Prediction Models

From development to validation to clinical impact

Karel G.M. Moons

UMC Utrecht

Julius Center for Health Sciences and Primary Care,

www.juliuscenter.nl

Summary of various papers/consortia

- BMJ: series 4 papers on prognostic modelling (2009)
- Clin Chem: evaluation of biomarkers (2010)
- Heart: 2 papers on prediction modelling (2012)
- Clin Chem: series 4 papers on diagnostic research (2012)
- BMJ and Plos Med: PROGRESS series 4 papers prognostic research (2013)

Prediction

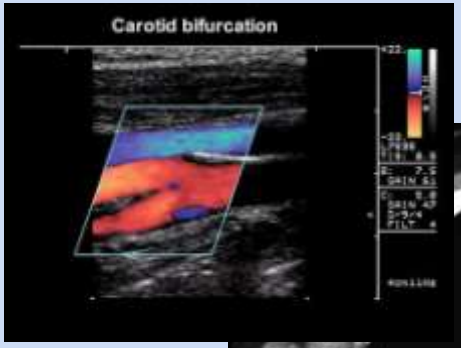
- Prediction = foreseeing / foretelling
  - … (probability) of something that is yet unknown
- Largely 2 situations in medicine:
  - … probability of future conditions/situations = prognosis
  - … probability of result of a more invasive/costly reference (gold) standard test that is not yet done = diagnosis

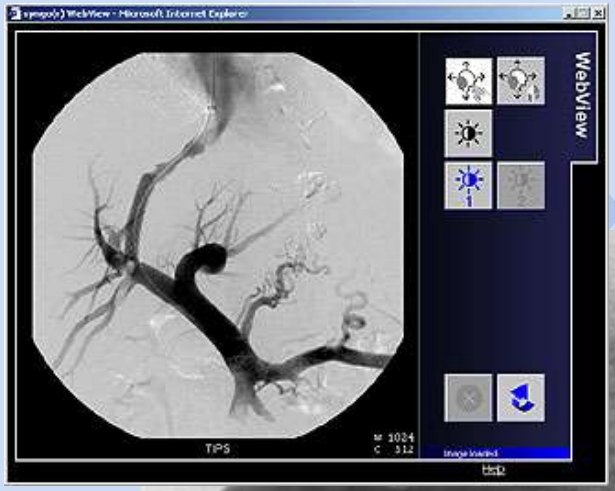
Prediction is done with predictors

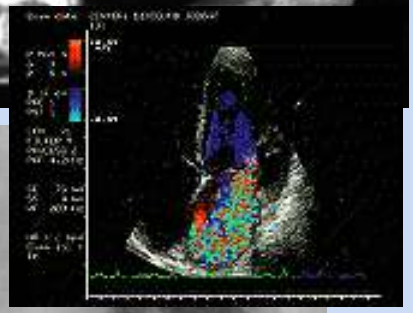
- = variables measured in subject → obtained from:
  - Patient history
  - Physical examination
  - Imaging tests
  - Elektrofysiology (ECG, EEG)
  - Blood/urine markers
  - Genetic markers
  - Disease characteristics

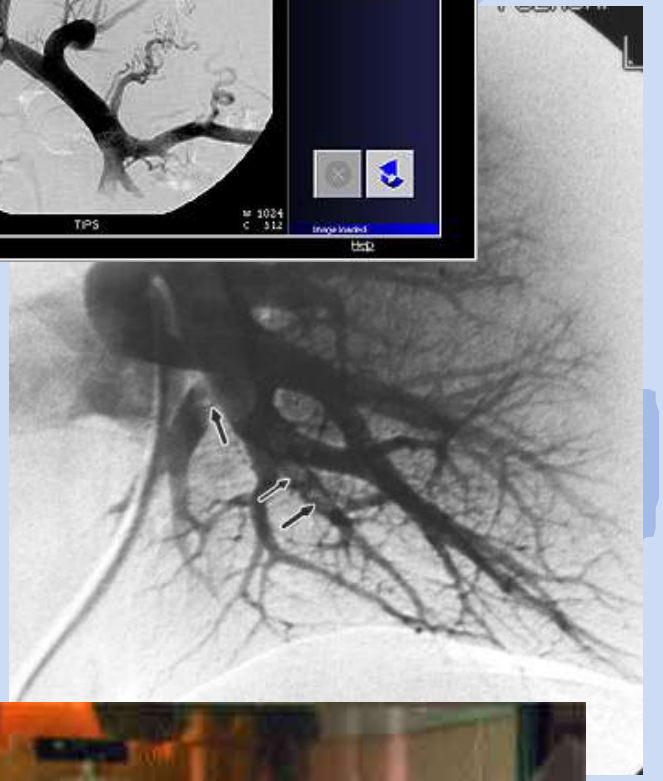
Finding new predictors/biomarkers/tests= HOT
Also in this field

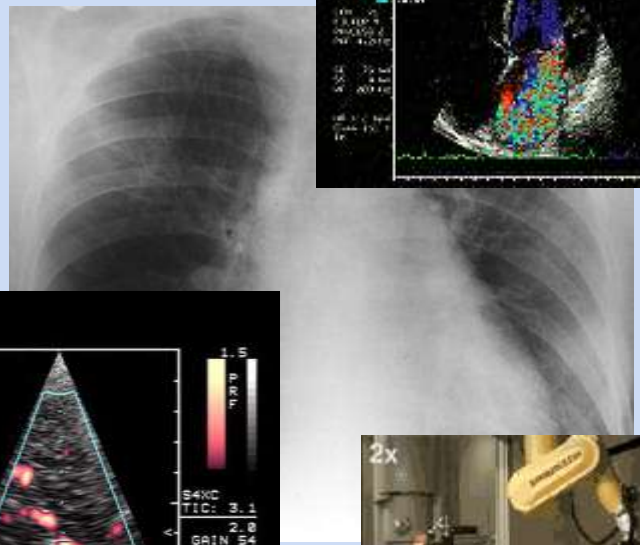
- # increases per day → greatly vary in
  - Predictive accuraccy
  - Invasiveness / burden
  - Measurement costs

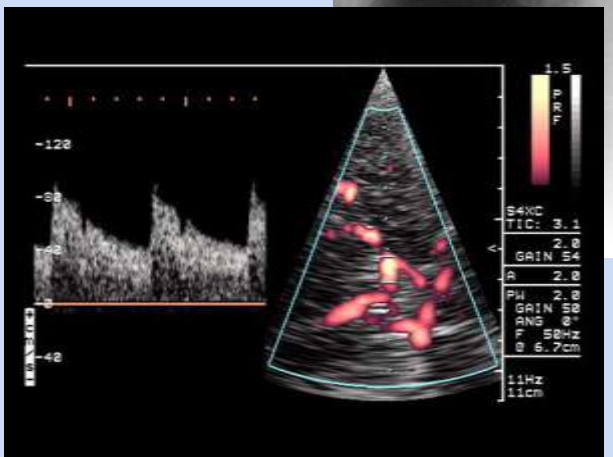

Carotid bifurcation


WebView
TIPS








-120
-80
-40
S4XC
TIC: 3.1
2.0
GAIN 54
A 2.0
PW 2.0
GAIN 50
ANG 0°
F 58Hz
@ 6.7cm
11Hz
11cm


2x

AXIS-SHIELD

About us Contacts Axis-Shield plc

Products \ NycoCard \ Tests

Home
Products
Disease Areas
News & Events
Distributors

Login for distributors
login

NycoCard CRP test

The NycoCard CRP test is a 2-minute Point of Care test to indicate bacterial or viral cause of infection. NycoCard CRP measures C-reactive protein (CRP), an acute phase protein that increases rapidly after onset of infection.

Test specific information

- Sample volume: 5 µL
- Assay time: 2 minutes
- Sample material: Whole blood, serum or plasma
- Measuring range: 8 - 250 mg/L for whole blood samples and 5 - 150 mg/L for serum and plasma samples
- Stability at room temperature: 4 weeks
- Kit size: 24 and 48 tests
- NycoCard CRP Control: Positive control provided with the kit

Clinical use of NycoCard CRP

- Reduces unnecessary use of antibiotics
- More rapid induction of treatment
- Fewer hospital admissions
- Healthcare cost savings

Helena Catalog 1-800-231-5663

ColoCARE®

ColoCARE is the leading throw-in-the-bowl test for detecting pre-symptomatic occult bleeding caused by gastrointestinal diseases. It is **safer, easier** and more pleasant to use than traditional guaiac slide tests. Simply place a ColoCARE test pad in the toilet after a bowel movement, watch for a color change, then flush the pad away. It's **clean** and **disposable,** easy for elderly patients to see and interpret, and **extremely sensitive,** with no increase in the false positive rate. It is more **cost-effective** than guaiac slide tests because it requires no stool handling, no chemical developers, no laboratory processing, and no mailing of biohazards. Elimination of stool handling overcomes the number one patient objection to occult blood testing, resulting in wider use of the test and leading to greater success in early detection of pathological conditions.
The test pad consists of biodegradable paper chemically treated with a chromogen. The pad is floated on the water surface in the toilet bowl. If detectable blood is present, the hemoglobin reacts with the chromogen, and a blue and/or green color reaction occurs. The test pad has three reaction sites: a large test square and two smaller control squares to verify the system functions properly.

Pubmed 'Biomarkers': 621854 hits

Proteomics
Genomics
Metabolomics

Practice

- **Hardly any diagnosis/prognosis based on single variable (test/marker result)**
  - doctors measure many variables → combine them → estimate diagnostic + prognostic probabilities
- **Markers/tests only part (sometimes small) of diagnostic, prognostic and treatment-effect predictions**
- **Desired knowledge/evidence for professionals:**
  - Does next test/marker has added value to what they already know from the patient (easy variables)?
  - Or simply: Does it provide added predictive value?

New markers/tests
(actually all non-pharmaceuticals)

- **Problem: Simply enter market**
  - Drugs rigorous phased approach
  - Not diagnostic/prognostic tests: Very liberal guidelines
    - Only safety (KEMA/DEKRA → CE approval)
    - Not: Diagnostic or prognostic accuracy → let alone added value
    - = OUR JOB!!!
- **Consequences of fast market access…**

New markers/tests

- 1. High availability
  - Only increase ('omics' area) and 'point of care' markers/tests
- 2. Overtesting
  - Reasons: patient satisfaction; fear legal consequences; belief that new 'toys' always better
  - Overtesting → unnecessary burden to doctors, patients, budgets
    - Health care resources not used for those who need most
  - Incorrect use: Swan-Ganz; ICP monitoring; preoperative ECG → Only increase in 'omics' area and point of care tests

Circulation
JOURNAL OF THE AMERICAN HEART ASSOCIATION
American Heart Association®
Learn and Live℠

Hlatky et al, 2009

Criteria for Evaluation of Novel Markers of Cardiovascular Risk

Focus on prognostic cardiovascular markers

Phased approach

- From single testing → do marker levels differ between subjects with vs. without outcome?...
- ... to... Quantify added value to existing predictors using so-called multivariable (clinical) prediction models
- ...to... Quantify impact/clinical usefulness of such prediction models on decision making and thus patient outcomes

Central issue in current marker research

- Key words:
  - Added value → using multivariable analysis and prediction models
  - Clinical usefulness

- NOT: developing/searching new biomarker kits → many out there for same patients or outcomes
  - Review (Riley et al): 131 biomarkers for prognosis of neuroblastoma (in just few years) → can't be all relevant
  - Challenge for new markers is to beat existing strong predictors
  - OUR JOB to quantify that!

