The Performance of Matrix-Based Risk Models for Rapid Radiographic Progression in an Observational Cohort of Established Rheumatoid Arthritis Patients

Siri Lillegraven[1,2] MD, Femke H.M. Prince[1] MD PhD MSCE, Nancy A. Shadick[1] MD MPH, Espen A. Haavardsholm[2] MD PhD, Michelle L. Frits[1] BA, Christine K. Iannaccone[1] MPH, Tore K. Kvien[2] MD PhD, Michael E. Weinblatt[1] MD, Daniel H. Solomon[1] MD MPH

[1] Brigham and Women's Hospital, Boston, MA, USA
[2] Diakonhjemmet Hospital, Oslo, Norway

Background/Purpose: Matrix-based risk models have been proposed as a clinical tool to predict rapid radiographic progression (RRP, defined as ≥ 5 units change in Sharp score/year) in rheumatoid arthritis (RA). The three current models have been based on data from clinical trials in early RA, and each model includes 4 of the following variables: treatment, erosiveness, seropositivity, swollen joint count, CRP and smoking. Predictive probabilities for RRP are calculated, similar to what is seen in the Framingham risk score. We tested the performance of the three risk models in an observational cohort with established RA.

Method: Patients were recruited from BRASS, an observational RA cohort with treatment according to clinical practice. 478 patients had hand radiographs (scored according to the Sharp method) at baseline and 2 years and received disease modifying anti-rheumatic drugs (DMARDs). Three models were assessed: A) the ASPIRE model[1], B) the BeSt model[2] and C) a model from the second year of the SWEFOT trial[3]. We classified patients according to synthetic DMARDs vs. biologic treatment. Patients were classified as cases/non-cases for RRP and allocated to the correct matrix cell for each model, with a corresponding predicted risk of RRP. The mean predicted probability for cases and non-cases, the net reclassification improvement (NRI) with continuous outcome, integrated discrimination improvement (IDI) and area under the receiver operating curve (AUC) was calculated for each model.

Result: The median (IQR) age for the 478 patients was 59 years (50, 66), disease duration 12 years (4, 23), swollen joint count 6 (2, 13) and tender joint count 7 (1, 4). 84% were female and 86% had presence of erosions at baseline. The proportion of patients with RRP was 32/271 in the synthetic DMARD group and 21/207 in the biologic DMARD group (either as monotherapy or in combination with synthetic DMARDs). Model statistics summarized in the table indicated that Model B fit our data best, but none of the models separated cases and non-cases well in this cohort.

	Discrimination			***Classification***		
	Model A	***Model B***	***Model C***		***Model A vs. model B***	***Model B vs. Model C***
Mean predicted probability cases/non-cases	9.8 / 9.1	29.6 / 20.2	22.3 / 20.2	**IDI * ‡**	8.7	-7.3
AUC	0.631	0.716	0.617	**NRI with continuous outcome * #**	42	-32.8

* For both NRI and IDI a positive number indicates better performance of the comparator model (mentioned last in the column heading)
‡ IDI compares the difference in mean predicted probability between cases and non-cases for two models
# NRI compares the ability of two models to correctly classify cases and non-cases

Conclusion: Matrix risk models developed in randomized clinical trials had limited value in this observational cohort of RA patients with established disease. Limitations of the study include lack of feet radiographs, which might have led to fewer patients being classified as RRP, and potential confounding by indication. The value of matrix risk models for RRP might be greater in early RA, larger studies and cohorts necessary to develop and test such models, or models for specifically developed for established RA may be needed.

References:

[1] Vastesaeger et al, Rheumatology 2009
[2] Visser et al, ARD 2010
[3] Engström et al, EULAR 2011 (abstract)

Question: The models are now named A, B and C throughout the abstract, partly to avoid annoying the people who have developed the models since our conclusion is that the performance of the models in BRASS is not that good. Should we instead call it the ASPIRE, BeSt and SWEFOT models?