Developments in modern hemophilia care

Hassan, S.

Chapter 5

Performance of a clinical risk prediction model for inhibitor formation in severe hemophilia A

Shermarke Hassan, Roberta Palla, Carla Valsecchi, Isabella Garagiola, Amal El-Beshlawy, Mohsen Elalfy, Vijay Ramanan, Peyman Eshghi, Mehran Karimi, Samantha C. Gouw, Pier M. Mannucci, Frits R. Rosendaal, Flora Peyvandi

Haemophilia. 2021;27(4):e441-e449

Abstract

Background

There is a need to identify patients with hemophilia who have a very low or high risk of developing inhibitors. These patients could be candidates for personalized treatment-strategies.

Aims

The aim of this study was to externally validate a previously published prediction model for inhibitor development and to develop a new prediction model that incorporates novel predictors.

Methods

The population consisted of 251 previously untreated or minimally treated patients with severe hemophilia A enrolled in the SIPPET study. The outcome was inhibitor formation. Model discrimination was measured using the C-statistic and model calibration was assessed with a calibration plot. The new model was internally validated using bootstrap resampling.

Results

Firstly, the previously published prediction model was validated. It consisted of three variables: family history of inhibitor development, *F8* gene mutation and intensity of first treatment with factor VIII (FVIII). The C-statistic was 0.53 (95%CI: 0.46-0.60) and calibration was limited.

Furthermore, a new prediction model was developed that consisted of four predictors: *F8* gene mutation, intensity of first treatment with FVIII, the presence of factor VIII non-neutralizing antibodies before treatment initiation and lastly, FVIII product type (recombinant vs. plasma-derived). The C-statistic was 0.66 (95CI: 0.57-0.75) and calibration was moderate. Using a model cut-off point of 10%, positive- and negative predictive values were 0.22 and 0.95, respectively.

Conclusion

Performance of all prediction models was limited. However, the new model with all predictors may be useful for identifying a small number of patients with a low risk of inhibitor formation.

Background

A major treatment complication in hemophilia A is the formation of neutralizing antibodies against factor VIII (also called inhibitors) which render subsequent treatment with factor VIII (FVIII) ineffective and are associated with increased morbidity/mortality.[1] There is a need to identify patients with a very low/high risk of developing inhibitors as these patients could be candidates for personalized treatment-strategies.[2]

Two published prediction models for inhibitor formation have been suggested for clinical use.[3, 4] The second model[4] (a modified version of the earlier model[3]), was developed using data from the CANAL study and PedNet registry. The study population consisted of 825 previously untreated patients (PUPs) with severe hemophilia A, followed from 1-50 days of exposure to FVIII (EDs). The model contained three predictors: family history of inhibitors, *F8* gene mutation and intensity of the first FVIII treatment episode. The model C-statistic was 0.69 (95% CI 0.65–0.73). The calibration plot overestimated the inhibitor risk in the higher ranges of inhibitor incidences (› 0.55). This model urgently needs to be externally validated in another dataset.

New risk factors for inhibitor formation have been identified using the SIPPET study cohort.[5-7] Firstly, the use of recombinant FVIII (rFVIII) was associated with a higher inhibitor risk than plasma-derived FVIII (pdFVIII) (hazard ratio: 1.87, 95CI: 1.17-2.96).[5]

Furthermore, the presence of non-neutralizing anti-FVIII antibodies (NNAs) before FVIII exposure was associated with an increased risk of inhibitor formation in previously untreated and minimally treated patients with severe hemophilia A (HR: 1.83, CI95: 0.84-3.99).[7] Studies have also shown that NNAs are detectable in non-hemophilic subjects. (most of whom were never exposed to blood components such as fresh-frozen plasma).[8] This suggests that some autoreactivity against endogenous FVIII is relatively common.[9]

Lastly, a genetic analysis showed that inhibitor prediction based on FVIII mutation could be improved by also accounting for FVIII antigen production.[6] A new model incorporating these new data could be useful for clinical practice.

The first aim of this study was to externally validate the latest published prediction model for inhibitor development.[4] The second aim was to develop a new clinical prediction model that incorporates novel predictors.

