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

Borislav D Dimitrov*, Nicola Motterlini, Tom Fahey

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, clinical management, especially in primary care. Three obligatory phases exist in CPR development before it can be implemented (**Figure 1**). An algorithm of CPR derivation and validation, including updating is summarised on a flow-chart*.

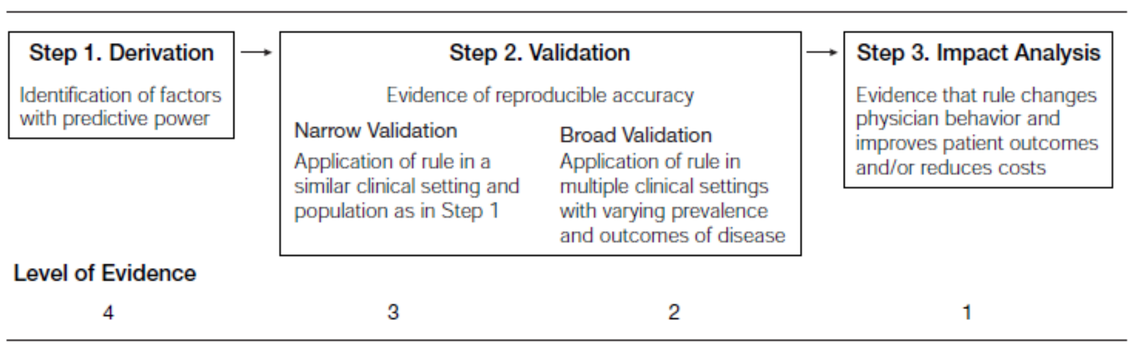

Fig.1 Steps in the development of a clinical prediction rule

CPR derivation is based on multivariable regression modelling to compute predicted probabilities of outcome and stratify patients into subgroups (low, intermediate or high risk). After derivation, the CPR performance (discrimination, calibration) is assessed through validation. The internal validation may use new data from the same source, but the true evaluation of CPR performance and its generalisability ("transportability") requires data from another clinical setting (external validation). However, such calibration assessment cannot be done if no predicted values are published or accessible from the validation studies.

The ***aim*** of the present study was to introduce and describe a new, simple methodology which, using information from the derivation study (referred to as a "derivation model"), allowed a calculation of predicted values in the validation studies.

Methods

Our approach employed outcome distributions in derivation and validation studies of ABCD^2 rule to predict strokes at 7 days after transient ischaemic attack (TIA) according to 3 risk strata (scores 0-3, 4-5 and 6-7). We used the original distribution (Col.3, **Table 1**) as "derivation" (predictive) model to which a validation study had to be related. To predict the strokes in validation cohort, the proportionate risk estimate from derivation cohort was applied in each risk stratum: low (1.35%), intermediate (6.51%), high (11.30%) risk (Col.5). The strokes in the risk strata of the validation cohort as predicted by the ABCD^2 "derivation model" (Col.6) were then compared to the observed strokes (Col.7).

To confirm above findings, we obtained logistic regression (LR) model and its coefficients from derivation data. ABCD^2 variable was converted into two dummy, related dichotomous variables (score 4-5: "intermediate" and 6-7 score: "high" risk) as single predictors in a derivation study multivariable model. For instance, if ABCD^2 score in derivation study was either 0, 1, 2 or 3 points (score 0-3: low risk), this patient got Y = -4.29 resulting in probability of 1.35% (95% CI 0.6-2.8) for stroke in the next 7 days.

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

Stroke risk by ABCD^2 rule (score levels)	Derivation study: Patients (N)	Derivation study: Observed strokes n (%)	Validation study: Patients (N)	Validation study: Predicted incidence (%)*	Validation study: Predicted number (n)	Validation study: 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 used these "derivation study" coefficients to predict probabilities and strokes in the validation study. The linear estimate Y_{VAL} in the validation study was calculated as:

$$Y_{VAL} = \alpha_{DER} + \beta_{INT,DER} X_{INT,VAL} + \beta_{HIGH,DER} X_{HIGH,VAL}$$

where a_{DER} (-4.29) is intercept and $ß_{INT,DER}$ (1.63) and $ß_{HIGH,DER}$ (2.23) are coefficients from the derivation study. $X_{INT,VAL}$ 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:

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

We added individual probabilities P to predict strokes in each stratum: low risk 0-3 (expected=5.7), intermediate 4-5 (25.9) and high 6-7 (15.7), rounded as 6, 26 and 16.

