NEW SIMPLIFIED APPROACH TO THE POOLED ANALYSIS OF CALIBRATION OF CLINICAL PREDICTION RULES FOR SYSTEMATIC REVIEWS OF VALIDATION STUDIES

Borislav D Dimitrov [1,2*], Nicola Motterlini [2], Tom Fahey [2]

[1] Academic Unit of Primary Care and Population Sciences, University of Southampton, UK; [2] HRB Centre for Primary Care Research, Department of General Practice, Royal College of Surgeons in Ireland, Dublin, Ireland

Introduction

Clinical prediction rules (CPRs) are important tools for optimisation of diagnosis and clinical management, especially in primary care, with few obligatory phases before their implementation (**Figure 1**).

CPR derivation is based on multivariable regression modelling to compute predicted probabilities of outcome and stratify patients in various risk groups. The CPR performance (discrimination, calibration) and its generalisability ("transportability") are assessed through validation as illustrated by a *flow-chart** algorithm of CPR development.

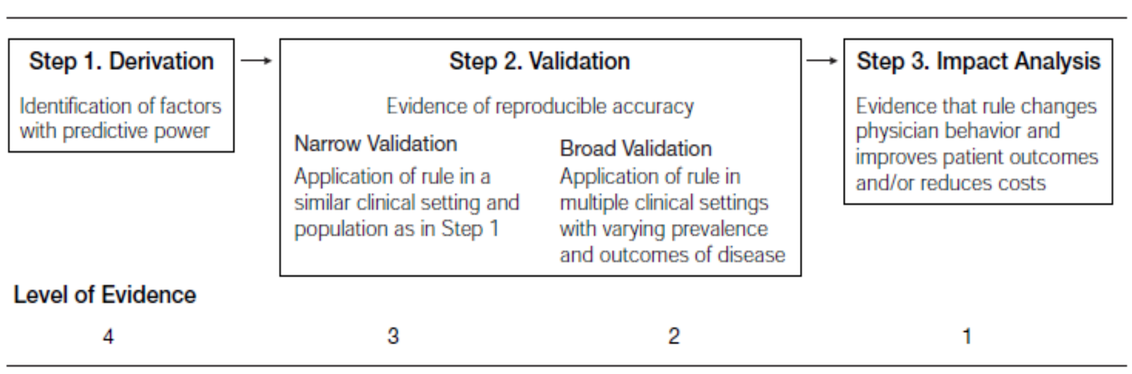

Figure 1 Steps in the development of a clinical prediction rule

Problem: The assessment of calibration cannot be done if no predicted values are published in or accessible from the validation studies.

Aim: The present work aimed at introducing a new, simple methodology which, by using derivation study information (via a "derivation model"), allows a calculation of predicted values in validation studies of CPRs.

Methods

We used ABCD^2 rule to predict strokes at 7 days following a transient ischaemic attack (TIA) according to 3 risk strata (scores 0-3, 4-5 and 6-7). The original distribution (Col.3, **Table 1**) was used as "derivation" (predictive) model - to predict the validation cohort strokes we applied the proportionate risk estimate (Col.5): low (1.35%), intermediate (6.51%), high (11.30%) risk. We compared the strokes in validation cohort as predicted by the "derivation" model (Col.6) to observed strokes (Col.7).

As a confirmation, the estimates in the validation study Y_{VAL} were calculated by a linear equation, where a_{DER} (-4.29) is intercept and $ß_{INT,DER}$ (1.63) and $ß_{HIGH,DER}$ (2.23) are derivation study coefficients obtained by LR model. $X_{INT,VAL}$ and $X_{HIGH,VAL}$ are ABCD^2 values as dummy variables ("intermediate" and "high" risk) from the validation study. In this way the predicted probability of stroke at level P, i.e., for each simulated patient, was computed as:

$$P = \frac{e^{Y_{VAL}}}{(1+e^{Y_{VAL}})}$$

Table 1 Observed and predicted number of strokes in validation sample [n=962, California Clinic cohort] using the distribution patterns of derivation sample strokes [n=1707, California ED; & n=209, Oxford population-based cohorts; *Johnston et al, 2007*] as a predictive model

	Derivation study		Validation study			
Stroke risk by ABCD^2 rule (score levels)	Patients (N)	Observed strokes n (%)	Patients (N)	Predicted incidence (%)*	Predicted number (n)	Observed number (n)**
Low risk (0-3 points)	520	7 (1.35)	426	1.35	5.8 (≈6)	2
Intermediate risk (4-5 points)	921	60 (6.51%)	397	6.51	25.8 (≈26)	17
High risk (6-7 points)	469	53 (11.30%)	139	11.30	15.7 (≈16)	10

Note: *Stroke incidence in each risk stratum of the validation study (data from California, USA) according to the distribution patterns of stroke in the original, derivation study (as used as a predictive model); **actual number of strokes as reported in each stratum of risk.

We added individual Ps to predict strokes: low 0-3 (expected=5.7), intermediate 4-5 (25.9) and high 6-7 (15.7) risk, i.e., 6, 26 and 16. The meta-analysis of predicted:observed ratios in the validation studies provided pooled RRs, measures of discrimination, calibration and heterogeneity with fixed and random-effects estimates. Calibration was corrected by updating the intercept of the original LR model - we "adjusted" the mean predicted probability to become equal to the frequency of observed outcomes.