Quantifing independent/added value of markers requires multivariable (clinical prediction) modeling approach

Multivariable clinical prediction models

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 irrita-bility	No response	Grimace	Cry or cough
Color	Blue or pale	Body pink, ex-tremities blue	Completely pink

Σ = Apgar score (0-10)

Five-Minute Apgar Score	No. of Live Births	Neonatal Death	Relative Risk (95% CI)
		no. (rate per 1000 births)	
0–3	86	21 (244)	1460 (835–2555)
4–6	561	5 (9)	53 (20–140)
7–10	131,581	22 (0.2)	1

*Infants with five-minute Apgar scores of 7 to 10 served as the reference group. CI denotes confidence interval.

10 year risk of CVD

$$S_0(age) = \exp\{-(\exp(\alpha))(age-20)^p\}$$
$$S_0(age+10) = \exp\{-(\exp(\alpha))(age-10)^p\}^* \quad (1)$$

$$w = \beta_{chol}(cholesterol-6) + \beta_{SBP}(SBP-120) + \beta_{smoker}(current) \quad (2)$$

$$S(age) = \{S_0(age)\}^{\exp(w)}$$
$$S(age+10) = \{S_0(age+10)\}^{\exp(w)} \quad (3)$$

$$S_{10}(age) = S(age+10)/S(age) \quad (4)$$

$$Risk_{10} = 1 - S_{10}(age) \quad (5)$$

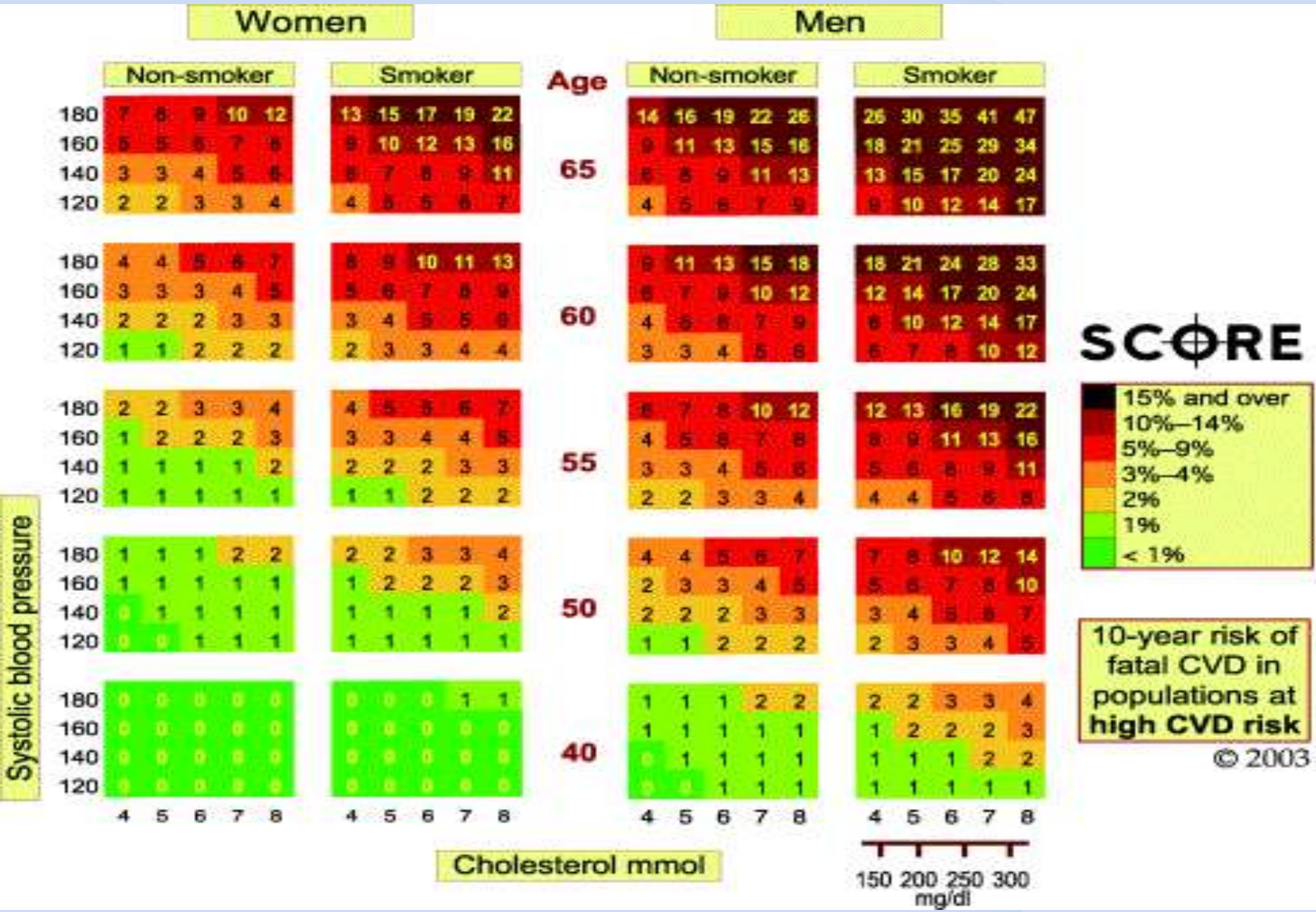
$$CVDRisk_{10}(age) = [CHDRisk(age)] + [Non\text{-}CHDRisk(age)] \quad (6)$$

Table A Coefficients for Eq. (1)

		CHD		Non-CHD CVD	
		α	p	α	p
Low risk	Men	−22.1	4.71	−26.7	5.64
	Women	−29.8	6.36	−31.0	6.62
High risk	Men	−21.0	4.62	−25.7	5.47
	Women	−28.7	6.23	−30.0	6.42

Table B Coefficients for Eq. (2)

	CHD	Non-CHD CVD
Current smoker	0.71	0.63
Cholesterol (mmol/L)	0.24	0.02
Systolic BP (mmHg)	0.018	0.022


Women
Men
Non-smoker
Smoker
Non-smoker
Smoker
Age
65
60
55
50
40
Systolic blood pressure
180
160
140
120
4 5 6 7 8
Cholesterol mmol
150 200 250 300
mg/dl
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
© 2003

Prediction (model) is not obscure = not restricted to medicine

Clinical prediction models

- Convert predictor values of subject to an absolute probability…
  - …of having (!) a particular disease → diagnosis
  - … of developing (!) particular health state → prognosis
    - … within a certain time (hours, days, weeks, years)
    - Dying, complication, disease progression, hospitalised, quality of life, pain, therapy response

Clinical prediction models

- Predictors (for both aims) are:
  - history taking
  - physical examination
  - tests (imaging, ECG, blood markers, genetic 'markers')
  - disease severity
  - received therapies

Prognostic prediction models

- **Sometimes distinction**
  - Prognostic markers/models = baseline prognosis
  - Predictive markers/models = therapy respons
  - Same requirements for design, analysis, reporting
- … Plus: does not matter whether predictor is answer to simple question; blood/urine marker; imaging; ECG; genomics; metabolomics, etc.

Clinical prediction model

- Presented as:
  - Mathematical formula requiring computer – certainly dynamic predicton models
  - Simple scoring rules (Apgar)
  - Score charts / Nomograms (SCORE / Framingham)

Why using prediction models?

- It is very difficult to make an accurate prediction, especially about the future (Niels Bohr (1885-1962))

- Diseases have multiple causes, presentations and courses (McShane LM 2005; Riley RD 2003. Moons, BMJ 2009)

  - A patient's diagnosis and prognosis rarely based on single predictor

  - 'Impossible' for human brain to disentangle and weigh all contributing factors, and to adjust for their mutual influence

  - Our weather (wo)man can also not do this!

Conservation of momentum:

$$\frac{du}{dt} - \left(f + u\frac{\tan\phi}{a}\right)v = -\frac{1}{a\cos\phi}\frac{1}{\rho}\frac{\partial p}{\partial\lambda} + F_\lambda$$

$$\frac{dv}{dt} + \left(f + u\frac{\tan\phi}{a}\right)u = -\frac{1}{\rho a}\frac{\partial p}{\partial\phi} + F_\phi$$

Hydrostatic approximation:

$$g = -\frac{1}{\rho}\frac{\partial p}{\partial z}$$

Conservation of mass:

$$\frac{\partial\rho}{\partial t} = -\frac{1}{a\cos\phi}\left(\frac{\partial}{\partial\lambda}(\rho u) + \frac{\partial}{\partial\phi}(\rho v\cos\phi)\right) - \frac{\partial}{\partial z}(\rho w)$$

Conservation of energy:

$$C_p\frac{dT}{dt} - \frac{1}{\rho}\frac{dp}{dt} = Q$$

State equation (atmosphere):

