3,4] Doug G Altman,[5] Karel G M Moons,[3,4] Gary S Collins[5]

Multivariate meta-analysis of individual participant data helped externally validate the performance and implementation of a prediction model

Kym I.E. Snell[a], Harry Hua[b], Thomas P.A. Debray[c,d], Joie Ensor[e], Maxime P. Look[f], Karel G.M. Moons[c,d], Richard D. Riley[e,*]

BMJ Open

Empirical evidence of the impact of study characteristics on the performanc of prediction models: a meta-epidemiological study

Johanna A A G Damen,[1,2] Thomas P A Debray,[1,2] Romin Pajouheshnia,[2] Johannes B Reitsma,[1,2] Rob J P M Scholten,[1,2] Karel G M Moons,[1,2] Lotty Hooft[1]

Poor validation = expected

(Reilly Ann Int Med 2009; Moons BMJ 2009 + Heart 2012;Steyerberg Plos Med 2013)

- No matter what reason for poor validation – developing immediately another model means:
  - Neglecting previous models/studies
  - Prediction research becomes completely particularistic
    - Every country, setting, hospital, subgroup 'own' model
  - Validation datasets often smaller → even less generalisable models
  - Perhaps new model needed: but likely not!
- Easy to adjust existing model using validation dataset
  - rather than fitting new model → notably when validation set is small(er)

What evidence do we need to start using prediction models in practice?

Steps in prediction modelling

BMJ series 2009; HEART series 2012; PROGRESS series BMJ + PLOS MED 2013, TRIPOD Ann Intern Med 2015

- 1. Developing prediction model from a particular dataset
- 2. Validate/test the predictive accuracy of previously developed model in (data of) other subjects
- 3. Adjust/tailor model to local situation/care setting using the validation dataset
- 4. Quantify impact of using a model on decision making and patient outcomes

3. Adjusting prediction models

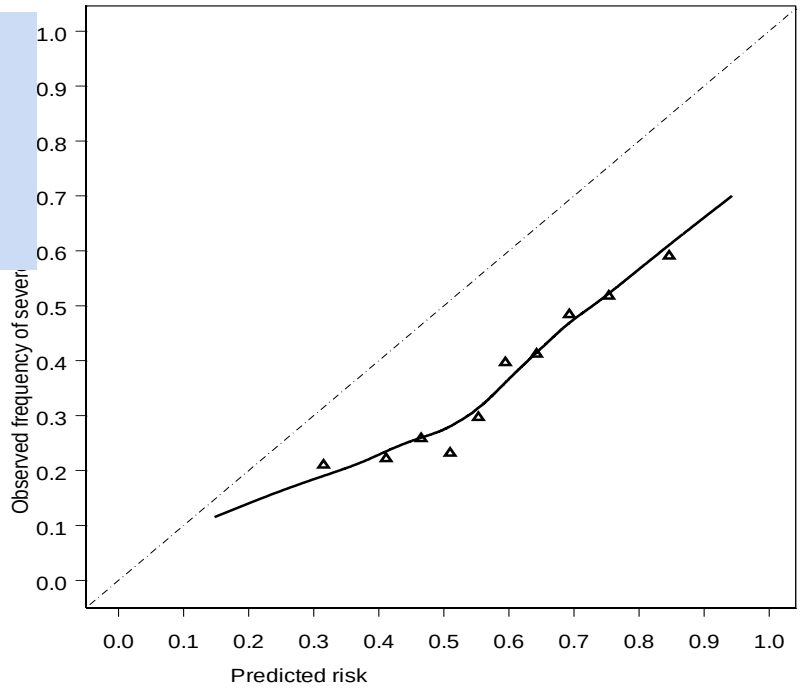
Houwelingen Stat Med 2000; Steyerberg Stat Med 2004; KJM Janssen JCE 2008+CJA 2008; D Toll JCE 2008; Moons Heart 2012)