Results

We described a new, simplified approach to compute predicted values and derive "predicted:observed" ratio of outcomes in validation studies of CPRs.

As an example, the analysis of ABCD^2 rule employed simulated individual-patients data (IPD) sets from the derivation and validation studies (136 to 1054 patients). The Forest plot (**Figure 2**) illustrates summary calibration estimates (pooled RRs, 95% CIs).

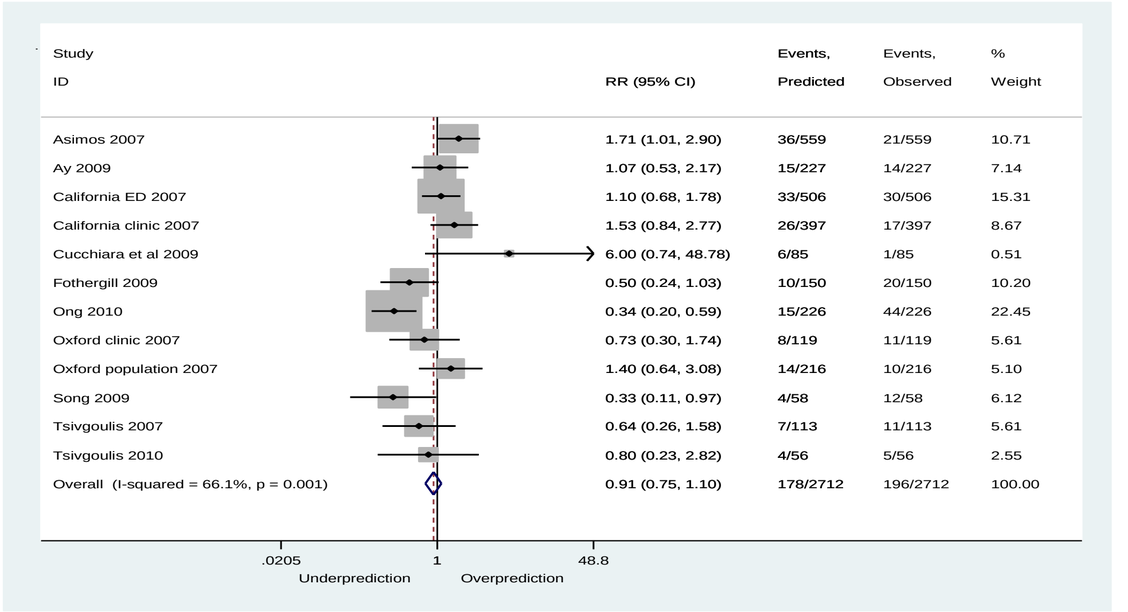

Figure 2 Forest plots of ABCD^2 rule (intermediate risk) – original CPR (fixed effects)

Along good discrimination (c-statistics Є 0.608-0.819), we identified low calibration levels (slight under-prediction of stroke risk with RR Є 0.73-0.91), with increased heterogeneity (18.3-66.1%) at different ABCD^2 levels (**Table 2**).

After adjustment of the original model intercept, while discrimination has not improved further, better calibration (Hosmer-Lemeshow "goodness-of-fit" p-values) and improved pooled estimates (RR Є 0.90-1.06), with narrower 95%CIs and zero heterogeneity, were achieved (Table 2).

Table 2 Meta analysis with pooled RRs and 95% CIs* from the validation studies of the ABCD^2 rule - comparison between our new approach (original CPR) and an updated logistic regression models

Stroke risk by ABCD^2 rule (score levels)	No adjustment (original CPR)			Adjustment of intercept	
	I^2	Fixed effects	Random effects	I^2	Fixed effects
Low risk (0-3 points)	18.3%	0.73 (0.45-1.20)	0.78 (0.41-1.48)	0.0%	0.90 (0.57-1.41)
Intermediate risk (4-5 points)	66.1%	0.91 (0.75-1.11)	0.88 (0.61-1.28)	0.0%	1.06 (0.88-1.28)
High risk (6-7 points)	52.6%	0.85 (0.68-1.06)	0.79 (0.55-1.15)	0.0%	0.95 (0.77-1.17)

Note: *Abbreviations: RR, risk ratio, CI, confidence interval; I^2, coefficient of heterogeneity.

Discussion

What does this methodological study contribute?

Our new approach is very useful in the development of CPRs for:

- Derivation of predicted values by everyone, when using data from a derivation study alone, without the need for highly-specialised knowledge or sophisticated statistical software;
- Assessment of calibration and meta-analysis of validation studies;
- Signalling mis-calibration and its improvement by updating;
- Confirmation of construct validity by comparison with predicted values, obtained by other methods;
- Further testing, refinement and improvement of CPR transportability.

Acknowledgement: This work was supported by the Health Research Board (HRB) of Ireland under Grant HRC/2007/1.

***Correspondence to:** Dr Borislav Dimitrov, MD MSc SMHM DM/PhD, Senior Lecturer in Medical Statistics, Academic Unit of Primary Care and Population Sciences, University of Southampton, UK, email: b.dimitrov@soton.ac.uk *(The flow-chart and references are available from the authors)*