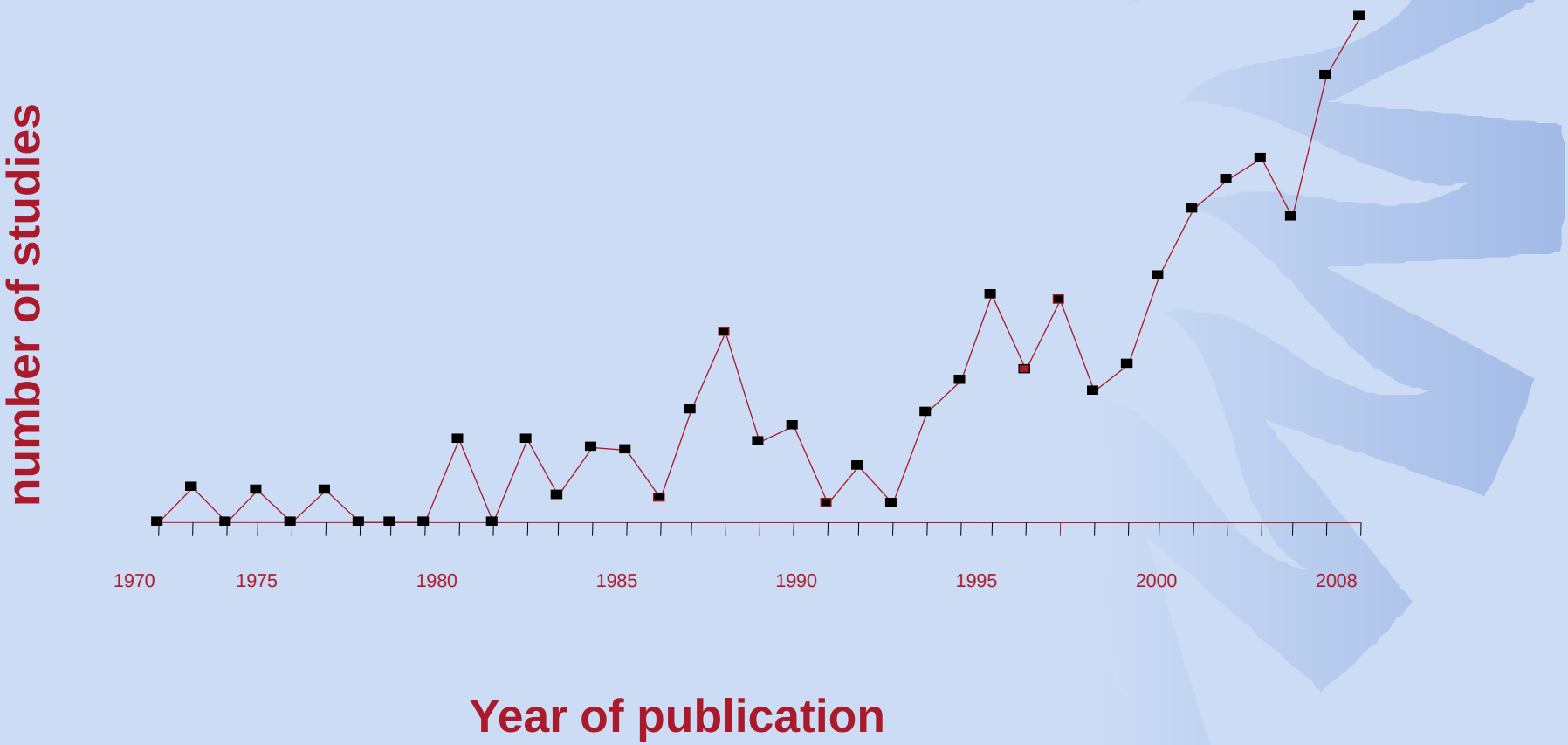
$$p = \rho RT$$

Why using prediction models?

- **… Not meant to replace physician by a computer, but to complement their clinical intuition**

- **Assumption:**
  - Accurately/objectively estimated probabilities…
  - …improve physicians' behaviour / decision making …
  - … and thus patient outcome

Prediction models are hot
(Steyerberg 2009)



10.000's examples

- Apgar score
- Framingham risk score
- SCORE
- Euroscore (cardiac surgery)
- Goldman risk index (chest pain)
- Over 60 models for cancer prognosis (e.g. Gail model)
- Over 100 models for TBI patients
- APACHE score , SAPS score (IC models)
- Ottawa ankle and knee rules
- Reynolds risk score

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.

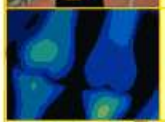
To get started, choose one of the diseases below.

What is your risk?

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

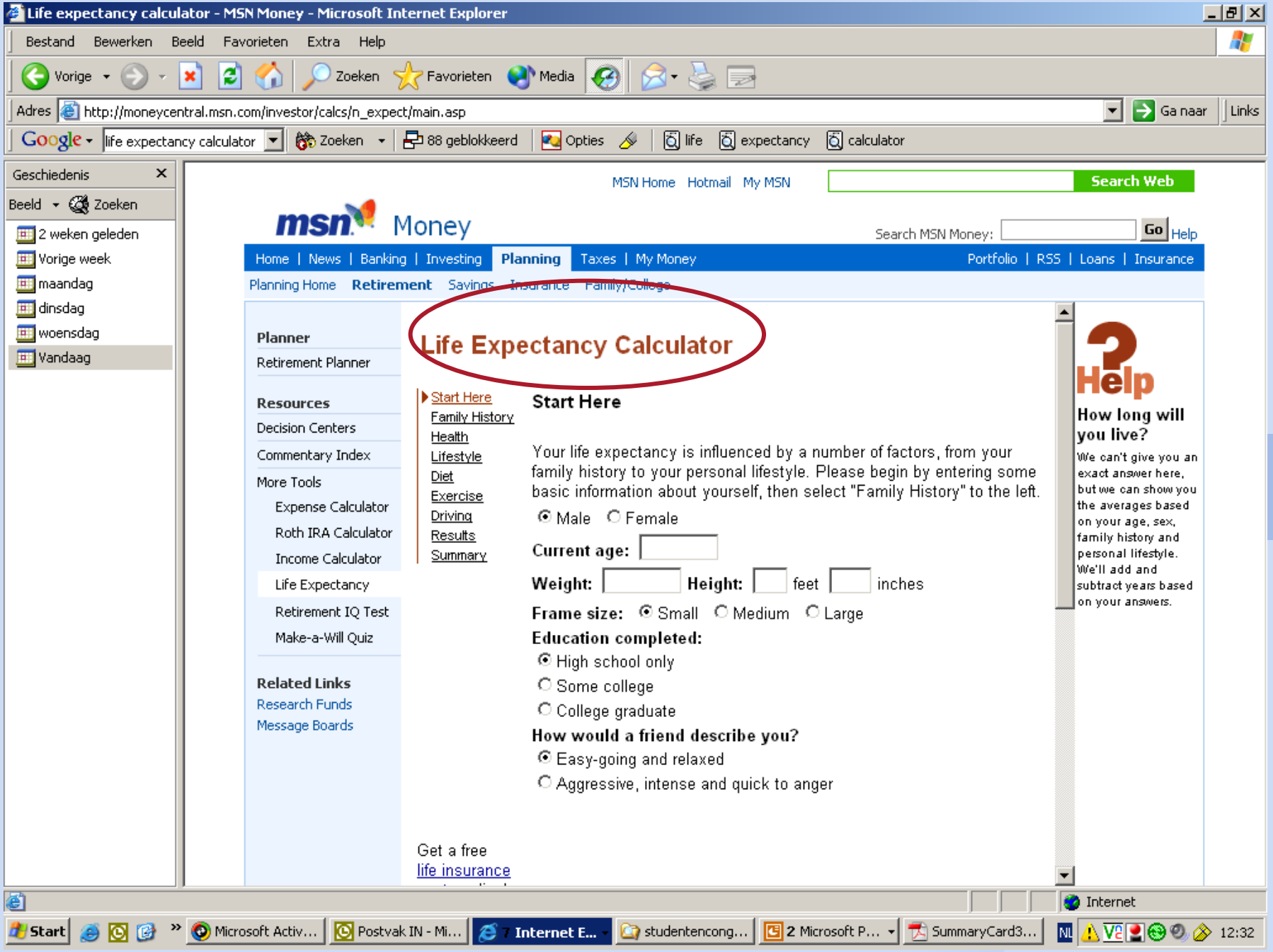
Diabetes: Over 18 million in the U.S. suffer from it. Take steps now to lower your risk. — What's your diabetes risk?

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?


Life expectancy calculator - MSN Money - Microsoft Internet Explorer
Bestand Bewerken Beeld Favorieten Extra Help
Vorige Zoeken Favorieten Media
Adres http://moneycentral.msn.com/investor/calcs/n_expect/main.asp
Ga naar Links
Google life expectancy calculator Zoeken 88 geblokkeerd Opties life expectancy calculator
Geschiedenis
Beeld Zoeken
2 weken geleden
Vorige week
maandag
dinsdag
woensdag
Vandaag
MSN Home Hotmail My MSN
Search Web
msn Money
Search MSN Money: Go Help
Home | News | Banking | Investing | Planning | Taxes | My Money
Portfolio | RSS | Loans | Insurance
Planning Home Retirement Savings Insurance Family/College
Planner
Retirement Planner
Resources
Decision Centers
Commentary Index
More Tools
Expense Calculator
Roth IRA Calculator
Income Calculator
Life Expectancy
Retirement IQ Test
Make-a-Will Quiz
Related Links
Research Funds
Message Boards
Life Expectancy Calculator
Start Here
Family History
Health
Lifestyle
Diet
Exercise
Driving
Results
Summary
Start Here
Your life expectancy is influenced by a number of factors, from your family history to your personal lifestyle. Please begin by entering some basic information about yourself, then select "Family History" to the left.
Male Female
Current age:
Weight: Height: feet inches
Frame size: Small Medium Large
Education completed:
High school only
Some college
College graduate
How would a friend describe you?
Easy-going and relaxed
Aggressive, intense and quick to anger
Get a free
life insurance
Help
How long will you live?
We can't give you an exact answer here, but we can show you the averages based on your age, sex, family history and personal lifestyle. We'll add and subtract years based on your answers.
Internet
Start
Microsoft Activ...
Postvak IN - Mi...
Internet E...
studentencong...
2 Microsoft P...
SummaryCard3...
12:32

Bank of Scotland

Bankruptcy Prediction Models

No one has ever
claimed that the results were not valid.

To try this model
yourself go to Business Bankruptcy Predictor.

What evidence needed to apply prediction models in practice?

Steps in prediction modeling

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

- 1. Developing the prediction model
- 2. Validate the model in other subjects
- 3. Update existing models to local situations
- 4. Quantify impact of using a model on doctor's decision making and patient outcome (cost-effectiveness)

1. Development studies

- Many reviews (G Collins 2010/2011; S Mallet 2010;W Bouwmeester 2012) show that majority of prediction models still poorly developed → in all disciplines

- In fact: no real challenges anymore → Much literature:

  - Design (Grobbee&Hoes 2009; BMJ series 2009; Heart series 2012; Plos Med series 2013)

  - Analysis including quantifying added value of new test (Royston BMJ 2009;Books by Harrell 2001; Steyerberg 2008; Royston&Sauerbrei 2009; others)

1. Development study characteristics

FEW HIGHLIGHTS

DESIGN

(Moons BMJ 2009 + Heart 2012)

1. Inherently multivariable

- In practice: diagnosis and prognosis rarely done by single test/marker/etc.
- Diagnostic and prognostic research should provide evidence on …
  - 1. Which are true diagnostic and prognostic predictors
  - 2. Whether new predictor truly adds predictive information to easy to obtain predictors
  - 3. Outcome probabilities for (different) predictor combinations or tools to estimate these probabilities
    - All require multivariable approach in design + analysis