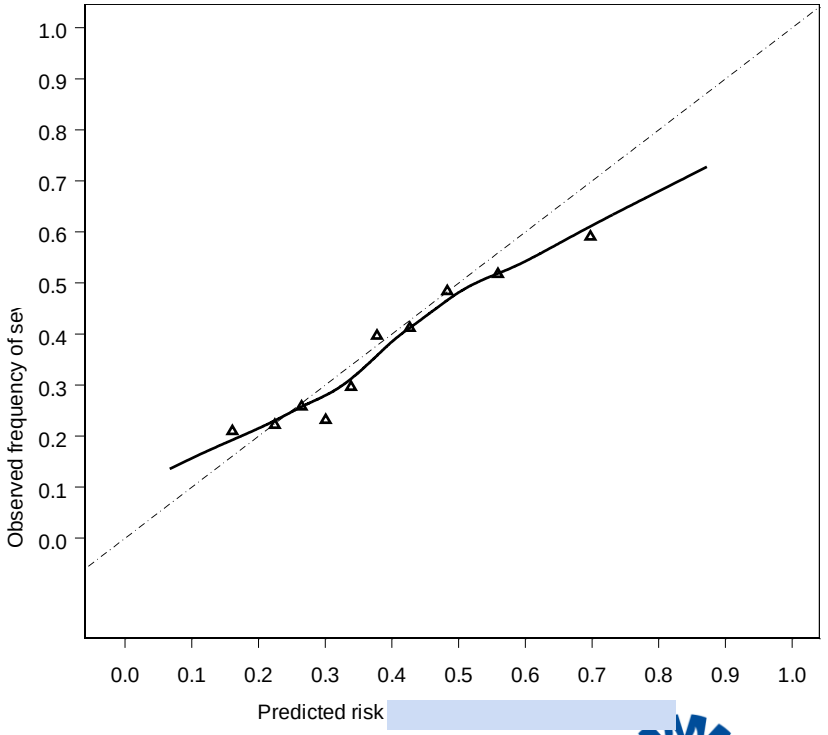
- **Adjusting can be simple and ranges from:**
  - Simple adjustment of base line risk/hazard (intercept)
  - Adjusting regression coefficients of predictors in model
  - Adding previously missed or new predictors/biomarkers
  - Refitting

3. Adjusting prediction models

Houwelingen Stat Med 2000; Steyerberg Stat Med 2004; KJM Janssen JCE 2008+CJA 2008; D Toll JCE 2008; Moons Heart 2012

- Adjusting for difference in overall outcome frequency (intercept adjustment) is often sufficient





- If (also) slope different → adjust predictor weights

3. Adjusting prediction models

Houwelingen Stat Med 2000; Steyerberg Stat Med 2004; KJM Janssen JCE 2008+CJA 2008; D Toll JCE 2008; Moons Heart 2012

- **Updating is particularly important when:**
  - new predictors found → added to existing models
    - CRP to Framingham risk model
  - new era / new setting

- **Updating done after (!) a model's (external) validation → if unsatisfactory accuracy**
    - **Not recommend updating without first validating**

If validation of existing models in our data is unsatisfactory …

…and updating could not fix the job…then

… Develop our new model

What evidence do we need before using prediction models?

4 Steps in prediction modelling

BMJ series 2009; HEART series 2012; PROGRESS series BMJ + PLOS MED 2013, TRIPOD Ann Intern Med 2015

- 1. Developing prediction model from a particular dataset

- 2. Validate/test the predictive accuracy of previously developed model in (data of) other subjects
- 3. Adjust/tailor model to local situation/care setting using the validation dataset
- 4. Quantify impact of using a model on decision making and patient outcomes

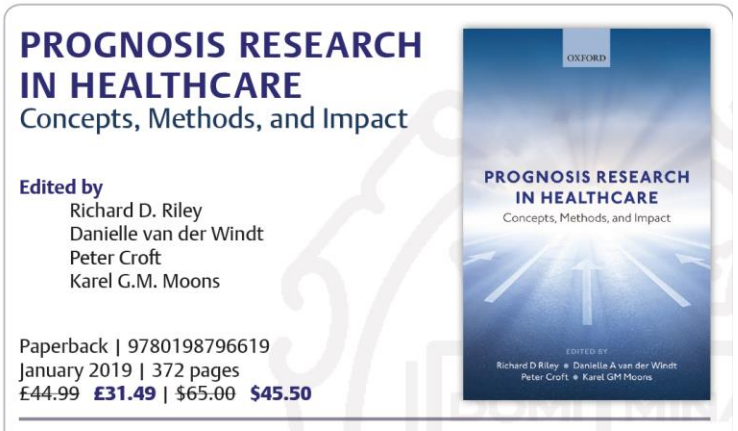
1. Developing a prediction model

No real challenges anymore

So much literature:

Design: Book Grobbee & Hoes 2009; BMJ series 2009; Heart series 2012; PROGRESS series 2013; TRIPOD 2015.