Methods

Study design and population

Data from the SIPPET study were used.[5] The SIPPET study enrolled 251 severe (FVIII:C ‹ 1%) hemophilia A patients without previous treatment with FVIII or only minimal treatment with blood components. Patients were followed-up for 50 EDs or 3 years of observation (whichever came first). The cumulative number of EDs to FVIII was used as the timescale.

Defining outcome and predictor variables

Validation of 2015 model

The outcome, inhibitor formation, was defined as any inhibitor higher than 0.4 Bethesda Units (BU), measured using the Bethesda assay with Nijmegen modification. The 2015 prediction model consisted of three predictors; family history of inhibitors, *F8* gene mutation and intensity of the first treatment with FVIII.[4]

Family history of inhibitors was analyzed as a categorical variable (not applicable/negative, positive, unknown). Family history of inhibitors was classified as 'not applicable' when the patient had a negative family history of hemophilia.

F8 gene mutation was defined as a categorical variable (missense mutations, null mutations, other, unknown). The category 'null mutations' consisted of deletions of › 200 base pairs, nonsense mutations, intron 22 inversions and intron 1 inversions. The category 'other mutations' consisted of small deletions of ‹ 200 base pairs, insertions and splice site defects.

Intensity of first treatment was a continuous variable defined as the product of the number of consecutive EDs at first treatment (ranging from the first ED up to the 10th consecutive ED) and the mean daily dose in IU/kg of FVIII used during this period. The result was expressed as a fraction of 50 IU/kg. (As an example, an individual who was treated for 5 consecutive EDs with a mean daily dose of 75 IU/kg would have a value of 5 EDs x (75 IU/kg ÷ 50 IU/kg) = 5 x 1.5 = 7.5.)

Development of new model

To improve clinical applicability, high-titer inhibitor formation, defined as a peak inhibitor titer of at least 5 Bethesda units, was used as the outcome.

On the basis of literature and subject-matter knowledge, four predictors were considered: intensity of the first treatment with FVIII, *F8* gene mutation, NNA status before treatment initiation and treatment with pdFVIII or rFVIII.

Treatment intensity was defined as being treated for at least 2 consecutive EDs at first treatment. For *F8* gene mutation, we used the classification by Spena et al.[6] In this classification, in silico predicted null mutations were reclassified as non-null if there were detectable FVIII antigen levels. Missing values were encoded as a separate category labeled 'unknown'. NNA status before treatment initiation was analyzed as a dichotomous variable (negative or positive), according to cut-off values of the NNA assay (≥ 1.64mg/mL of specific anti-FVIII IgG[7]). Treatment type was defined as treatment with either plasma-derived FVIII (pdFVIII) or recombinant-derived FVIII (rFVIII).[5]

Statistical analysis

Validation of 2015 model
The predicted risk of inhibitor formation was calculated for each individual in the SIPPET study, using the formula described in the original paper.[4]

Development of new model
Three different models were fit using logistic regression. The first two models were developed to be used before any FVIII exposure; the first model contained only *F8* gene mutation as a predictor, the second model also included NNA status.

The third model was developed to predict inhibitor risk just after the first treatment episode and consisted of *F8* gene mutation, NNA status, treatment intensity and treatment type. Variable selection was based on the strength of the predictors as well as subject-matter knowledge. Family history was difficult to ascertain correctly and was therefore not included as a predictor. For treatment intensity, we chose 2 ED's instead of 5 ED's because the aim was to develop a model that could be implemented almost immediately after the start of treatment. Consequently, patients with an inhibitor event in the first 2 EDs were excluded from the analysis of the full model.

Internal validation of the new model using a bootstrapping procedure
To correct for overfitting, a uniform shrinkage factor was estimated using the bootstrap resampling method.[10] Next, model coefficients were multiplied by the shrinkage factor and the model constant was re-estimated with the shrunken coefficients.

Evaluating model performance
Discrimination is the level to which a model can distinguish between patients developing and not developing the outcome. Discriminative power of each model was assessed with the C-statistic. The C-statistic can be calculated by taking all possible pairs in which one subject developed the outcome and the other did not. Pairs in which the patient with the outcome also had a higher predicted risk of the outcome are called concordant pairs. The higher the proportion of concordant pairs among all pairs, the higher the C-statistic. The C-statistic can range from 0.5 (no discrimination) to 1 (perfect discrimination).