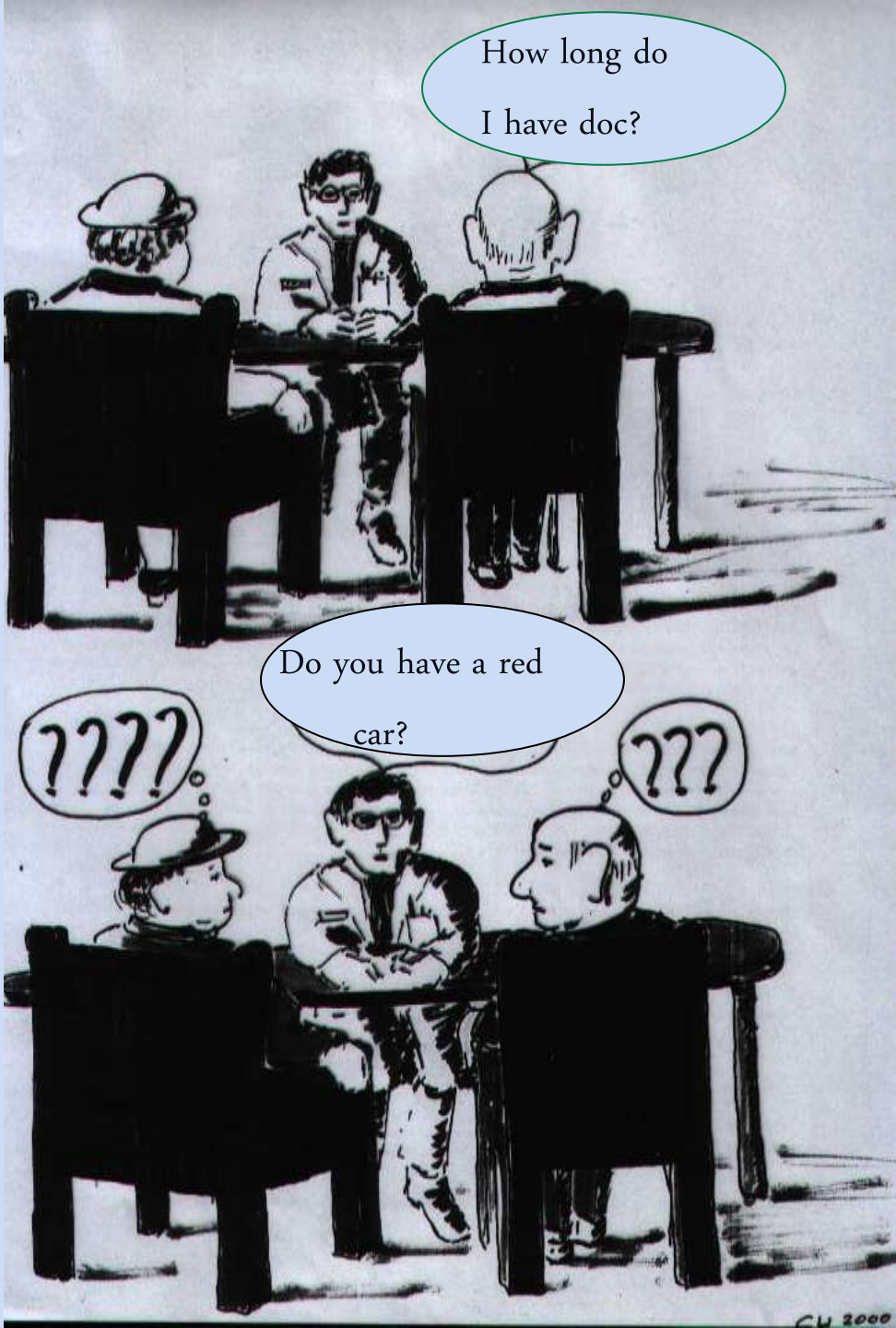
2. Prediction research != aetiologic research

- despite clear similarities in design+analysis (Brotman, 2005)

Different aims

- Aetiologic: explain whether outcome occurrence can be attributed to particular risk factor → pathofysiology
  - Adjusted for other risk factors
- Prediction: (simply) to predict as accurate as possible
  - Predictive analysis gives insight in causality but is aim nor requirement

- Aetiology: predictors theoretically in causal pathway
- Prediction: all variables potentially related to outcome can be studied
  - E.g. imaging test results, biomarkers
- Every causal factor is predictor but not v.v.



2. Prognostic research != aetiologic research

Different analysis + presentation

- Both (same) multivariable models … but different results reported from the output
- Prediction studies: absolute probabilities of disease presence/occurrence
  - Etiologic studies: Relative risk estimates / odds ratios
- Prediction studies: calibration, discrimination, (re)classification
  - Non-issue in etiology

3. Subject Selection

- **Ideal: cohort study (may be obtained from RCT) on subjects with same characteristic, i.e. …**
  - **…Suspected of a disease (diagnosis)**
  - **…Having a disease, lying at IC, being pregnant, being born (prognosis)**
  - **Prospective cohort (preferred)**
    - Retrospective dominate literature unfortunately (McShane 2005; Riley 2003)

3. Subject selection
Case control / Case cohort

- **Ideal for causal not for prediction studies**

- **No absolute probabilities**

- **Exception: nested-case-control or case-cohort study (*Biesheuvel et al, BMC 2008; Rutjes et al Clin Chem 2005; Ganna Am J Epi 2012)***

  - **Sample fraction known (weight controls with inverse sample fraction) → Ideal if:**

    - Predictor expensive (genetic marker, reading images)
    - Retrospective analysis stored data / human material
      - Biomarkers!!!

3. Subject selection randomised trial data

- When Ry is ineffective: combine both groups
- If Ry effective
  - only control group (limited power)
  - combine → include treatment(s) as seperate predictor
    - Ry studied on (independent) predictive effect
- Generalisability/external validation issue

4. Candidate predictors

- Prediction research = to serve practice

- Predictors well defined, standardized, reproducible to enhance generalisability + applicability

  - Care with predictors requiring interpretation

    - Imaging test results → study observers rather than test results

5. Outcome

- Preferably patient-relevant outcome
  - Event, remission disease, death, pain, growth
  - Intermediates (IC stay, physiology aspects) unhelpful
    - Except clear association with patient outcome; E.g. CD4 count in HIV ; athersclerosis % for CAD.
- Define time (F-up) period in which outcomes measured
- Measure outcomes without knowledge of predictors (except death)

6. Required number of subjects

- Multivariable prediction research → no rules for power calculations
  - Too many candidate predictors compared to # events → risk of optimistic predictive performance + improper variable selection
- Ideally hundreds of events
- Suggested (at least) 10 events per predictor
  - *Peduzzi J Clin Epi 1995, Concato J Clin Epi 1995; Harrell 2001.*

1. Development study characteristics

(Steyerberg Book 2009, Royston BMJ 2009, Moons BMJ 2009)

ANALYSIS

Typical Multivariable Prediction Study

- Define all potential predictors one could think of
- Select cohort members
- Measure in each patient all potential predictors plus the outcome
- Univariable analysis: select significant ones ($p<0.05$, perhaps 0.10)
- 'Throw' these in multivariable model (Logistic or Cox)
- Remove non-significant ones ($p > 0.05$) = final model
- Interpret estimated regression coefficients (OR's) selected predictors
- Estimate ROC area (and – if lucky – calibration) of model
- Use regression coefficients to make easy sum score (nomogram)
- Presented as the prediction model for the studied outcome

Typical Multivariable Prediction Study

- **Selected predictors: too extreme regression coefficients**
  - Spurious predictors (by chance large OR in data)
  - Multiple testing
- **Missing important predictors**
  - By chance low OR in data
- **Predictive accuracy of model in data too optimistic**
  - Worse predictions (accuracy) in other/new patients
- **Reason**
  - Too many predictors, too little data
  - Same data used to select predictors and to estimate regression coefficients (data 'overused'/overfitted)

‘Two’ types prediction studies

- **Existing knowledge at time of study initiation should determine aim, design, analysis and presentation of model development study**

- **2 types of ‘prediction model development studies’:**

1. **If prior studies on most promising predictors already exist:**

   Fit and present the estimated predictor weights, and a final prediction model for future patients

Two types prediction studies

2. If yet limited knowledge on most likely predictors (and certainly with limited events in your data)

- Aim is not to fit and present a final prediction model
- Rather explorative (hypothesis generation) study → 'potential predictor finding'

  - To set the stage for future prediction studies

Model development
(much prior knowledge; type 1)

- *Harrell, Stat Med 1996 + 2001 (book); Royston + Sauerbrei 2008 (book); Steyerberg 2009 (book)*

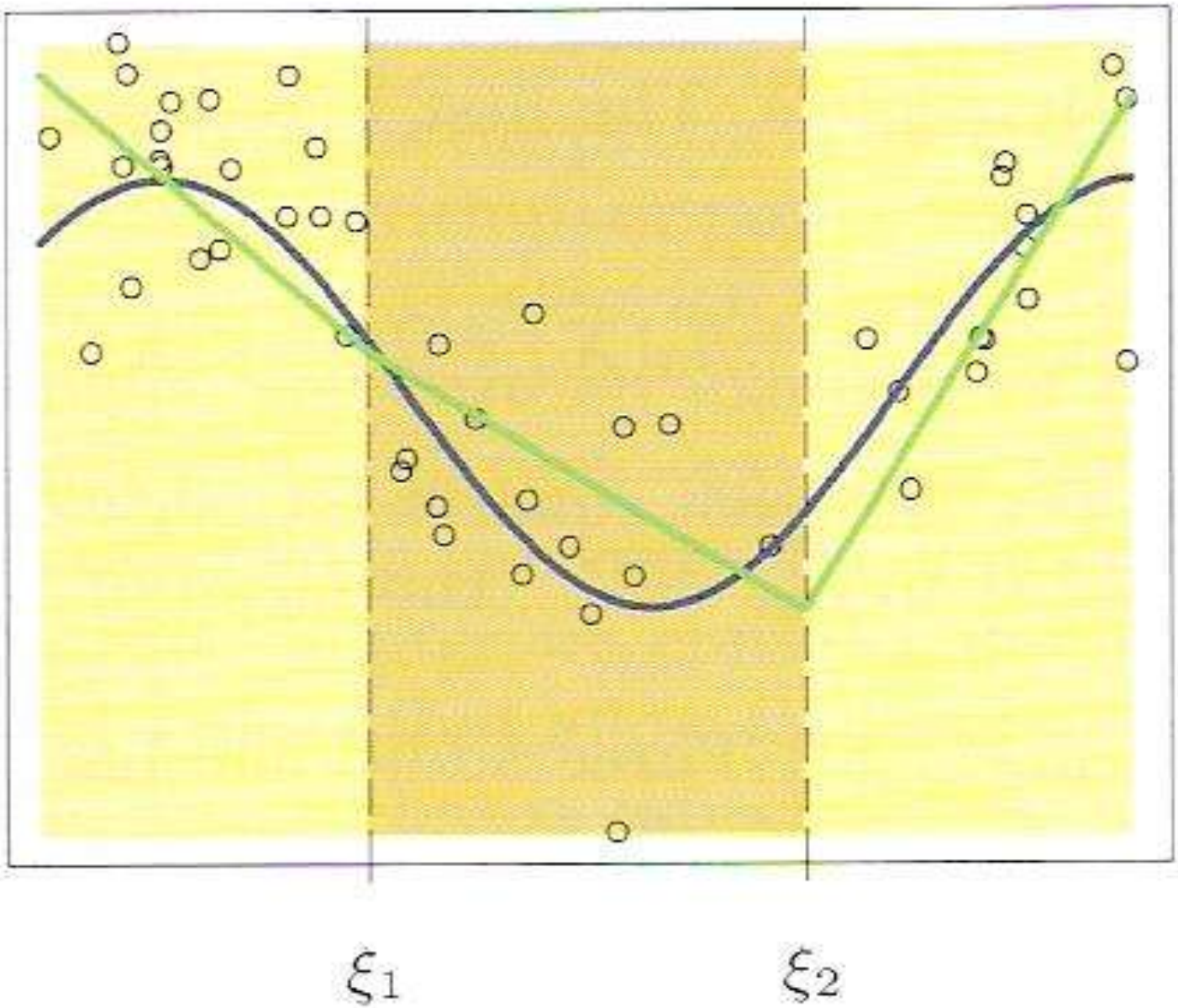
6 Steps (largely):

