UMC Utrecht
Julius Center

Meta-analysis of external validation studies

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

Prediction

- Turn available information about individuals into a statement about the probability:
  ...of having a particular disease → **diagnosis**
  ...of developing a particular event → **prognosis**
- Use of multiple predictors
  - Subject characteristics
  - History and physical examination results
  - Imaging results
  - (Bio)markers
- Typical aims
  - to inform patients and their families
  - to guide treatment and other clinical decisions
  - to create risk groups

What is a good model?

Accurate predictions

Good and consistent performance across different settings and populations

Ability to distinguish between low and high risk patients

Influence decision making

Numerous models for same target population + outcomes

1995

"We believe that the main reasons why doctors reject published prognostic models are lack of clinical credibility and lack of evidence that a prognostic model can support decisions about patient care."

Wyatt JC, Altman DG. Commentary: Prognostic models: clinically useful or quickly forgotten? BMJ. 1995;311(7019):1539–41.

Numerous models for same target population + outcomes

1995 2012

"Comparing risk prediction models should be routine when deriving a new model for the same purpose"

Collins GS, Moons KGM. Comparing risk prediction models. BMJ. 2012;344:e3186–e3186.

Numerous models for same target population + outcomes

"There is an excess of models predicting incident CVD in the general population. The usefulness of most of the models remains unclear."

Damen JAAG, Hooft L, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. BMJ. 2016;353:i2416.

Numerous models for same target population + outcomes

Formal guidance for systematic review and meta-analysis

RESEARCH METHODS AND REPORTING

CrossMark
click for updates

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

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

Motivating example

Previous guidance focused on meta-analysis of logistic regression models. Hence, focus of today is on **survival models**.

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`)

Statistical framework - data extraction

Key problem: Poor and inconsistent reporting of prediction model performance.

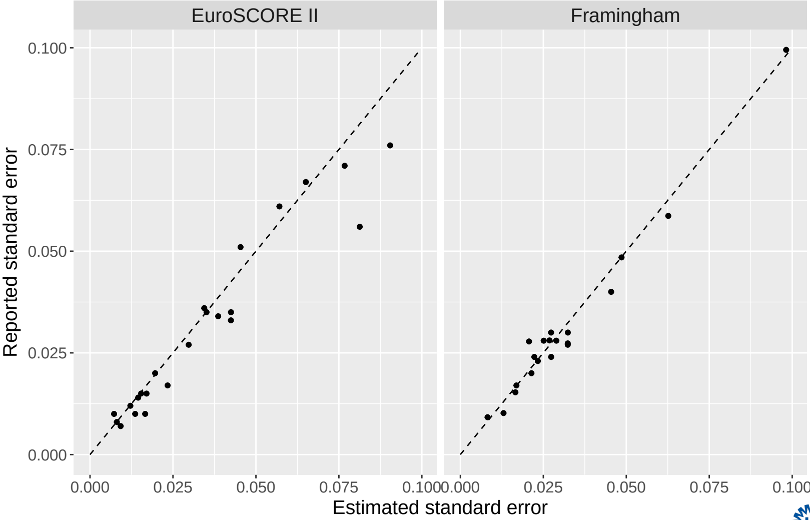
Standard error of the c-statistic

$$\mathrm{SE}(c) \approx \sqrt{\frac{c\,(1-c)\left[1 + n^{*}\frac{1-c}{2-c} + \frac{m^{*}c}{1+c}\right]}{mn}}$$

with c the reported c-statistic, n the number of observed events, m the total number of non-events and $m^{*} = n^{*} = \frac{1}{2}(m+n) - 1$.

Newcombe RG. Confidence intervals for an effect size measure based on the Mann-Whitney statistic. Part 2: asymptotic methods and evaluation. Stat Med. 2006;25(4):559–73.