Calibration refers to the degree to which predicted and observed outcomes are similar. Calibration of each model was reported visually in a calibration plot, with expected outcome probabilities plotted against observed outcome frequencies, for each quintile of predicted risk. Furthermore, a LOWESS (Locally Weighted Scatterplot Smoothing) line was estimated to examine calibration across the whole range.

Sensitivity, specificity, positive and negative predictive values were calculated for different cut-off values of the new model.

Handling missing values
Missing values for any of the predictors or outcome variable in the SIPPET dataset were imputed using multivariate imputation by chained equations. Model coefficients of each imputed dataset, their C-statistics and corresponding standard errors were pooled using Rubin's rules to obtain the final estimates.[11] Internal validation using bootstrap resampling was performed within each imputed dataset. The results (i.e. the calibration intercept, slope, shrinkage factor and optimism corrected C-statistic) were also pooled using Rubin's rules. The calibration plot was constructed by combining the imputed datasets and fitting the shrunken model to this pooled dataset.

Statistical packages
The data was prepared for analysis using IBM SPSS statistics version 25. Analysis were performed using R version 3.1.0.

Results

General information

Characteristics of the validation cohort are shown in Table 1. Overall, 76/251 patients developed an inhibitor, 50/76 inhibitor patients had a high-titer inhibitor. Further-

Table 1: Patient characteristics.

Predictors of 2015 model	**All patients (N = 251)**	**inhibitor-negative (N = 175)**	**inhibitor-positive (N = 76)**
F8 gene mutation type (Hashemi 2015)			
Missense	22 (8.8%)	18 (10.3%)	4 (5.3%)
Null	166 (66.1%)	111 (63.4%)	55 (72.4%)
Other	46 (18.3)	33 (18.9%)	13 (17.1%)
Unknown	17 (6.8%)	13 (7.4%)	4 (5.3%)
Family history			
Negative/not applicable	205 (81.7%)	140 (80.0%)	65 (85.5%)
Positive	24 (9.6%)	19 (10.9%)	5 (6.6%)
Unknown	22 (8.8%)	16 (9.1%)	6 (7.9%)
Intensive treatment†			
Mean (SD)	0.82 (SD: 5.9)	0.96 (SD: 6.6)	0.48 (SD: 0.9)
Predictors of new model	**All patients (N = 251)**	**High-titer inhibitor-negative (N = 201)**	**High-titer inhibitor-positive (N = 50)**
Study treatment			
pd-FVIII	125 (49.8%)	105 (52.2%)	20 (40.0%)
rec-FVIII	126 (50.2%)	96 (47.8%)	30 (60.0%)
Pre-treatment NNA status‡			
Negative	219 (92.4%)	178 (94.2%)	41 (85.4%)
Positive	18 (7.6%)	11 (5.8%)	7 (14.6%)
At least 2 consecutive EDs at first treatment			
No	210 (83.7%)	175 (87.1%)	35 (70.0%)
Yes	41 (16.3%)	26 (12.9%)	15 (30.0%)
F8 gene mutation type (Spena 2018)			
Missense	42 (16.7%)	39 (19.4%)	3 (6.0%)
Null	189 (75.3%)	144 (71.6%)	45 (90.0%)
Unknown	20 (8.0%)	18 (9.0%)	2 (4.0%)

† Mean intensive treatment was a continuous variable defined as the product of the number of consecutive EDs at first treatment (ranging from the first ED up to the 10th consecutive ED) and the mean daily dose in IU/kg of FVIII used during this period. The result was expressed as a fraction of 50 IU/kg.

‡ Pre-treatment NNA status: 14 missing values overall (5.6%).

more, 75% of patients had a *F8* null mutation, 9.6% had a positive family history, 7.6% were NNA-positive and 16.3% were treated for at least 2 consecutive days at first treatment. NNA status was unknown in 14 patients and *F8* gene mutation was unknown in 20 patients.

