UMC Utrecht
Julius Center

A framework for meta-analysis of prediction models for binary and time-to-event outcomes

Thomas Debray, PhD

Julius Center for Health Sciences and Primary Care,
University Medical Center Utrecht, The Netherlands

Cochrane Netherlands, Utrecht, The Netherlands

T.Debray@umcutrecht.nl

Disclosure

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

Risk prediction

- Quantify individual prognosis
  (e.g. probability of developing an adverse event)
- Use of multiple prognostic/predictive factors
  - Subject characteristics
  - History and physical examination results
  - Imaging results
  - (Bio)markers

Identification of high risk individuals

SCORE

15% and over
10%–14%
5%–9%
3%–4%
2%
1%
<1%

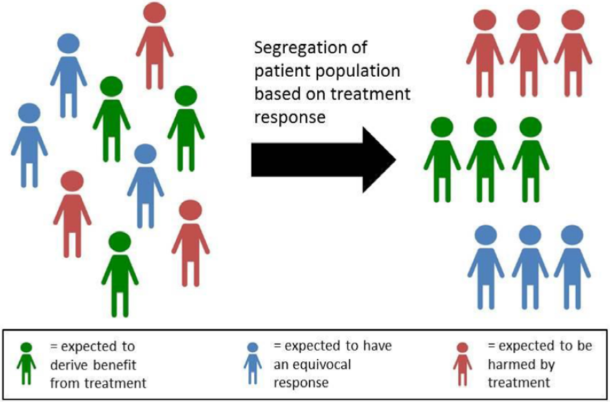
10-year risk of fatal CVD in populations at high CVD risk

		WOMEN										MEN									
Age	Systolic blood pressure (mmHg)	Non-smoker					Smoker					Non-smoker					Smoker				
		4	5	6	7	8	4	5	6	7	8	4	5	6	7	8	4	5	6	7	8
65	180	7	8	9	10	12	13	15	17	19	22	14	16	19	22	26	26	30	35	41	47
	160	5	5	6	7	8	9	10	12	13	16	9	11	13	15	16	18	21	25	29	34
	140	3	3	4	5	6	6	7	8	9	11	6	8	9	11	13	13	15	17	20	24
	120	2	2	3	3	4	4	5	5	6	7	4	5	6	7	9	9	10	12	14	17
60	180	4	4	5	6	7	8	9	10	11	13	9	11	13	15	18	18	21	24	28	33
	160	3	3	3	4	5	5	6	7	8	9	6	7	9	10	12	12	14	17	20	24
	140	2	2	2	3	3	3	4	5	5	6	4	5	6	7	9	8	10	12	14	17
	120	1	1	2	2	2	2	3	3	4	4	3	3	4	5	6	6	7	8	10	12
55	180	2	2	3	3	4	4	5	5	6	7	6	7	8	10	12	12	13	16	19	22
	160	1	2	2	2	3	3	3	4	4	5	4	5	6	7	8	8	9	11	13	16
	140	1	1	1	1	2	2	2	2	3	3	3	3	4	5	6	5	6	8	9	11
	120	1	1	1	1	1	1	1	2	2	2	2	2	3	3	4	4	4	5	6	8
50	180	1	1	1	2	2	2	2	3	3	4	4	4	5	6	7	7	8	10	12	14
	160	1	1	1	1	1	1	2	2	2	3	2	3	3	4	5	5	6	7	8	10
	140	0	1	1	1	1	1	1	1	1	2	2	2	2	3	3	3	4	5	6	7
	120	0	0	1	1	1	1	1	1	1	1	1	1	2	2	2	2	3	3	4	5
40	180	0	0	0	0	0	0	0	0	1	1	1	1	1	2	2	2	2	3	3	4
	160	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	2	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	1	1	1	1	1	1	1	1

Total cholesterol (mmol/L)

150 200 250 300
mg/dL

Prediction of individual treatment response



Source: Yeh RW & Kramer DB. Circulation. 2017;135:1097–1100

What is a good model?

Accurate predictions

Ability to distinguish between low and high risk patients

Good and consistent performance across different settings and populations

Influence decision making

Evaluation of model performance

Validation of prediction models increasingly common!