Estimated versus reported standard error of the c-statistic



Statistical framework - data extraction

Ratio of observed versus expected events

- Reference value: 1
- Observed survival probability ($S_{\mathrm{KM},t}$)
- Expected (predicted) event rate ($P_{\mathrm{E},t}$)

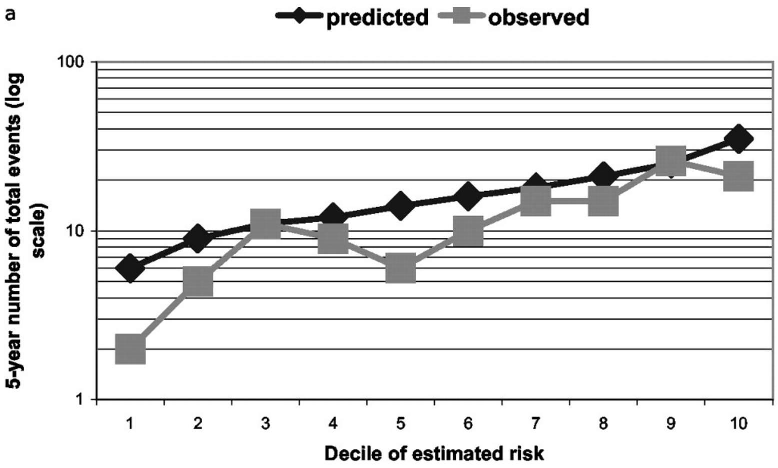
$$(\mathsf{O{:}E})_t = \frac{1 - S_{\mathrm{KM},t}}{P_{\mathrm{E},t}}$$

$$\mathrm{SE}(\mathsf{O{:}E})_t = \frac{1}{P_{\mathrm{E},t}} \mathrm{SE}(S_{\mathrm{KM},t})$$

If unavailable, $\mathrm{SE}(S_{\mathrm{KM},t})$ can be approximated as well:

$$\mathrm{SE}(S_{\mathrm{KM},t}) \approx \sqrt{\frac{S_{\mathrm{KM},t}(1 - S_{\mathrm{KM},t})}{N_t}}$$

Extraction of event rates: an example



Statistical framework - data extraction

Calibration slope

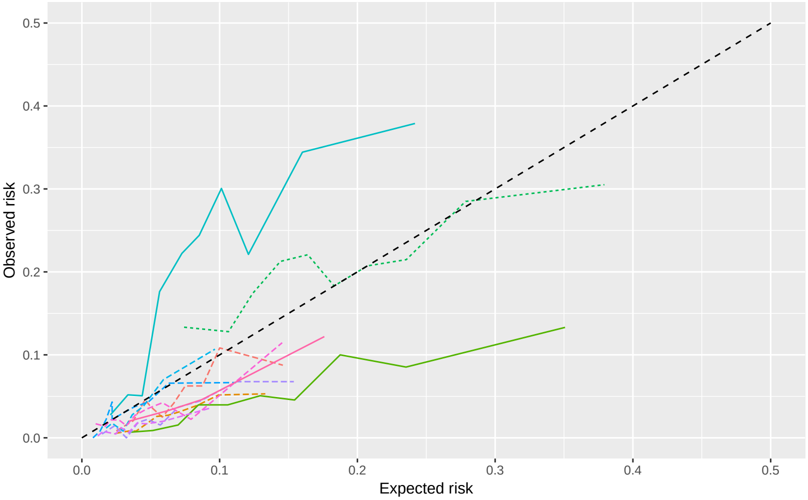
- Reference value: 1
- The calibration slope β is calculated as follows using IPD:

$$y_k \sim \mathrm{Bernoulli}(p_k)$$
$$\mathrm{logit}(p_k) = \alpha + \beta \, \mathrm{LP}_k$$

- When missing, we can derive β from reported event counts across j risk strata (e.g. as presented in calibration tables):

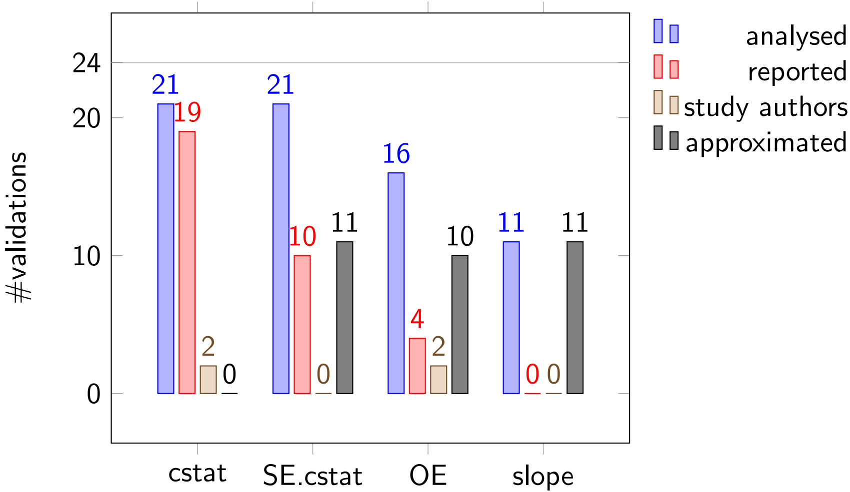
$$O_j \sim \mathrm{Binom}(N_j, p_{\mathrm{O},j})$$
$$\mathrm{logit}(p_{\mathrm{O},j}) = \eta + \beta \, \mathrm{logit}(P_{\mathrm{E},j})$$