External validation of the 2015 prediction model

Baseline characteristics of the 825 patients in the development cohort compared to the 251 patients in the validation cohort are shown in Table 2. In the development cohort, 228/825 (27.6%) of patients developed an inhibitor. The C-statistic in this cohort was 0.69 (95% CI 0.65 - 0.73). In our cohort, we found a C-statistic of 0.53 (0.46 - 0.60). Figure 1A shows the calibration plot of the risk score, as applied to the validation cohort. Overall calibration was limited, as the model highly overpredicts in the higher risk ranges.

Development of new prediction models; association between predictors and inhibitor formation

Table 3 shows the unadjusted and adjusted associations (of the full model) between each predictor and high-titer inhibitor formation. In the multivariable model, the strongest predictors were *F8* gene mutation type (odds ratio: 3.94) and NNA status (odds ratio: 3.38).

Development of prediction models before exposure to FVIII products

The C-statistic of the model with only *F8* gene mutation was 0.59 (95CI: 0.54 - 0.64). The C-statistic of the model with only F8 gene mutation and NNA status at treatment initiation was 0.61 (95CI: 0.52 - 0.71).

Development of full prediction model

The C-statistic of the full model was 0.66 (95CI: 0.57-0.75). The shrunken regression coefficients of the final logistic model are shown in Table 4. Figure 1B shows the optimism-corrected calibration plot of the new model. Overall calibration was low to moderate, as the model underpredicted in the higher risk ranges. The predicted inhibitor risk for an individual in the SIPPET cohort ranged from 6% to 62%. Table 5 shows the incidence of inhibitor development across different categories of predicted risk. Table 6 shows the sensitivity, specificity, positive and negative predictive values of the model for different model cut-off points. The positive predictive value was very low when using the low- and medium cut-off values and slightly higher but still low for the high cut-off value. Conversely, the negative predictive value was high for all three model cut-off points.

Discussion

Main findings

A published inhibitor prediction model showed limited performance in our cohort. Furthermore, the performance of a new model that included novel predictors was also limited.

External validation of 2015 model

The limited performance of the old model may partly be explained by differences in patient characteristics between development- and validation cohorts. Curiously, a positive family history of inhibitors was more common among non-inhibitor patients in our cohort (which reduced model performance). Family history was often difficult to ascertain, which could explain the aforementioned results. However, we were able to include the *F8* gene mutation in our model. (which explains a large of part of familial inhibitor risk) Similarly, mean treatment intensity (which is consistently reported to be associated with inhibitor development) was also higher in non-inhibitor patients.

Compared to the observational development cohort, some patients may have been underrepresented as the SIPPET trial was interventional. For example, obtaining informed consent for participation before any FVIII exposure might have been more difficult for patients with a negative family history of hemophilia presenting with acute severe trauma at the emergency department. Similarly, neonates with an intracranial bleed would have been more difficult to enroll if family history of hemophilia was unknown. Unfortunately, patients with a negative family history of hemophilia and patients with a positive family history of hemophilia/negative family history of inhibitors were combined into one category (family history 'Negative/not applicable') in the 2015 model. (Table 2) It was therefore not possible to directly compare the proportion of patients with a negative family history of hemophilia in the SIPPET cohort vs. the development cohort.

Furthermore, the 2015 model used a stepwise predictor selection procedure, which is known to produce overfitted models.[12] However, the study partially corrected for this by shrinking the final model coefficients through bootstrapping.

Lastly, the poor calibration in the higher risk range (over 50%) was mostly due to the very low number of patients in this area.

Table 2. Comparison of participant characteristics in the development cohort and the validation cohort.

Characteristic	2015 model development cohort (n = 825)†	SIPPET cohort (n = 251)‡
Age in months		
Median (IQR)	10 (6-14)	15 (9-29)
Inhibitor status		
Negative (%)	587 (72.4)	175 (69.7)
Positive (%)	228 (27.6)	76 (30.3)
F8 gene mutation type		
% Missense	12	9
% Null	59	66
% Other	17	18
% Unknown	13	7
Family history		
% Negative/not applicable	83	82
% Positive	9	10
% Unknown	8	9
Treatment intensity		
Mean (SD)§	NR	0.82 (5.9)

† CANAL study/PedNet registry.
‡ SIPPET study. NR: not reported in the original article.
§ Intensity of first treatment was a continuous variable defined as the product of the number of consecutive EDs at first treatment (ranging from the first ED up to the 10th consecutive ED) and the mean daily dose in IU/kg of FVIII used during this period. The result was expressed as a fraction of 50 IU/kg.