- **38 validations** of Framingham Risk Score (Wilson 1998)
- **31 validations** of Pooled Cohort Equations: estimate risk of cardiovascular disease
- **22 validations** of EuroSCORE II: estimate risk of operative mortality in patients undergoing cardiac surgery
- **19 validations** of CHA2DS2-VASc: estimate stroke risk in patients with atrial fibrillation

Need for systematic review & meta-analysis

Validation studies often yield conflicting results due to

- Differences in studied populations
- Differences in methodological standards

Recent guidance (BMJ 2017)

RESEARCH METHODS AND REPORTING

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

Thomas P A Debray,[1,2] Johanna A A G Damen,[1,2] Kym I E Snell,[3] Joie Ensor,[3] Lotty Hooft,[1,2] Johannes B Reitsma,[1,2] Richard D Riley,[3] Karel G M Moons[1,2]

Motivating example

Framingham Risk Score (Wilson et al. 1998)

- Model type: Cox regression
- Outcome: Fatal or non-fatal coronary heart disease (CHD)
- Timing: Initial CHD within 10 years
- Evidence: 24 validations in male populations

Summarize estimates of model performance

- Concordance statistic (`cstat`)
- Ratio of observed versus expected events (`OE`)
- Calibration slope (`slope`)

Focus of today on data extraction and meta-analysis

Data extraction

Poor and inconsistent reporting of prediction model performance.

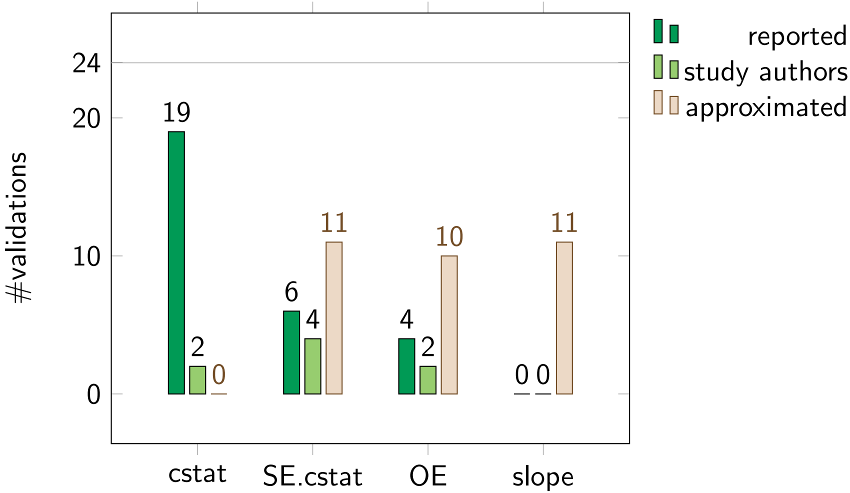
- Poor study design
- Inappropriate handling and acknowledgement of missing data
- Calibration often omitted from the publication

Data extraction

Need to restore missing information

- `cstat` can be approximated from the distribution of the linear predictor
- SE of `cstat` can be approximated from the c-statistic, total sample size and total # evens
- `OE` and its SE can be estimated from reported event counts and/or survival curves
- `slope` and its SE can be estimated from reported event counts across risk strata (e.g. as presented in calibration tables)

Data extraction: motivating example



For 10 studies, calibration performance was only available for <10 years follow-up.

Meta-analysis

- Performance measures such as `cstat` and `OE` are very sensitive to patient spectrum, and therefore likely to vary across studies
- Additional uncertainty due to approximations
- Need for weakly informative prior distributions in Bayesian estimation

Meta-analysis of the c-statistic

Statistical model

$$\text{logit}(c_i) \sim \mathcal{N}\left(\mu_{\text{discr}}, \text{Var}\left(\text{logit}(c_i)\right) + \tau^2_{\text{discr}}\right)$$

Estimation	K	**Summary**	**95% CI**	**95% PI**
REML	21	0.69	0.66 – 0.71	0.59 – 0.77
Bayesian (Unif)	24	0.69	0.66 – 0.71	0.59 – 0.78
Bayesian (Student-t)	24	0.69	0.66 – 0.71	0.59 – 0.78

For 3 studies, we did not have information on c_i but could nevertheless approximate $\text{SE}(c_i)$.