Extracted risk estimates for the Framingham Risk Score



The diagonal line indicates perfect calibration. Risk estimates were reported for 5 years follow-up (dashed lines), 7.5 years follow-up (dotted lines) and 10 years follow-up (full lines).

Statistical framework - data extraction



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

Statistical framework - meta-analysis models

Meta-analysis of the c-statistic

- Need to apply logit transformation

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

- Use delta method to derive $\mathrm{Var}\left(\mathrm{logit}(c_i)\right)$ from $\mathrm{SE}(c_i)$
- For Bayesian models, we propose weakly informative priors:
  - $\mu_{\mathrm{discr}} \sim \mathcal{N}(0, 10^6)$
  - $\tau_{\mathrm{discr}} \sim \mathrm{Unif}(0, 2)$
  - $\tau_{\mathrm{discr}} \sim$ Student-$t(0, 0.5^2, 3)$ $T[0, 10]$

  based on empirical data from 26 meta-analyses (with $\hat{\tau}_{\mathrm{discr}}$ ranging from 0 to 0.50).

Statistical framework - meta-analysis models

Meta-analysis of the c-statistic

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 $\mathrm{SE}(c_i)$.

Statistical framework - meta-analysis models

Meta-analysis of the total O:E ratio

We can use different models to account for sampling variability:

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

Option 2
$$O_i \sim \text{Binom}\left(N_i, p_{\text{O},i}\right)$$
$$E_i \sim \text{Binom}\left(N_i, p_{\text{E},i}\right)$$
$$\ln\left(p_{\text{O},i}/p_{\text{E},i}\right) \sim \mathcal{N}\left(\mu_{\text{cal.OE}}, \tau^2_{\text{cal.OE}}\right)$$

Option 3
$$O_i \sim \text{Poisson}\left(E_i \exp(\eta_i)\right)$$
$$\eta_i \sim \mathcal{N}\left(\mu_{\text{cal.OE}}, \tau^2_{\text{cal.OE}}\right)$$

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

Statistical framework - meta-analysis models

Meta-analysis of the total O:E ratio (c' ed)

- For Bayesian models, we propose weakly informative priors:
  - $\mu_{\text{cal.OE}} \sim \mathcal{N}(0, 10^6)$
  - $\tau_{\text{cal.OE}} \sim \text{Unif}(0, 2)$
  - $\tau_{\text{cal.OE}} \sim \text{Student-}t(0, 1.5^2, 3)\ T[0, 10]$

  based on empirical data from 16 meta-analyses
  (with $\hat{\tau}_{\text{cal.OE}}$ ranging from 0 to 1.39).
- Possible to extrapolate event rates at time I to time t using:

$$p_t = 1 - S_{\text{KM},t} = 1 - \exp\left(\frac{t \ln(1 - p_I)}{I}\right)$$

Straightforward to integrate in Bayesian estimation framework

Statistical framework - meta-analysis models

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).

Statistical framework - meta-analysis models

Meta-analysis of the calibration slope

- No transformations needed
- Rely on binomial approximation

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

- For Bayesian models, we propose weakly informative priors:
  - $\mu_{\mathrm{cal.slope}} \sim \mathcal{N}(0, 10^6)$
  - $\tau_{\mathrm{cal.slope}} \sim \mathrm{Unif}(0, 2)$
  - $\tau_{\mathrm{cal.slope}} \sim \text{Student-}t(0, 1.5^2, 3)\ T[0, 10]$

Statistical framework - meta-analysis models

Meta-analysis of the calibration 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).

Final remarks

- Substantial efforts often needed to restore missing information
- Bayesian estimation methods recommended to fully propagate uncertainty arising from data restoration
- Development of R package `metamisc` to assist in data preparation and meta-analysis
- Presence of statistical heterogeneity most likely
- Straightforward extension to meta-regression and multivariate meta-analysis