Overall, whether the 2015 model underperforms in general, or is merely poorly generalizable to the type of patients enrolled in the SIPPET cohort remains an open question.

Development of pre-FVIII exposure prediction models

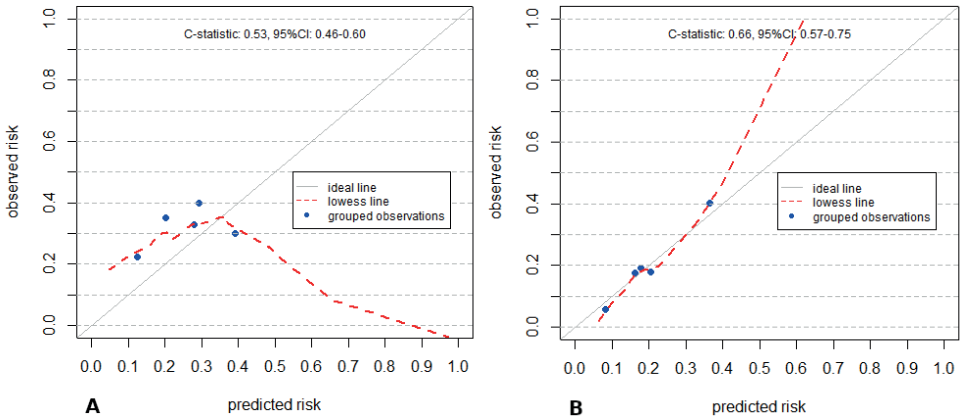
The two simple prediction models were chosen to contain only predictors measurable before FVIII exposure. Both models performed poorly. To construct an accurate pre-FVIII exposure prediction model, additional predictors that can be measured before treatment are necessary. (e.g. certain gene variants)

Development of full prediction model

The full model performed similarly to the 2015 model. The model included treatment intensity, which is consistently associated with inhibitor development.[13] However,

Figure 1. Calibration plot of 2015 model & new model.

Legend: The figure shows the calibration plot of the 2015 model (A), and of the new model (B). On the X-axis, the predicted probability of inhibitor formation according to the model is plotted against the observed risk on the Y-axis. (0 = no inhibitor, 1 = developed an inhibitor) The blue dots represent the proportion of patients experiencing an event, stratified by quintiles of increasing predicted risk. Quintiles with a higher predicted inhibitor risk should have a higher proportion of patients who develop the outcome. (i.e. a higher observed risk) Alternatively, a LOWESS (Locally Weighted Scatterplot Smoothing) line was estimated to examine calibration across the whole range. (shown here as a red dotted line) The grey line represents perfect prediction, meaning that the predicted risk is exactly the same as the observed risk across the whole range. Ideally, both the quintiles and the LOWESS line should lie exactly on top of the grey line.



our definition of treatment intensity (two consecutive EDs) has some limitations, as the second dose might have been a prophylactic dose. Also, instead of receiving one dose, some patients may have gotten two half doses over two days.

The association between FVIII product type and inhibitor development was not statistically significant due to a lack of power caused by not having enough high-titer inhibitor events. This predictor was still included based on previous literature and subject matter-knowledge, as models with predictors selected solely using significance levels perform poorly when externally validated.[14]

However, model performance was still very limited. The maximum predicted inhibitor risk was 62% and, except for one outlier, no patients had a predicted inhibitor risk

over 40%. Therefore, prediction in the higher risk ranges was not possible. However, calibration in the lower risk ranges was acceptable, and the negative predictive value of the model using the lowest model cut-off of 10% was 95%. (i.e. of the 41 patients with a predicted risk below 10% only two developed an inhibitor) Therefore, we can conclude that the model is useful for identifying low-risk patients. However, only 16% of patients fell into this low-risk category. These were all patients with a *F8* non-null mutation or an unknown *F8* mutation, no detectable NNAs before treatment initiation, and who were not intensively treated at first treatment.