A meta-analysis of predicted strokes and predicted:observed ratios in the validation studies, is performed *(pooled RRs, measures of discrimination, calibration, heterogeneity with fixed and random-effects models)*. Updating the intercept of the original LR model corrects the calibration - it "adjusts" the mean predicted probability to become equal to the frequency of observed outcome. This can be achieved by fitting a LR model with the intercept as the only free parameter and the linear estimate Y_{VAL} as an offset variable *(slope is fixed at unity)*.

Results

The main result of our methodological work is the introduction of a new, simplified approach to compute predicted values and derive a "predicted:observed" ratio of the outcomes in order to assess the calibration in validation studies of CPRs. This was achieved by using only information from the derivation study (i.e., the "derivation model") (**Table 1**).

We used the ABCD^2 rule as an example and the approach was confirmed in terms of construct and congruent validities of the predicted estimates. The analysis included the derivation and validation studies of ABCD^2 rule that ranged in size from 136 to 1054 patients. Using simulated IPD sets, we fitted a CPR-based LR model with the derivation study coefficients to the data from a sample validation study. We were able to obtain the same predicted outcomes as computed by our new, simplified approach (not shown).

We obtained predicted outcomes and "predicted:observed" ratios, performed meta-analysis and assessed CPR performance in each study (not shown). Summary calibration estimates (pooled RRs, 95% CIs, etc.) were illustrated by Forest plots (see example: **Figures 2 & 3**). Although with good discrimination (c-statistics € 0.608-0.819), some studies had low calibration (slight under-prediction). The latter can be clearly observed when presented by the assumed risk of stroke at different levels of the ABCD^2 rule (low: 0-3, intermediate: 4-5 and high: 6-7 points, **Table 2**).

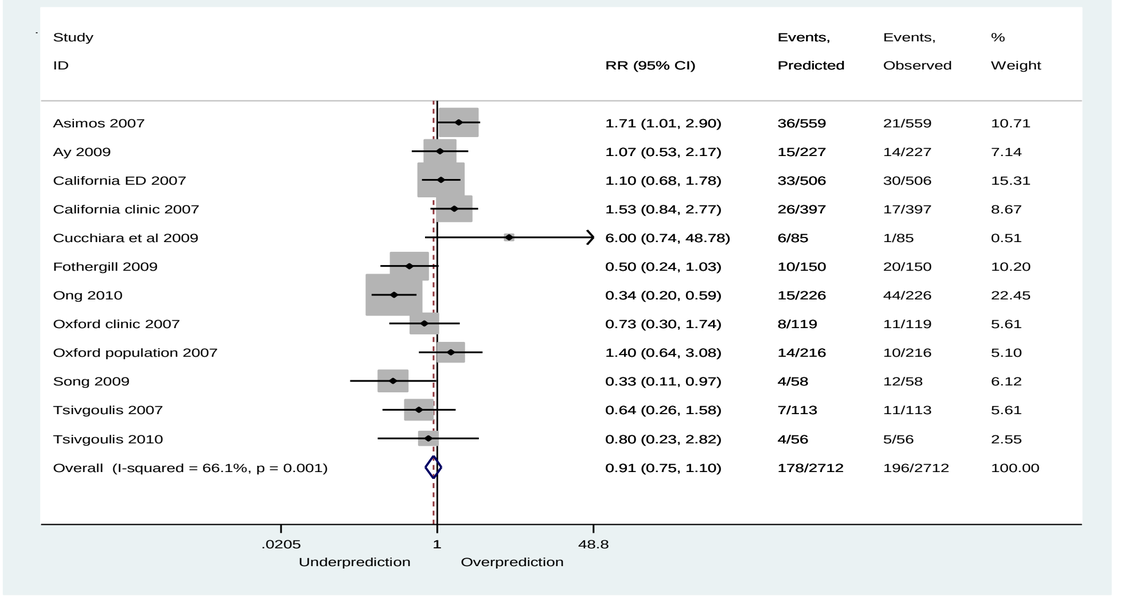

Fig.2 Forest plots of ABCD^2 rule (intermediate risk) – original CPR (fixed effects)