Meta-analysis of the total O:E ratio

We can use different models to account for sampling variability:

Option 1 $\ln(\mathrm{O{:}E})_i \sim \mathcal{N}\left(\mu_{\mathrm{cal.OE}}, \mathrm{Var}\left(\ln(\mathrm{O{:}E})_i\right) + \tau^2_{\mathrm{cal.OE}}\right)$

Option 2 $O_i \sim \mathrm{Binom}\left(N_i, p_{\mathrm{O},i}\right)$

$E_i \sim \mathrm{Binom}\left(N_i, p_{\mathrm{E},i}\right)$

$\ln\left(p_{\mathrm{O},i}/p_{\mathrm{E},i}\right) \sim \mathcal{N}\left(\mu_{\mathrm{cal.OE}}, \tau^2_{\mathrm{cal.OE}}\right)$

Option 3 $O_i \sim \mathrm{Poisson}\left(E_i \exp(\eta_i)\right)$

$\eta_i \sim \mathcal{N}\left(\mu_{\mathrm{cal.OE}}, \tau^2_{\mathrm{cal.OE}}\right)$

For all models, the interpretation of $\mu_{\mathrm{cal.OE}}$ is the same.

Meta-analysis of the total O:E ratio

Estimation	K	Summary	95% CI	95% PI	
REML[1]	6	0.56	0.28 – 1.16	0.09 – 3.62	
Bayesian[1] (Unif)	6	0.61	0.19 – 1.08	0.00 – 2.84	
Bayesian[1] (Student-t)	6	0.61	0.20 – 1.07	0.00 – 2.63	
ML[3]	6	0.56	0.25 – 1.26	0.03 – 11.29	$\star$
Bayesian[3] (Unif)	7	0.60	0.19 – 1.09	0.00 – 2.91	
Bayesian[3] (Student-t)	7	0.60	0.18 – 1.05	0.00 – 2.67	

When applying extrapolation, we have 10 additional studies for meta-analysis (similar results).

Meta-analysis of the calibration slope

Statistical model

$$O_{ij} \sim \mathrm{Binom}(N_{ij}, p_{\mathrm{O},ij})$$
$$\mathrm{logit}(p_{\mathrm{O},ij}) = \alpha_i + \beta_i \, \mathrm{logit}(P_{\mathrm{E},ij})$$
$$\beta_i \sim \mathcal{N}(\mu_{\mathrm{cal.slope}}, \tau^2_{\mathrm{cal.slope}})$$

Estimation	K	**Summary**	**95% CI**	**95% PI**
ML	3	1.03	0.90 – 1.16	0.20 – 1.87
Bayesian†	3	1.05	0.47 – 1.64	-0.01 – 2.22
Bayesian‡	3	1.05	0.51 – 1.65	-0.06 – 2.17

When applying extrapolation, we have 8 additional studies for meta-analysis (similar results but smaller intervals).

R package “metamisc”

- Assist in data preparation & meta-analysis
- Illustrative examples

metamisc: Diagnostic and Prognostic Meta-Analysis

Meta-analysis of diagnostic and prognostic modeling studies. Summarize estimates of diagnostic test accuracy and prediction model performance. Validate, update and combine published prediction models.

Version:	0.1.6
Depends:	R (≥ 2.10), stats, graphics
Imports:	metafor, mvtnorm, ellipse, lme4
Suggests:	runjags, rjags
Published:	2017-09-06
Author:	Thomas Debray [aut, cre], Valentijn de Jong [aut]
Maintainer:	Thomas Debray <thomas.debray at gmail.com>
License:	GPL-2
URL:	http://r-forge.r-project.org/projects/metamisc/
NeedsCompilation:	no
In views:	MetaAnalysis
CRAN checks:	metamisc results

Downloads:

Reference manual:	metamisc.pdf
Package source:	metamisc_0.1.6.tar.gz
Windows binaries:	r-devel: metamisc_0.1.6.zip, r-release: metamisc_0.1.6.zip, r-oldrel: metamisc_0.1.6.zip
OS X El Capitan binaries:	r-release: metamisc_0.1.6.tgz
OS X Mavericks binaries:	r-oldrel: metamisc_0.1.6.tgz
Old sources:	metamisc archive

Linking:

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

Final remarks

- Meta-analysis of model performance often feasible & helpful
- Despite poor reporting, key performance estimates can be retrieved and summarized
- Bayesian estimation methods recommended to fully propagate uncertainty arising from data restoration
- Presence of statistical heterogeneity most likely
- Straightforward extension to meta-regression and multivariate meta-analysis