The model did not include genetic risk factors other than *F8* gene mutation, and this could have impacted performance. Furthermore, we found no association between family history and inhibitor development in the SIPPET cohort. This result was probably biased, as family history was difficult to ascertain correctly in our cohort (which mostly consisted of patient from the developing world). Therefore, we decided to exclude this predictor from the model.

NNAs are not routinely measured in clinical practice which limits practical implementation of this model.

Information on ethnicity was not included in the model, as most research on ethnicity and inhibitor formation has focused on African-American/Latino populations, and these ethnicities are very uncommon within the SIPPET cohort. Furthermore, many patients within the SIPPET cohort self-identified as "white" (e.g. patients from Egypt, Iran, Saudi-Arabia etc.), while the original studies on ethnicity mostly enrolled "white" patients from a predominantly European background (i.e. from Europe or North-America), which complicates between-study comparisons.

Lastly, the performance of the new model after external validation in a different population remains unknown.

Implications for clinical practice

The overall performance of the original prediction model, as well as the newly developed models was limited. However, the newly developed full model performed relatively well when identifying patients with a low risk of inhibitor formation.

Currently, pre-authorization trials evaluating FVIII therapeutics often enroll previously treated patients (PTPs) who have been exposed to FVIII for more than 50 EDs. Enrolling previously untreated patients (PUPs) with a low predicted risk of inhibitor formation

Table 3. Unadjusted and adjusted association between each predictor and outcome.

Characteristic	All patients (N = 251)	High-titer inhibitor-positive (N = 50)	Univariate Odds Ratio (95%CI)	Multivariable Odds Ratio (95%CI)‡
Study treatment				
pdFVIII	125 (49.8%)	20 (40.0%)	Ref	Ref
rFVIII	126 (50.2%)	30 (60.0%)	1.64 (0.88-3.12)	1.46 (0.75-2.84)
Pre-treatment NNA status†				
Negative	219 (92.4%)	41 (85.4%)	Ref	Ref
Positive	18 (7.6%)	7 (14.6%)	2.76 (0.97-7.46)	3.38 (1.17-9.80)
At least 2 consecutive EDs at first treatment§				
No	209 (83.6%)	34 (69.4%)	Ref	Ref
Yes	41 (16.4%)	15 (30.6%)	2.96 (1.41-6.15)	3.20 (1.47-6.97)
F8 gene mutation type (Spena 2018)				
Missense	42 (16.7%)	3 (6.0%)	Ref	Ref
Null	189 (75.3%)	45 (90.0%)	4.06 (1.39-17.36)	3.94 (1.13-13.73)
Unknown	20 (8.0%)	2 (4.0%)	1.44 (0.18-9.5)	1.38 (0.20-9.37)

† Pre-treatment NNA status: 14 missing values overall (5.6%) ‡: For the multivariable model, missing values were imputed using multiple imputation, one patient with an inhibitor event in the first 2 EDs was excluded from the analysis, so the total sample size for this analysis was 250. §: 1 missing value, due to one patient being excluded from the analysis due to experiencing an inhibitor event in the first 2 EDs of treatment.

might be considered as an alternative, as the study population is a better match for the target population that will actually use the treatment after market approval (not just PTPs but also PUPs). However, due to the difficulty of enrolling such a rare group of patients (only 16% of PUPs), this approach is not practically feasible. For non-factor replacement therapy, this score would not be useful, as these drugs don't elicit anti-FVIII antibodies. The most important use-case for this prediction model would be after market approval. Novel therapeutics are relatively expensive compared to FVIII, and many patients will continue to be treated with FVIII. A score such as this could be used to select low-risk patients who can be safely treated with regular FVIII concentrates (which are relatively cheap).

Table 4. Final logistic regression model.