- ***1. Preselect candidate predictors***
  - Depend on existing knowledge
  - Not use univariable preselection
    - *Based on predictor-outcome*
    - *rather correlations between predictors*

Model development

- 2. Evaluate data quality
  - Missing values
  - Combine predictors (certainly if limited data)
  - Keep continuous predictors as such
    - Dichotomising leads to loss of information
    - In practice: patient has a certain value: not just high or low
  - Check relation with outcome (Altman+Royston, BMJ 2006)
    - Splines or fractional polynomials

Continuous Piecewise Linear

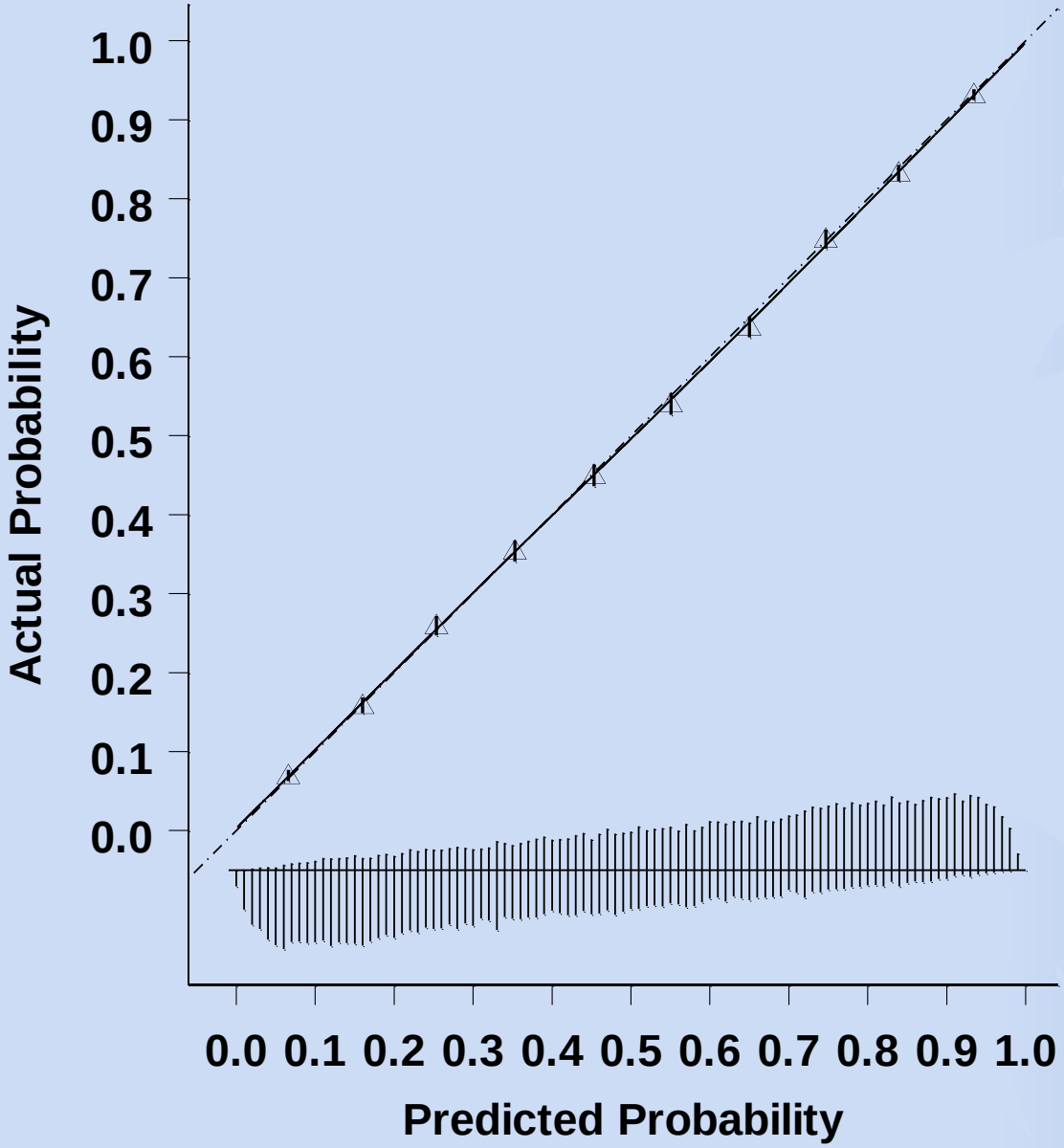


Model development

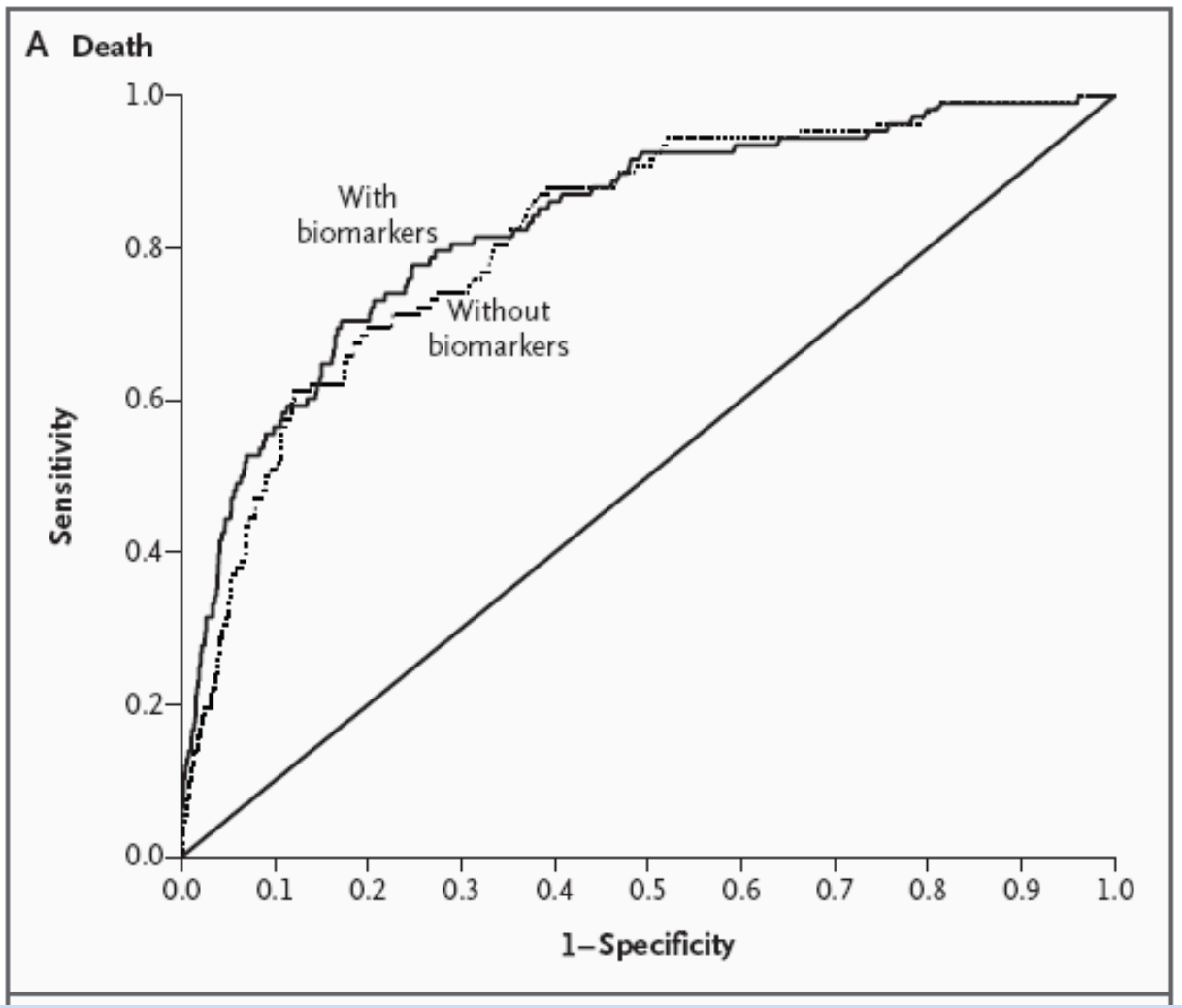
- **3. Choose predictor selection strategy**
  - No consensus and much debate
  - Two main approaches
    - ***Full model (Harrell)***
      - *Avoids: overfitting, selection wrong predictors, correct SE's*
      - *Disadvantages: not easy to define (much prior knowledge)*
    - ***Backward elimination***
      - *Higher p-value*
      - *Bootstrap and shrink (if needed)*

Model development

- **4. Estimate model performance**
  - Calibration (for specific time point in case of survival models)
    - Plot (not H-L test → seldom significant)
  - Discrimination
    - C-statistic (ROC area for logistic regression)
  - (Re)classification
    - NRI → in case of model comparison / addition of new predictor (Pencina Stat Med 2008) → requires thresholds
    - Often arbitrary
    - IDI / Decision curve analysis


Actual Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Predicted Probability

Biomarkers (CRP, etc) as predictors of cardiovascular morbidity/mortality



AUC 0.76

AUC 0.77

Wang TJ, et al. NEJM

Table II. Reclassification among people who experience a CHD event and those who do not experience a CHD event on follow-up.