Analysis: Royston BMJ 2009 + Books by Harrell 2001; Steyerberg 2008; Royston & Sauerbrei 2009.



What evidence do we need before using prediction models?

4 Steps in prediction modelling

BMJ series 2009; HEART series 2012; PROGRES series BMJ + PLOS MED 2013, TRIPOD (Ann Intern Med 2015)

- 1. Developing prediction model
- 2. Validate the predictive accuracy of developed model in (data of) other subjects
- 3. Adjust/tailor model to local situation/care setting

- 4. Quantify impact of using a model on decision making and patient outcomes

4. Model impact studies

Campbell BMJ 2000; Reilly+Evans Ann Int M. 2006; Moons BMJ 2009 + Heart 2012

- Recall assumption of prediction models:
  - accurately estimated probabilities…
  - …improve physicians' decision making/behaviour…
  - … and thus patient outcome

- … studied in so-called model impact studies

4. Model impact studies

Campbell BMJ 2000; Reilly+Evans Ann Int M. 2006; Moons BMJ 2009 + Heart 2012

- **Aim: Whether actual use of prediction model truly improves …**
  - … Decision making behaviour (treatment indications) …
  - … Patient outcome or healthcare costs …

  … as compared to not using such model

- **Impact studies are comparative, intervention studies**
  - Intervention = model use + subsequent (treatment) actions based on model predictions
  - In sharp (!) contrast to previous prediction model phases

4. Model impact studies

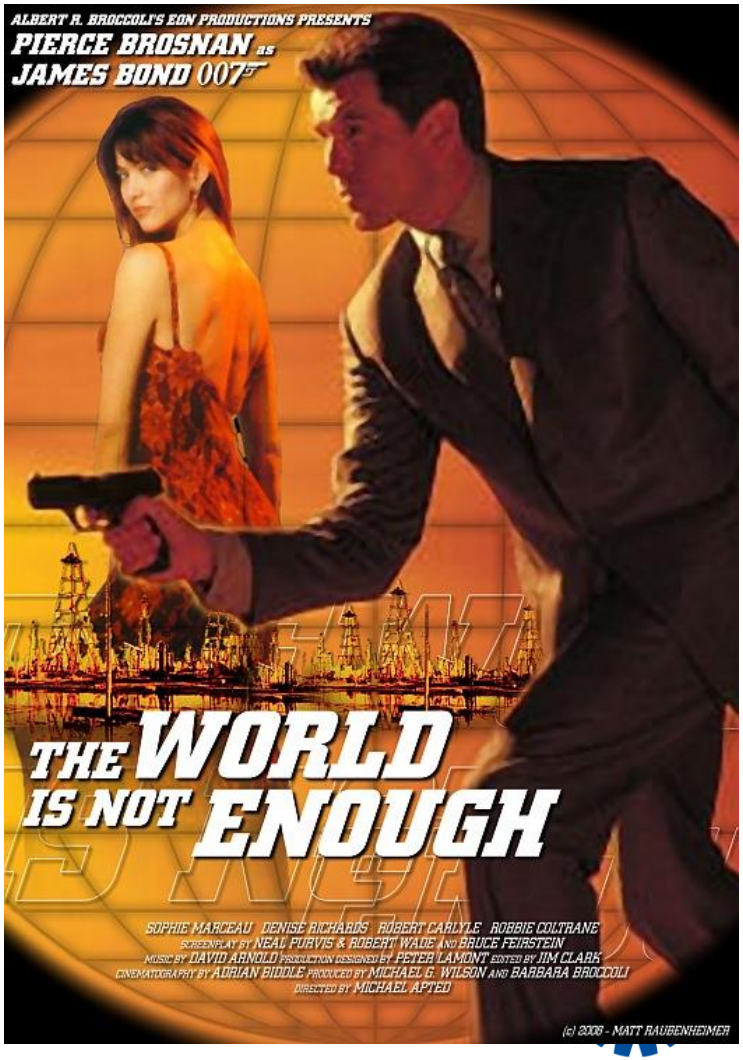
Campbell BMJ 2000; Reilly+Evans Ann Int M. 2006; Moons BMJ 2009 + Heart 2012

- Quantifying effects on patient outcomes:
  - Reflex = randomized comparison
  - This time good reflex: best design indeed RCT
    - Preferably cluster RCT (e.g. stepped wedge) trial
    - Not randomising patients
      - Learning effects of doctors → reduced contrast
    - Randomising practices
      - Less contamination across doctors in same practice → reduced contrast

