ARTICLE

Bayesian Forecasting Utilizing Bleeding Information to Support Dose Individualization of Factor VIII

João A. Abrantes[1], Alexander Solms[2], Dirk Garmann[3], Elisabet I. Nielsen[1], Siv Jönsson[1] and Mats O. Karlsson[1],*

Bayesian forecasting for dose individualization of prophylactic factor VIII replacement therapy using pharmacokinetic samples is challenged by large interindividual variability in the bleeding risk. A pharmacokinetic-repeated time-to-event model-based forecasting approach was developed to contrast the ability to predict the future occurrence of bleeds based on individual (i) pharmacokinetic, (ii) bleeding, and (iii) pharmacokinetic, bleeding and covariate information using observed data from the Long-Term Efficacy Open-Label Program in Severe Hemophilia A Disease (LEOPOLD) clinical trials (172 severe hemophilia A patients taking prophylactic treatment). The predictive performance assessed by the area under receiver operating characteristic (ROC) curves was 0.67 (95% confidence interval (CI), 0.65–0.69), 0.78 (95% CI, 0.76–0.80), and 0.79 (95% CI, 0.77–0.81) for patients ≥ 12 years when using pharmacokinetics, bleeds, and all data, respectively, suggesting that individual bleed information adds value to the optimization of prophylactic dosing regimens in severe hemophilia A. Further steps to optimize the proposed tool for factor VIII dose adaptation in the clinic are required.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
☑ In recent years, model-based therapeutic drug monitoring using individual pharmacokinetic samples for dose individualization of factor VIII replacement products in prophylactic treatment has received ample attention. However, the large interindividual variability in bleeding risk partly limits the value of the approach.

WHAT QUESTION DID THIS STUDY ADDRESS?
☑ An integrated pharmacokinetic-repeated time-to-event model was used in Bayesian forecasting of bleeds to contrast different sources of patient information (factor VIII activity observations, bleeds and factor VIII activity, and bleeds and covariates) in their ability to predict the occurrence of bleeds.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?
☑ Individual bleed information may add value to the optimization of prophylactic dosing regimens in severe hemophilia A.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?
☑ The model-based approach suggested in this work considers bleeds, the efficacy end point of interest, and can be further developed to individualize dosing regimens.

Hemophilia A is a congenital X-linked bleeding disorder caused by a reduced or defective coagulation factor VIII (FVIII), and the severity is inversely correlated with the amount of functional FVIII an individual is able to produce. Patients with severe hemophilia (endogenous plasma FVIII activity < 1 IU/dL) experience spontaneous bleeds into the joints or muscles, starting typically during the first year of age.[1] Long-term prophylaxis with replacement FVIII products has been shown to be more effective than on-demand (episodic) treatment to prevent bleeding and joint damage. However, this strategy is costly and relies on the intravenous administration of FVIII products two to three times weekly to maintain a trough FVIII activity above 1 IU/dL. This threshold is supported by the finding that a greater time spent below this level leads to the occurrence of more spontaneous bleeds

and that patients with moderate hemophilia (1–5 IU/dL) bleed less frequently.[2,3]

In recent years, model-based therapeutic drug monitoring based on individual plasma FVIII activity observations has been encouraged as a way to optimize dosing regimens in prophylaxis (pharmacokinetic (PK)-based/driven/guided/tailored prophylaxis).[4,5] Briefly, this method uses Bayesian estimation to produce individual empirical Bayes estimates (EBEs) of PK parameters based on patient information (dose; covariate information, e.g., weight; and individual plasma FVIII activity measurements) and a prior population PK model. The population PK of FVIII activity has been extensively studied using a two-compartment model.[6–8] Concerning FVIII activity measurement, two or three blood samples are recommended to be collected at least 12 hours

[1]Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden; [2]Bayer, Berlin, Germany; [3]Bayer, Wuppertal, Germany. *Correspondence: Mats O. Karlsson (mats.karlsson@farmbio.uu.se)

apart in the period 4–48 hours postdose.[5,9] Among the recent efforts toward PK-based prophylaxis is the creation of tools for Bayesian forecasting of PK parameters for dose individualization specific to hemophilia products, for instance, the Web-Accessible Population Pharmacokinetic Service—Hemophilia (WAPPS-Hemo, McMaster University, Hamilton, Ontario, Canada; www.wapps-hemo.org) or the medical device myPKFit® (Shire Pharmaceutical Holdings Ireland Limited, Dublin, Ireland; www.mypkfit.com).[10,11]