We identified low calibration levels (slight under-prediction with RR € 0.73-0.91), possibly due to high heterogeneity (18.8%-66.1%) which were refined by adjustment of the original model intercept to take into account the different incidence rates (*a priori* probabilities). While discrimination has not improved further, better calibration Hosmer-Lemeshow "goodness-of-fit" p-values and improved pooled estimates (RR € 0.90-1.06, Table 2), with narrower 95%CIs and zero heterogeneity, were achieved.

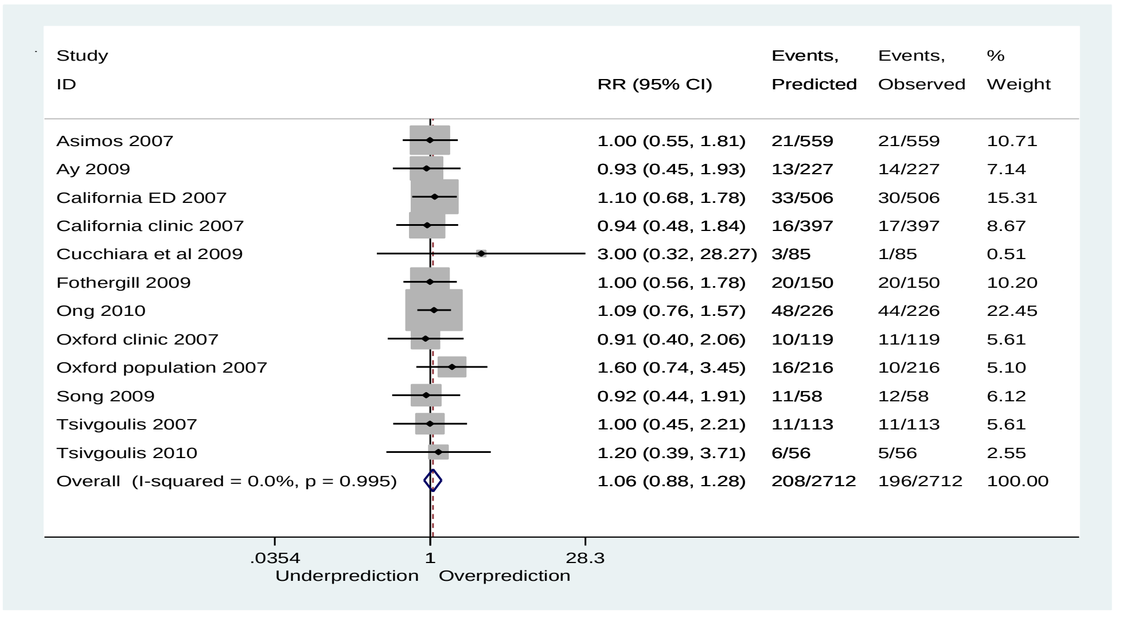

Fig.3 Forest plots of ABCD^2 rule (intermediate risk) – intercept adjustment (fixed effects)

Discussion

Our new approach is very useful in predicting outcomes by CPRs for: (i) assessment of calibration and/or subsequent inclusion of validation studies in meta-analysis; (ii) signalling mis-calibration and its improvement by updating; (iii) comparison with predicted values, computed by other models (e.g., confirmation of construct validity in ABCD^2 rule); and (iv) further testing, refinement and improvement in terms of transportability for other CPRs or in different populations.

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): I^2	No adjustment (original CPR): Fixed effects	No adjustment (original CPR): Random effects	Adjustment of intercept: I^2	Adjustment of intercept: 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.

In summary, when predicted outcomes in validation studies are neither published nor accessible or sufficient, such predicted values can be easily obtained by our simplified method. Our new approach describes and justifies how this can be achieved by everyone, when using data from a derivation study alone, without any further requirements for highly-specialised knowledge or sophisticated statistical software.

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

***Current address:** Dr Borislav Dimitrov, 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)*