4. Model impact studies

Campbell BMJ 2000; Reilly+Evans Ann Int M. 2006; Moons BMJ 2009 + Heart 2012

- **Disadvantages Cluster RCTs:**
  - Long duration → Certainly if patient outcomes occur late in time
  - Large studies (costs)
  - Prediction model always studied in combination with current treatments
    - If new treatment → new RCT
- **100.000's clinical prediction models → increase per day**
- **Not enough resources - budget + subjects to study them all in long term, expensive cluster RCT**



4. Model impact studies

Moons BMJ 2009 + Heart 2012; Hendriksen JTH 2013

- Need alternative approaches to separate chaff from wheat
- To determine which models are completely useless and which may …
  - …Change decision making
  - … Change patient outcomes
- Simple approaches to determine whether a model may/may not change decision making + patient outcomes

4. Model impact studies

Moons BMJ 2009 + Heart 2012; Hendriksen JTH 2013

- **1. Cross sectional randomised study**
  - Treatment decision = outcome (no f-up)
  - Outcome never changes if physicians/patients don't change behavior based on model predictions
  - If changes decision making → Still need to quantify whether change in therapeutic decisions actually change patient outcomes

4. Model impact studies

Moons BMJ 2009 + Heart 2012; Hendriksen JTH 2013

- **2. Risk-Benefit analysis**
  - Risk-Benefit (markov) models:
    - Linked evidence approach -- combining Model's predictive accuracy studies + Treatment effect evidence
    - Use predictive probabilities of (validated) model

      +
    - Results of beneftis + risks of existing therapies for that disorder (ideally obtained from RCTs)
    - → To quantify effect of actually using the model with model-directed therapies → on patient outcome (+ cost-effectiveness)

4. Model impact studies

Moons BMJ 2009 + Heart 2012; Hendriksen JTH 2013

- **Indication of expected risks/benefits when introducing model combined with subsequent therapies**
  - + test specific scenarios (e.g. multiple treatment-probability thresholds) or subgroups
  - + whether empirical study is (not) indicated – chaff from wheat
  - + How to enrich RCT design

ELSEVIER

Journal of Clinical Epidemiology 62 (2009) 1248–1252

Journal of Clinical Epidemiology

SYSTEMATIC REVIEW

Decision analysis to complete diagnostic research by closing the gap between test characteristics and cost-effectiveness

Joanna D. Schaafsma[a,*], Yolanda van der Graaf[b], Gabriel J.E. Rinkel[a], Erik Buskens[b,c]

Koffijberg *et al. BMC Medical Research Methodology* 2013, **13**:12
http://www.biomedcentral.com/1471-2288/13/12

BMC Medical Research Methodology

RESEARCH ARTICLE Open Access

From accuracy to patient outcome and cost-effectiveness evaluations of diagnostic tests and biomarkers: an exemplary modelling study

Hendrik Koffijberg[1*], Bas van Zaane[2] and Karel GM Moons[1,2]

4. Model impact studies

Moons BMJ 2009 + Heart 2012; Hendriksen JTH 2013

3. Before-After study

- Compare patient outcomes in period before introducing model to the period after introducing
- E.g. Wells rule for DVT; Ottawa ankle/knee rules

4. Geographical comparison or historical control group

- Disadvantages 3+4: both observational
  - Geographical differences or time changes in therapeutic guidelines/therapies
  - Confounding by indication / case mix differences → adjustment in analysis (like all non-randomized intervention studies)

Take home messages

- **Indeed theoretically 4 consecutive phases of prediction modelling**
  - Development, validation, adjusting (updating), impact assessment
- **But way too much developed models for same outcome or target population**
  - Too much focus on development → 'innovation' / 'own' model
- **If behind your dataset: don't start with phase 1 = developing a model**
  - Do first good systematic review (SR) -- guidance available
  - Then validate these existing models

Take home messages

- **Validation is not refitting original model or repeat analysis of development study in your data**
  - Testing the model ‘as ít is’ in your data
  - Requires proper reporting of original developed models, plus how predictors and outcomes defined/measured
  - not reporting of simplified scores only
  - No random-split sample validation
  - Rather by time, geography, setting/clinical domain
  - Validation is not aiming to find same predictive accuracy as in development set → rather: acceptable accuracy

Take home messages

- **Validation often shows poor accuracy → don't panic → try update first (easy) → suppress your 'development reflex'**