Model without HDL	Model with HDL			
Frequency (Row per cent)	<6 per cent	6–20 per cent	>20 per cent	Total
Participants who experience a CHD Event				
<6 per cent	39 (72.22)	15 (27.78)	0 (0.00)	54
6–20 per cent	4 (3.81)	87 (82.86)	14 (13.33)	105
>20 per cent	0 (0.00)	3 (12.50)	21 (87.50)	24
Total	43	105	35	183
Participants who do not experience a CHD Event				
<6 per cent	1959 (93.24)	142 (6.76)	0 (0.00)	2101
6–20 per cent	148 (16.78)	703 (79.71)	31 (3.51)	882
>20 per cent	1 (1.02)	25 (25.51)	72 (73.47)	98
Total	2108	870	103	3081

$$\frac{\#events\_moving\_up}{\#events} - \frac{\#events\_moving\_down}{\#events}$$

$$\frac{\#nonevents\_moving\_up}{\#nonevents} - \frac{\#nonevents\_moving\_down}{\#nonevents}$$

Table II. Reclassification among people who experience a CHD event and those who do not experience a CHD event on follow-up.

Model without HDL	Model with HDL			
Frequency (Row per cent)	<6 per cent	6–20 per cent	>20 per cent	Total
Participants who experience a CHD Event				
<6 per cent	39 (72.22)	15 (27.78)	0 (0.00)	54
6–20 per cent	4 (3.81)	87 (82.86)	14 (13.33)	105
>20 per cent	0 (0.00)	3 (12.50)	21 (87.50)	24
Total	43	105	35	183
Participants who do not experience a CHD Event				
<6 per cent	1959 (93.24)	142 (6.76)	0 (0.00)	2101
6–20 per cent	148 (16.78)	703 (79.71)	31 (3.51)	882
>20 per cent	1 (1.02)	25 (25.51)	72 (73.47)	98
Total	2108	870	103	3081

- Net gain in reclassification events = (29-7)/183=0.12
- Net gain in reclassification nonevents = (173-174)/3081= -0.001
- NRI = 0.12 - -0.001= 0.121 ($p < 0.001$)
- Addition of HDL improved the classification of events with 12%

Model development

- ***5. Check overfitting / optimism***

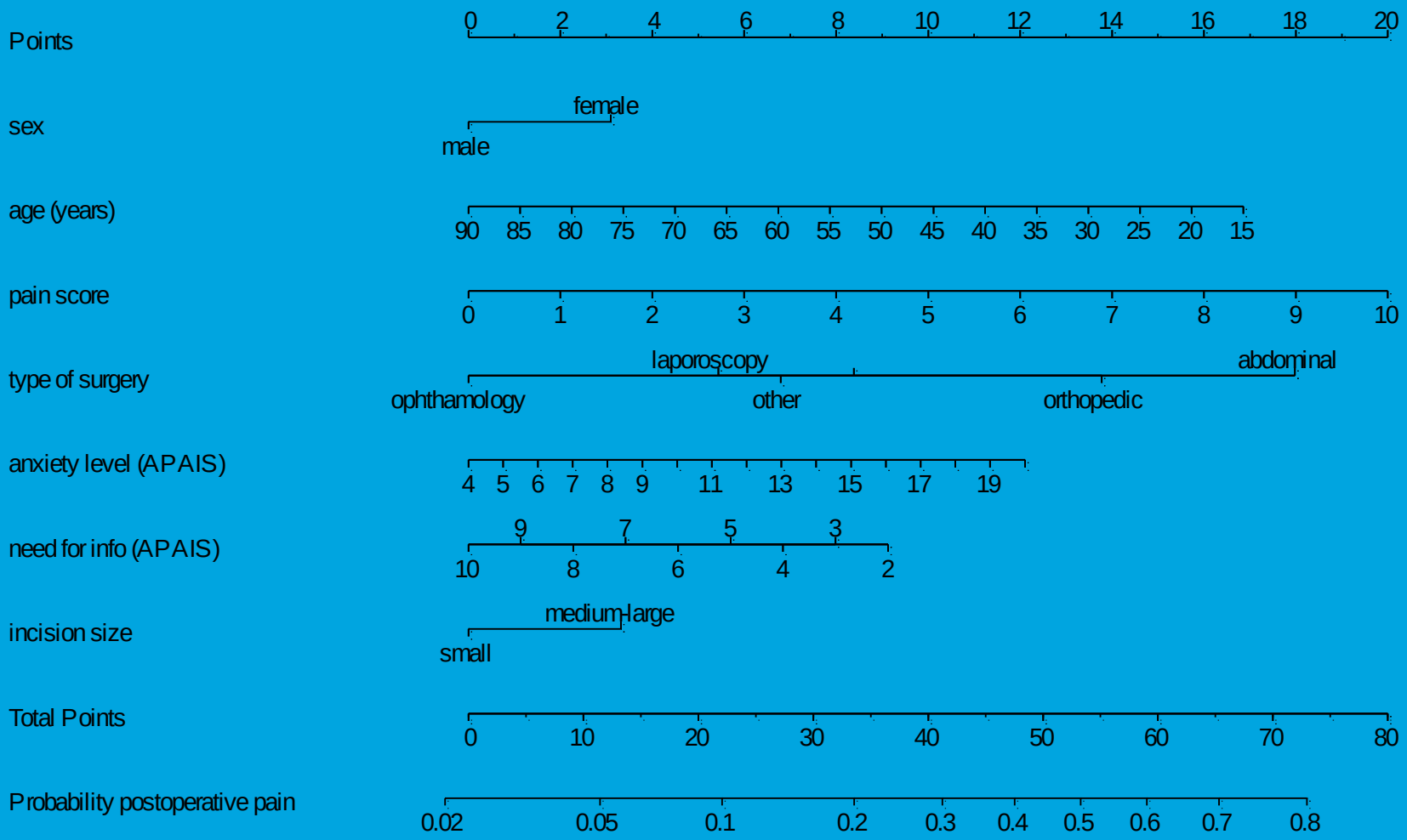
  - Adjust / shrink OR's and beta's

    - Heuristic shrinkage (van Houwelingen JC)

    - Bootstrapping techniques

Model development

- 6. Model presentation
  - *Original (shrunk beta's)*
  - *Nomogram*
  - *Simplified rule (like Apgar score)*
    - *Multiply beta's with 10 and round*
      - *Continuous variables first multply with value and then round*
    - *Give probabilities across scores a*
    - *Give c-statistic of simplified rule (loose accuracy usually)*


Points
0 2 4 6 8 10 12 14 16 18 20
sex
female
male
age (years)
90 85 80 75 70 65 60 55 50 45 40 35 30 25 20 15
pain score
0 1 2 3 4 5 6 7 8 9 10
type of surgery
laporoscopy
abdominal
ophthamology
other
orthopedic
anxiety level (APAIS)
4 5 6 7 8 9 11 13 15 17 19
need for info (APAIS)
9 7 5 3
10 8 6 4 2
incision size
medium-large
small
Total Points
0 10 20 30 40 50 60 70 80
Probability postoperative pain
0.02 0.05 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8

Points	0	1	2	3	4	5	6	7	8	12	13
Age, years*		<15	25	35	45	55	>65				
Primary Education only	N				Y						
Being single	N		Y								
Number of complaints	1			2				3+			
# consults past 12 months*	0	2	5	8	11	15	18	21	>24		
Depression past 12 months	N									Y	
Number of life events	0		1	2							3+

Age beta was 0.01; Consults beta was 0.03.

Model development

True challenges

- *Dealing with repeated measurements (predictors) /* time varying covariates
  - Missing values in these
- *Dealing with clustered data (IPD MA)*
- *Dealing with undergone treatments in case of prognostic prediction modeling*

What evidence needed to apply models in practice?

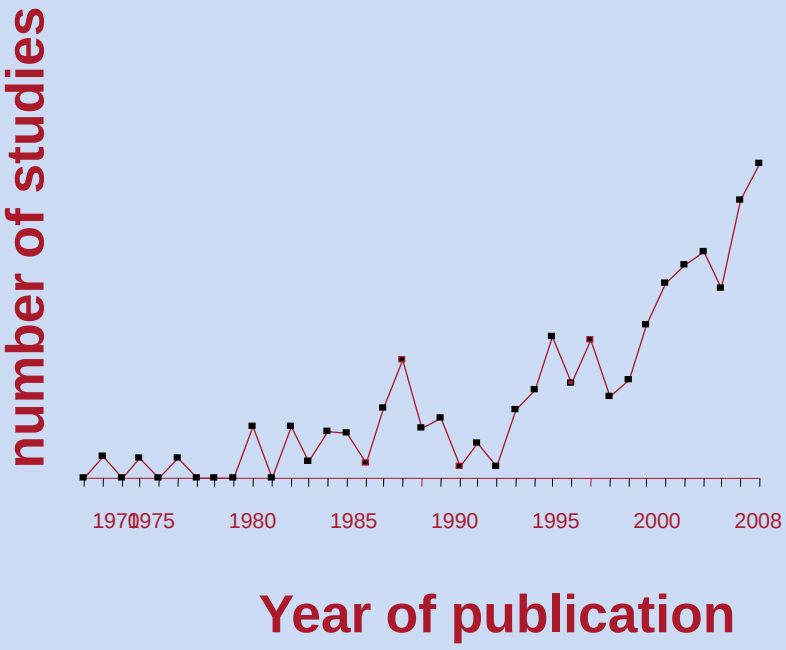
Steps in prediction modeling

- 1. Developing the prediction model
- **2. Validate the model in other subjects**
- 3. Update existing models to local situation
- 4. Quantify model's impact on doctor's decision making and on patient outcome (cost-effectiveness)

Phase 2. Validation studies

Unfortunately scarce

In contrast to development studies: sexy



Phase 2. Validation study characteristics

(Steyerberg + Moons Plos Med 2013, Altman Stat Med 2000+ BMJ 2009; Moons Heart 2012)

- Aim: to demonstrate accuracy/performance of original model in subjects not used to develop model
  - Calibration, discrimination (c-index), (re)classification

- Validating a model is not …
  - …Repeat analysis in new data and check if you come up with same predictors, regr. coeffs, predictive performance

  - …Fit the previously found predictors/model and estimate its predictive performance

Phase 2. Validation study characteristics

(Steyerberg + Moons Plos Med 2013, Altman Stat Med 2000+ BMJ 2009; Moons Heart 2012, JTH 2013)

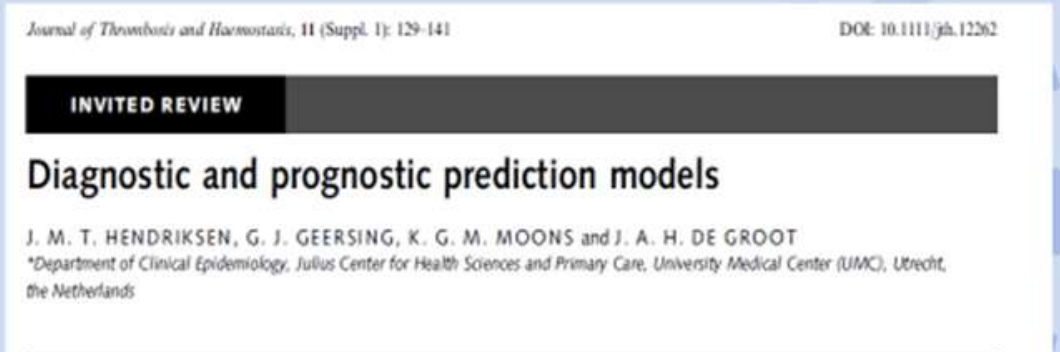
- Use original developed model → apply (!) to new data → Compare predicted with observed outcomes
  - Discrimination, calibration and (re)classification