Pharmacokinetic-based prophylaxis has been reported to be at least as effective and safe as standard prophylaxis (based only on body weight), with a similar or lower FVIII consumption and fewer infusions (both approaches involve rounding up the dose to the full vial).[12–17] However, maintaining plasma FVIII activity values above 1 IU/dL is not always sufficient to prevent bleed occurrence in all patients, particularly in those more vulnerable to bleeding or with high physical activity.[12,18–21] Other measures of exposure have been associated with bleeding, for instance, peak levels, area under the FVIII activity-time curve, or time above 20 IU/dL, acknowledging the impact of variability on the individual PK and bleeding risk.[22] Recently, eight plasma FVIII activity targets for patients with certain individual clinical characteristics, lifestyles, or those undergoing surgery were suggested based on expert consensus; however, they are still lacking prospective assessment.[23] Apart from variability in the PK of FVIII activity, procoagulant or anticoagulant factors, physical activity levels, and comorbidities may be some of the factors behind the high variability observed in bleeding. However, thus far model-based individual treatment optimization of FVIII products in prophylaxis relies solely on individual PK and body size data.

Recently, the relationship between the FVIII activity-time profile, covariates, and bleeding patterns in severe hemophilia A patients on prophylactic treatment with recombinant FVIII was characterized using an integrated PK-repeated time-to-event (PK-RTTE) model.[24] This model consisted of a population PK model describing the longitudinal plasma FVIII activity, a repeated time-to-event model describing the time to the occurrence of bleeds, and a covariate model describing the association of covariates with PK and bleeding risk. Covariates were found to only explain a minor portion of the markedly large unexplained interindividual variability in the bleeding hazard.

The aims of this study were to employ the previously developed integrated PK-RTTE model in Bayesian forecasting of bleeding events and to contrast different sources of patient information in their ability to predict the future occurrence of bleeds in severe hemophilia A patients receiving prophylactic FVIII replacement therapy.

METHODS

This study was based on observed data collected during clinical trials. Initially, EBEs were obtained by Bayesian forecasting, employing a PK-RTTE model and alternative information scenarios. Second, the EBEs were used to calculate the probability for each patient to have a bleed in the upcoming 24-hour period. Third, the individual bleeding probability was compared with the actual occurrence of a bleed in the upcoming 24-hour period. Finally, the predictive performance given by the alternative information scenarios was compared with probability threshold-dependent and threshold-independent analyses with graphical illustrations.

Data

Dose information, plasma FVIII activity, time of bleeds, and covariate data collected during the Long-Term Efficacy Open-Label Program in Severe Hemophilia A Disease (LEOPOLD) program were used for this evaluation. This three-part comprehensive clinical trial program aimed at assessing the PK, efficacy, and safety of a full-length recombinant human FVIII, BAY 81-8973 (octocog alfa, Kovaltry®; Bayer, Berkeley, CA, USA) in patients with severe hemophilia A.[25–27] The trials LEOPOLD I and LEOPOLD II included previously treated patients aged 12–64 years, and LEOPOLD kids included patients aged < 12 years. The individual treatment schedule was stablished at the discretion of the investigator within the ranges 20–50 IU/kg 2–3×/week (LEOPOLD I), 20–40 IU/kg 2–3×/week (LEOPOLD II), and 25–50 IU/kg at least 2×/week (LEOPOLD kids). In the LEOPOLD I and II trials, the individual treatment schedule established at study start was maintained throughout the study. In the LEOPOLD kids trial, the dosing regimen could be adapted to the individual needs during the study. In our analysis, the data were divided into the following two subgroups: patients aged < 12 years, including data from patients enrolled in LEOPOLD kids; and patients aged ≥ 12 years, including data from patients enrolled in LEOPOLD I and LEOPOLD II.