- **If still after updating unsatisfactory performance**
  - Try adjusting original model based on your data

- **If remains unsatisfactory: develop new model + validate**
  - Development No real challenges anymore

Take home messages

- **Impact assessment – not directly jump to RCT**
  - Use alternative approaches to see whether model may lead to improved decision making + patient outcome
- **No developed model applied/guideline without at least 1 external validation → preferably with impact assessment**
- **Validation, Updating, Development, Impact → Report your modelling study well**

Annals of Internal Medicine RESEARCH AND REPORTING METHODS

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

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

Take home messages

Preffered steps in prediction modelling

- 1. Systematic review existing prediction model for your domain or outcome at interest
- 2. Validate/test the predictive accuracy of these retrieved models in (data of) other subjects
- 3. Adjust/tailor model to local situation using the validation dataset
- 4. Developing prediction model from a particular dataset
- 5. Quantify impact of using a model on decision making and patient outcomes
- 6. If pass the above steps – empirical impact study

Next phase: Individual Participant Data (meta-) analyses

Debray TP et al + Riley et al: 2013, 2014, 2015

GUIDELINES AND GUIDANCE

Individual Participant Data (IPD) Meta-analyses of Diagnostic and Prognostic Modeling Studies: Guidance on Their Use

Thomas P. A. Debray[1,2]*, Richard D. Riley[3], Maroeska M. Rovers[4], Johannes B. Reitsma[1,2], Karel G. M. Moons[1,2], Cochrane IPD Meta-analysis Methods group[¶]

Ahmed *et al. BMC Medical Research Methodology* 2014, **14**:3
http://www.biomedcentral.com/1471-2288/14/3

BMC Medical Research Methodology

RESEARCH ARTICLE Open Access

Developing and validating risk prediction models in an individual participant data meta-analysis

Ikhlaaq Ahmed[1], Thomas PA Debray[2], Karel GM Moons[2] and Richard D Riley[3*]

A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis

Thomas P. A. Debray,[a*†] Karel G. M. Moons,[a] Ikhlaaq Ahmed,[b] Hendrik Koffijberg[a] and Richard David Riley[b]

External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges

Richard D Riley,[1] Joie Ensor,[1] Kym I E Snell,[2] Thomas P A Debray,[3,4] Doug G Altman,[5] Karel G M Moons,[3,4] Gary S Collins[5]

ELSEVIER

Journal of Clinical Epidemiology 66 (2013) 865–873

ORIGINAL ARTICLES

Individual participant data meta-analyses should not ignore clustering

Ghada Abo-Zaid[a], Boliang Guo[b], Jonathan J. Deeks[c], Thomas P.A. Debray[d], Ewout W. Steyerberg[e], Karel G.M. Moons[d], Richard David Riley[c,*]

Reporting of artificial intelligence prediction models

Data-driven technologies that form the basis of the digital health-care revolution provide potentially important opportunities to deliver improvements in individual care and to advance innovation in medical research. Digital health technologies include mobile devices and health apps (m-health), e-health

www.thelancet.com Vol 393 April 20, 2019

*Gary S Collins, Karel G M Moons

Thank you for your attention

Basic & Advanced Courses

MSC EPIDEMIOLOGY

Face to Face & Online

Accessible from all over the world
More than 50 courses…

- Basic and Modern Statistical methods
- Advanced Diagnostic Research
- Prognostic research
- Meta Analysis with Individual Participants Data

www.msc-epidemiology.eu

www.msc-epidemiology.online

UMC Utrecht Julius Center

Universiteit Utrecht

elevate

Gentlemen, Choose Your Models

Models are fine and statistics are dandy
But don't choose too quickly just cause they're handy
Stick to a model that's been through the mill
Don't try something new just for the thrill
A new shiny model is full of allure
But making it work is no sinecure.

The more complex the merrier does not follow
The voluminous output may be hard to swallow
Too many variables and too few cases
Is too much like duelling at ten paces

What's fit may be error rather than trend
And shrinkage will get you in the end.

Know what you're doing and do it well
Replictable findings are easy to sell
Be willing to progress one step at a time
A counterfeit dollar's worth less than a dime
Now that I've warned you I'm ready to stop
And let you get back to tending the shop.

PERSONNEL PSYCHOLOGY
1975, 28, 1–18.

UNDERPREDICTION FROM OVERFITTING: 45 YEARS OF SHRINKAGE[1]

ROBERT J. WHERRY, SR.