- Validation studies thus require that original, developed prediction models properly reported
  - Original beta's – plus intercept (parametric survival)
    - Not just simplified score (too often still done)
  - Clear definition and measurement method of predictors + outcome (so future researchers can repeat/use them)
  - Reporting guideline underway: **TRIPOD (end 2013)**

Phase 2. Types of Validation studies

(Steyerberg + Moons Plos Med 2013, Altman Stat Med 2000+ BMJ 2009; Moons Heart 2012)

- **4 (increasingly stringent) types:**
  1. Internal validation (in fact part of development phase)
  2. Temporal validation
  3. Geographical validation
  4. Other setting / domain (type of patients)



Journal of Thrombosis and Haemostasis, 11 (Suppl. 1): 129–141 DOI: 10.1111/jth.12262

INVITED REVIEW

Diagnostic and prognostic prediction models

J. M. T. HENDRIKSEN, G. J. GEERSING, K. G. M. MOONS and J. A. H. DE GROOT
*Department of Clinical Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center (UMC), Utrecht, the Netherlands

Types of Validation Studies

1. Internal validation (split sample, bootstrapping)
   - Not random split sample → no difference
   - Best = Bootstrapping
     - Note: not new data (Bleeker SE et al, JCE 2002)

2. Temporal validation
   - Same setting, measurements and investigators (often), but later in time
     - Many similarities → 'high' chance of good performance
   - Split sample: if large database -- split over time

Types of Validation Studies

3. Geographic

- Other centers + often other investigators
- Also often other protocols
- May be – if very large database or combination of data sets (= IPD meta analysis) -- split sample by country

4. Setting/domain/subgroup

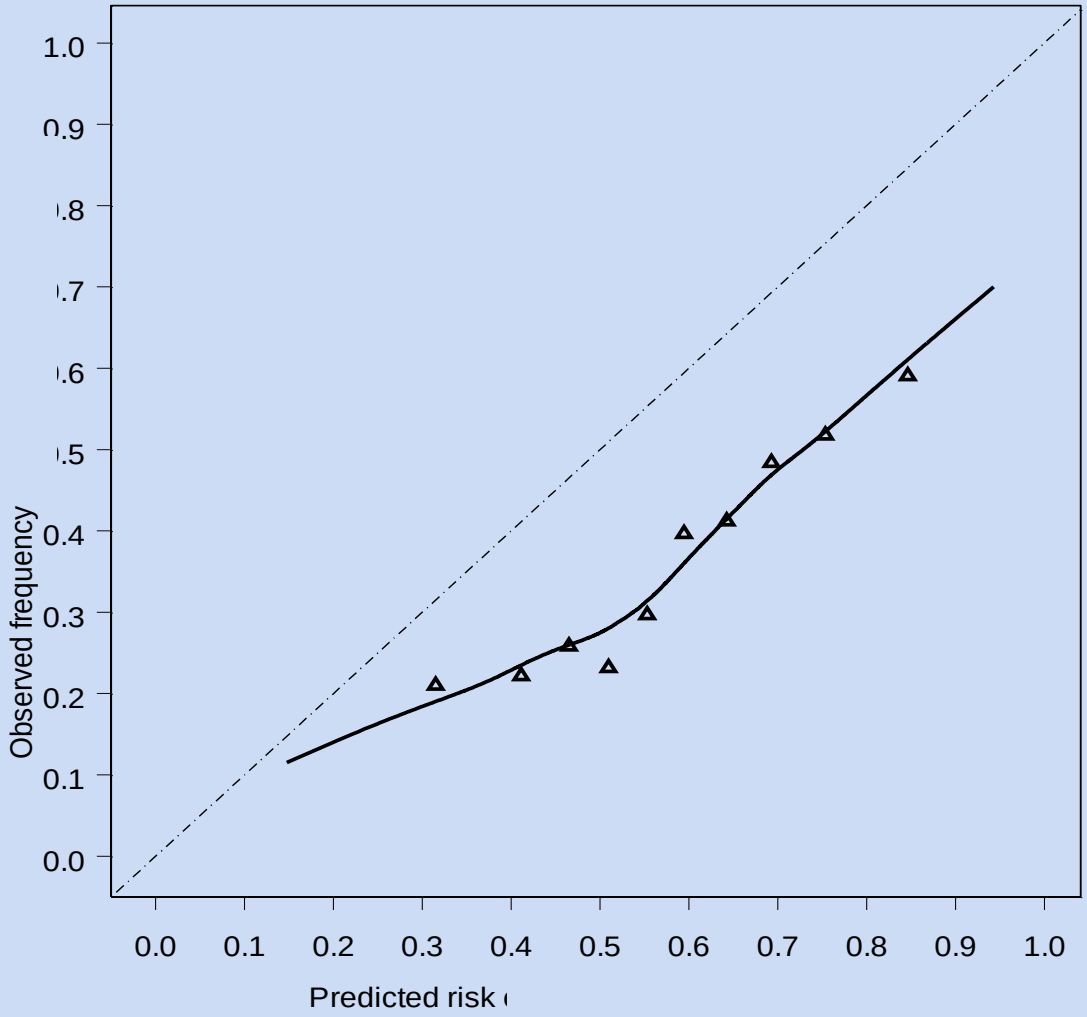
- Secondary → primary care
- Adults → children
- Men → women

Types of Validation Studies

- Note temporal, geographic and domain/setting validation can be done:
  - Prospectively
  - Retrospectively using large existing data sets
  - Often called 'external' validation

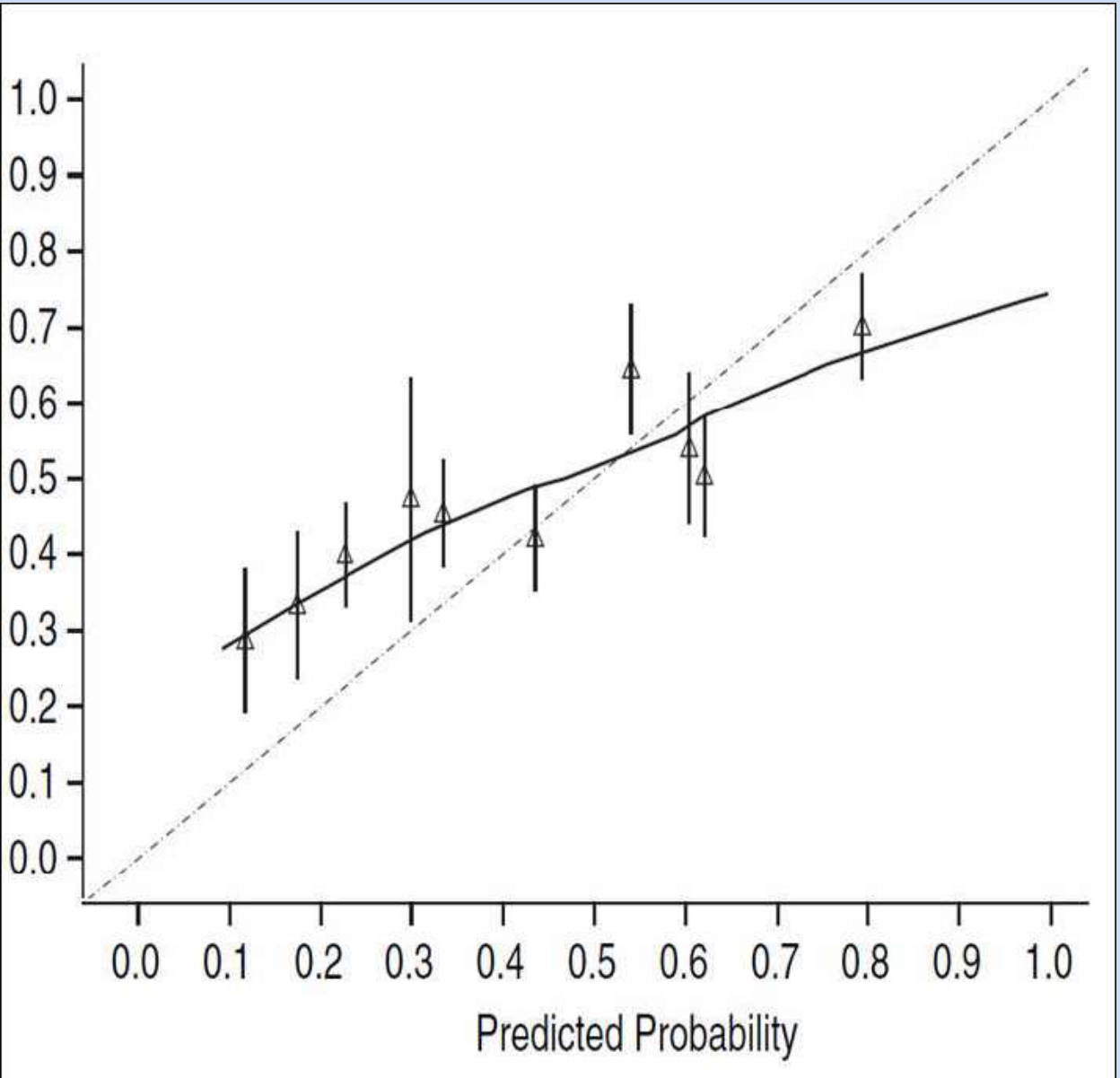
- YES: usually researchers find poor accuracy when validating existing model in their data
  - Key message: suppress your reflexes
  - Do not immediately fit (yet) a new model

Typical Result



- **Systematically too high predictions**
  - Higher outcome prevalence/incidence in development set
    - Intercept too large for new subjects

Typical 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)

Logical: reasons poor validation

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

- **1. Different outcome occurrence**
- **2. Different patients**

Reasons poor validation

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

3. Different interpretation of predictors

or (incorrect) proxies of predictors are used

4. Changes in care over time

- Improvement in measurements: e.g. imaging tests
  - Previous CTs less accurate than spiral CT for pulmonary embolism detection

5. Original model could have missed important predictor

Reasons poor validation

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

- **BUT: No matter what reason of poor validation:**
  - Reflex: one develops 'own new' model from their validation study data
  - >100 models for brain trauma; >60 models for breast cancer; >100 CVD risk in general population; > 100 diabetes models

- **Understandable:**
  - We finally learned the 'tricks' to develop models (in standard software)
  - 'Own' model makes you famous (Apgar; Goldman; Gail; Wells)
    - Validation is only to support (citation index of) others

Reasons poor validation

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

- **Unfortunate habit**
  - Previous knowledge neglected
  - Prediction research becomes completely particularistic
    - Every country, setting, hospital, subgroup, etc.
  - Validation data sets often smaller → even less generalisable models
  - Perhaps new model needed: but likely not!