Regression coefficients	
Intercept	-2.71
Treatment with rec-FVIII (TRT)	0.29
Positive for NNAs (NNA)	0.95
At least 2 consecutive EDs at treatment initiation (ED)	0.90
F8 gene null mutation (F8-null)	1.07
F8 gene mutation unknown (F8-unknown)	0.25

To calculate the individual risk of inhibitor formation, first calculate the linear predictor: (-2.71 + TRT * 0.29 + NNA * 0.95 + ED * 0.90 + F8-null * 1.07 + F8-unknown * 0.25). The formula is then as follows: 1 / (1+ exp(-(linear predictor))). As an example, the risk of inhibitor formation within 50 EDs for a patient treated with plasma-derived FVIII, who was positive at baseline for NNAs, who was treated for at least 2 consecutive EDs at treatment initiation, and whose *F8* mutation is unknown is 1 / (1+ exp(-(-2.71 + 0 * 0.29 + 1 * 0.95 + 1 * 0.90 + 0 * 1.07 + 1 * 0.25))) = 35%.

Table 5. Incidence of inhibitor development across different risk categories.

Predicted risk	**No of inhibitor-negative patients***	**Inhibitor events**	**Observed risk**
‹ 10%	39	2	4.9%
10-25%	134	29	17.8%
25-40%	24	13	35.1%
≥ 40%	4	5	55.6%

* For the construction of the new model, patients with an inhibitor event in the first 2 EDs were excluded. (also mentioned in the Methods section) This was the case for one out of 251 patients, the total number of patients used to construct the new model therefore equals 250.

Table 6. Sensitivity, specificity, positive and negative predictive values of the model for different model cut-off values.

Categories of predicted risk according to model	**Sensitivity**	**Specificity**	**Positive predictive value**	**Negative predictive value**
Low cut-off (10%)	0.96	0.19	0.22	0.95
Medium cut-off (25%)	0.37	0.86	0.39	0.85
High cut-off (›40%)	0.10	0.98	0.56	0.82

These results could be the first step in developing a model for this aim. However, these tools should not be used in clinical practice to select high-risk patients, as all models perform very poorly in this regard. For this reason, the new prediction model was not converted into a tool that could be used by clinicians. (for example, a nomogram or a score chart)

Implications for future research

All prediction models incorporated the most important pre-treatment risk factors. But even so, performance of these models was still unsatisfactory. However, these models did not incorporate time-varying predictors (e.g. the cumulative number of EDs, FVIII exposure frequency, on-demand vs. prophylactic treatment, exposure to FVIII during trauma or during surgery etc.). For example, much information could be gained by measuring the antibody response over time[15], as was done in a recent study by Reipert et al.[16] Interestingly, this study found that during treatment, the appearance of IgG1 antibodies, followed by IgG3 antibodies, was a strong biomarker of future inhibitor development. A different approach would be to incorporate genomic information at baseline, such as HLA class II haplotypes[17, 18] and/or gene variants of other genes previously associated with inhibitor formation (e.g. IL-10 and CTLA-4)[19].

Conclusion

Performance of old and new prediction models for inhibitor formation after external validation is limited. However, the new model with all predictors may be useful for identifying patients with a low risk of inhibitor formation. Further research is needed to obtain more precise prediction models for clinical use.

Acknowledgements

SH, RB, SG, PMM, CV, IG, AE, ME, VR, PE, MK, FRR and FP jointly designed the study, performed the research and contributed to writing the paper. SH analysed the data. We thank the SIPPET investigators for their help with patient recruitment and data collection (the full list of investigators is provided in the appendix). This work was partly supported by a research grant from the Leiden University Fund / Van Trigt Fund and by Ricerca Corrente of the Italian Ministry of Health.

References

1. Peyvandi F, Garagiola I, Young G. The past and future of haemophilia: diagnosis, treatments, and its complications. *Lancet (London, England)*. Feb 17 2016;doi:10.1016/s0140-6736(15)01123-x
2. Mannucci PM, Mancuso ME, Santagostino E, Franchini M. Innovative Pharmacological Therapies for the Hemophilias Not Based on Deficient Factor Replacement. *Seminars in thrombosis and hemostasis*. Jul 2016;42(5):526-32. doi:10.1055/s-0036-1571310
3. ter Avest PC, Fischer K, Mancuso ME, et al. Risk stratification for inhibitor development at first treatment for severe hemophilia A: a tool for clinical practice. *Journal of thrombosis and haemostasis : JTH*. Dec 2008;6(12):2048-54. doi:10.1111/j.1538-7836.2008.03187.x
4. Hashemi SM, Fischer K, Moons KG, van den Berg HM. Improved prediction of inhibitor development in previously untreated patients with severe haemophilia A. *Haemophilia : the official journal of the World Federation of Hemophilia*. Mar 2015;21(2):227-33. doi:10.1111/hae.12566
5. Peyvandi F, Mannucci PM, Garagiola I, et al. A Randomized Trial of Factor VIII and Neutralizing Antibodies in Hemophilia A. *N Engl J Med*. May 26 2016;374(21):2054-64. doi:10.1056/NEJMoa1516437
6. Spena S, Garagiola I, Cannavo A, et al. Prediction of factor VIII inhibitor development in the SIPPET cohort by mutational analysis and factor VIII antigen measurement. *J Thromb Haemost*. Apr 2018;16(4):778-790. doi:10.1111/jth.13961
7. Cannavo A, Valsecchi C, Garagiola I, et al. Nonneutralizing antibodies against factor VIII and risk of inhibitor development in severe hemophilia A. *Blood*. Mar 09 2017;129(10):1245-1250. doi:10.1182/blood-2016-06-720086
8. Abdi A, Bordbar MR, Hassan S, et al. Prevalence and Incidence of Non-neutralizing Antibodies in Congenital Hemophilia A- A Systematic Review and Meta-Analysis. *Front Immunol*. 2020;11:563. doi:10.3389/fimmu.2020.00563
9. Lacroix-Desmazes S, Voorberg J, Lillicrap D, Scott DW, Pratt KP. Tolerating Factor VIII: Recent Progress. *Front Immunol*. 2019;10:2991. doi:10.3389/fimmu.2019.02991
10. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. Aug 2001;54(8):774-81.
11. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Med Res Methodol*. Jul 28 2009;9:57. doi:10.1186/1471-2288-9-57

12. Steyerberg EW, Uno H, Ioannidis JPA, van Calster B. Poor performance of clinical prediction models: the harm of commonly applied methods. *J Clin Epidemiol*. Jun 2018;98:133-143. doi:10.1016/j.jclinepi.2017.11.013
13. Garagiola I, Palla R, Peyvandi F. Risk factors for inhibitor development in severe hemophilia a. *Thromb Res*. Aug 2018;168:20-27. doi:10.1016/j.thromres.2018.05.027
14. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. Aug 1 2014;35(29):1925-31. doi:10.1093/eurheartj/ehu207
15. Hofbauer CJ, Whelan SF, Hirschler M, et al. Affinity of FVIII-specific antibodies reveals major differences between neutralizing and nonneutralizing antibodies in humans. *Blood*. Feb 12 2015;125(7):1180-8. doi:10.1182/blood-2014-09-598268
16. Reipert BM, Gangadharan B, Hofbauer CJ, et al. The prospective Hemophilia Inhibitor PUP Study reveals distinct antibody signatures prior to FVIII inhibitor development. *Blood Adv*. Nov 24 2020;4(22):5785-5796. doi:10.1182/bloodadvances.2020002731
17. Shepherd AJ, Skelton S, Sansom CE, Gomez K, Moss DS, Hart DP. A large-scale computational study of inhibitor risk in non-severe haemophilia A. *British journal of haematology*. Feb 2015;168(3):413-20. doi:10.1111/bjh.13131
18. Pandey GS, Yanover C, Miller-Jenkins LM, et al. Endogenous factor VIII synthesis from the intron 22-inverted F8 locus may modulate the immunogenicity of replacement therapy for hemophilia A. *Nature medicine*. Oct 2013;19(10):1318-24. doi:10.1038/nm.3270
19. Bachelet D, Albert T, Mbogning C, et al. Risk stratification integrating genetic data for factor VIII inhibitor development in patients with severe hemophilia A. *PloS one*. 2019;14(6):e0218258. doi:10.1371/journal.pone.0218258