What evidence needed to apply models in practice?

Steps in prediction modeling

- 1. Developing the prediction model
- 2. Validate the model in other subjects
- **3. Update existing models to local situation**
- 4. Quantify model's impact on doctor's decision making and on patient outcome (cost-effectiveness)

Phase 3. Updating prediction models

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

- Update/adjust existing model with new data → rather than fitting ('our') new model
  - Certainly if validation set is relatively small(er)

- Updating is particularly important when:
  - new predictors found → added to existing models
    - **CRP to Framingham risk model**
  - new data/patients available → dynamic prediction models

Phase 3. Updating prediction models

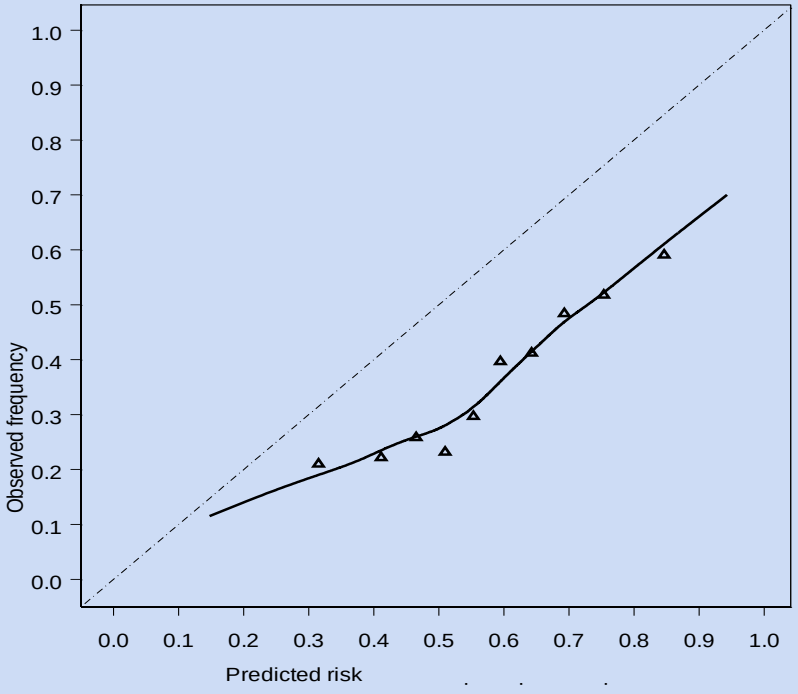
(Houwelingen Stat Med 2000; Steyerberg Stat Med 2004; KJM Janssen JCE 2008+CJA 2008; D Toll JCE 2008; Moons Heart 2012)

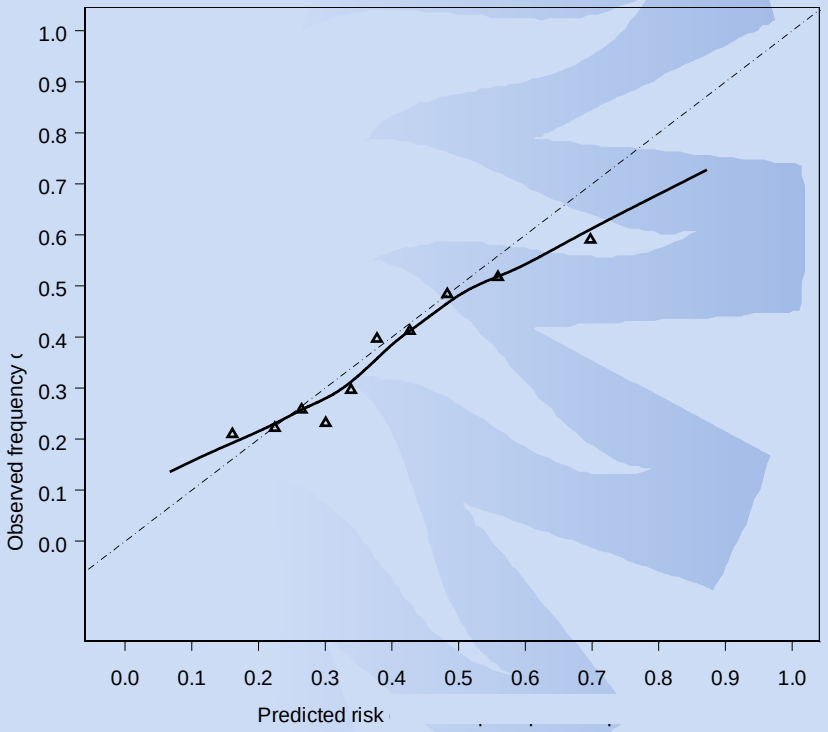
- **After validation existing model → unsatisfactory accuracy → update → ranges from:**
  - Simple adjustment of base line risk (intercept)
  - Adjusting the regression coefficients of predictors
    - All together in same way (if overfitted model)
    - Different adjustments
  - Adding previously missed or new predictors/markers

Phase 3. Updating prediction models

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

- Adjust for difference in overall prevalence/incidence (intercept adjustment) is often sufficient





- If also slope different → adjust predictor weights
- Or search for adding/new predictors

Phase 3. Updating prediction models

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

- **Final notes**

  - Updating done after (!) model (external) validation → if unsatisfactory accuracy in new subjects
    - Not recommend updating without first validating

  - Aim of validation studies is not to find similar predictive accuracy as in development set
    - But to find satisfactory accuracy in validation set
    - Depends on preferences/consequences of false predictions in validation situation
      - AUC of 0.60 is not per se bad

Phase 3. Updating prediction models

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

- **Final notes ctd**

  - For dynamic prediction models: validation and updating studies become even more important issues to address
    - Opportunity for continuous validation and updating

  - Challenge = dealing with missing predictor data

What evidence needed to apply prediction models in practice?

Steps in prediction modeling

- 1. Developing the prediction model
- 2. Validate the model in other subjects
- 3. Update existing models to local situation
- *4.* Quantify impact of using model/test/marker/test strategy on doctor's decision making and patient outcomes

Phase 4. Impact studies

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

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

- **… studied in so-called Impact studies**

Phase 4. Impact studies

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

- **Aim:** Whether actual use (!) of prediction model/test/marker truly improves …
  - … Physicians behaviour (treatment indications) …
  - … Patient outcome or health-care-costs …

  … as compared to not using such model/marker/test

- Impact studies are thus intervention studies
  - Intervention = use and subsequent treatment actions based on the model predictions

Phase 4. Impact studies

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

- **Design = like intervention studies**
  - When 'effects of some intervention on patient outcome' is mentioned → reflex = comparative stu → good reflex !
    - In sharp (!) contrast to previous prediction modeling phases
  - Second reflex = randomized comparison
  - Indeed: best design = RCT
    - Preferably cluster RCT (e.g. stepped wedge) trial *(Moons BMJ 2009 + Heart 2012)*
    - Randomising practices
      - Less contamination across doctors in same practice → reduced contrast
    - Not randomising patients
      - Learning effects of doctors → reduced contrast

Phase 4. Impact studies

(Campbell BMJ 2000; Reilly and 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 cluster RCT***

- **Thousands clinical prediction models → increase per day**
  - Simply not enough resources (budget plus people) to study them all in a long term, expensiv cluster RCT

Phase 4. Impact studies

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

- Before reflexing to RCTs → Alternative, cheaper/easier designs:
  - To better indicate which tests/markers/models should indeed undergo an RCT

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

Phase 4. Impact studies

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

- **2. Modeling study**

  - Risk-Benefits (decision) models:

    - Linked evidence approach -- combining predictive accuracy studies and RCTs
    - Use predictive probabilities of validated model

      +

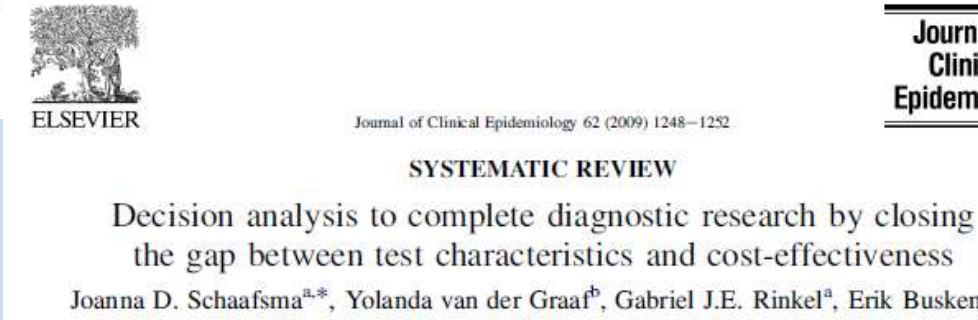
    - Results of beneftis and risks of existing therapies for that disease (e.g. obtained from RCTs)

    - → To quantify effect of actually using the model (or test/marker) with model-directed therapies on patient outcome

Phase 4. Impact studies

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

- Gives indication of expected risks/benefits when introducing model/test/marker combined with therapies
  - plus its cost-effectiveness
  - plus specific scenarios (e.g. treatment-probability thresholds) or subgroups may be tested
- Gives indication:
  - whether a real RCT is indicated or not
  - How to enrich the RCT design -- Eg excluding/focusing specific groups



Phase 4. Impact studies

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

- **3. Before-After study**
  - Compare patient outcomes in period before introducing model/test/marker with period afte introducing
  - E.g. Wells rule for DVT; Ottawa ankle/knee rule
- **4. External/historical control group**
- Disadvantages 3+4
  - Time changes (notaly in therapeutic guidelines/therapies)
  - Confounding by indication / case mix differences → adjustment in analysis (like non-randomized intervention studies)

Take home messages

- Number of markers increases per day
  - Simply enter market → overtesting → can't be all relevant

- No diagnosis or prognosis estimated by single test/marker
  - Marker always form (small) part of many results

- Added/independent value of a marker test is relevant to know for physicians → and thus to quantify in research
  - method: multivariable prediction modeling

Take home messages

- **Phased approach of prediction modeling:**
  - Development
  - Validation
  - Updating
  - Impact

- **Development No real challenges anymore → Notably:**
  - repeated or time varying predictors / missing data / multiple events
  - clustered data
  - 'confounding by treatment'

- **Validation studies much more needed**

Take home messages

- **Validation:**
  - Requires proper reporting of original developed models, plus how predictors and outcomes defined/measured
  - not only of simplified scores
  - 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 leads often to poor accuracy → do not panic → try an update first

- Impact studies are not per se large scale RCTs

- No developed model applied (or in guideline) without at least one external validation → preferably with some impact assessment

- We need more collaborative IPDs → to develop, externally validate and improve prediction models

  - **the more advanced our models – the higher this need**