A total of 51 patients were < 12 years and were observed for a median period of 6.1 months (range 3.8–7.2). Of these, 28 patients (55%) had at least one bleed, and in total 101 bleeds were reported during the study (median 1 bleed/patient, range 0–9). A total of 121 patients were ≥ 12 years and were observed for a median period of 12.1 months (range 3.08–13.1). Of these, 88 patients (73%) had at least one bleed, and a total of 532 bleeds were reported (median 2 bleeds/patient, range 0–33). When assessing the predictive performance, only 530 bleeds were considered as two bleeds occurred during day 1, and no forecast was available for this day. On day 1, 41 patients < 12 years (80% of total) and 107 patients ≥ 12 years (88% of total) had at least one FVIII observation (**Table 1**). In all studies, a FVIII activity measurement postdose (15–30 minutes) was available at the start of the observation period, and peak and trough samples were collected repeatedly during the course of the study to determine FVIII recovery. In addition, dense PK sampling was measured during the study in a subset of patients.

Furthermore, information concerning 11 covariates was available, including patient and study characteristics at the study start: age; body weight; body mass index; lean body weight (LBW); race; von Willebrand factor; number of bleeds in the 12 months prestudy period; previous therapy history (on demand or prophylaxis); number of target joints for bleeds; ratio of the number of bleeds in the 12-month prestudy period to the number of target joints for

Table 1 Number of patients, PK observations, and reported bleeds over time, stratified by age group

Age group	Month	No. patients in the study[a]	No. PK observations	No. bleeds
< 12 years	1	51	104	18
	2	51	98	15
	3	51	64	27
	4	50	4	12
	5	49	0	13
	6	46	20	13
	7	2	72	3
	8	0	2	0
	Total	51	364	101
≥ 12 years	1	121	131	57
	2	121	0	57
	3	121	18	53
	4	120	14	45
	5	119	0	58
	6	119	152	51
	7	119	355	22
	8	119	0	44
	9	119	0	26
	10	119	0	32
	11	118	0	34
	12	97	18	44
	13	2	110	9
	14	0	0	0
	Total	121	798	532

PK, pharmacokinetics.
[a]Number of patients still in the study at the end of the month.

bleeds at study start; and study. More information can be found in the PK-RTTE model development study and in the **Supplemental Text**.[24]

Model

The Bayesian prior model was the integrated population PK-RTTE model previously developed based on the LEOPOLD data.[24] The PK component consisted of a two-compartment model parameterized in terms of clearances (clearance (CL), inter-compartmental clearance (Q)) and volumes of distribution (central volume of distribution (V1), peripheral volume of distribution (V2)). Central CL and V1 were allometrically scaled using LBW. Interindividual variability was included on CL (30.2 percent coefficient of variation (%CV)), V1 (15.1%CV), and on residual unexplained variability (63.1% CV), with individual random effects η_i^{CL}, η_i^{V1}, and η_i^{ε}, respectively. The repeated time to the occurrence of bleeds was described in a parametric survival model using a Gompertz hazard function. The individual time-varying bleeding hazard ($h_i(t)$) was given by:

$$h_i(t) = 2.96 \cdot e^{-0.566 \cdot (t-1)} \cdot \left(1 - \frac{FVIII_i(t)}{FVIII_i(t) + 10.2}\right) \cdot e^{\eta_i^{h_i(t)}} \quad (1)$$

where 2.96 year^{-1} and −0.566 year^{-1} are the scale and the shape factors of the Gompertz distribution, respectively; t is the time from the start of the study; $FVIII_i(t)$ is the model-predicted individual time-varying FVIII activity; 10.2 IU/

dL is the FVIII activity level resulting in half-maximum inhibition of the hazard; and $\eta_i^{h_i(t)}$ is the individual random effect accounting for the individual deviation from the typical bleeding hazard (136% CV). Covariates were included using the full random effects modeling approach[28,29] in which covariate values were included as observations, their distributions modeled as random effects, and the interaction with the model parameters carried out through the covariance elements. For further information, see the **Supplemental Text**.

Bayesian forecasting

Empirical Bayes estimates of PK and bleeding hazard parameters were obtained through the minimization of a maximum *a posteriori* Bayesian objective function.[30] The EBEs were estimated for each patient based on the information available from the start of the study up to the end of each consecutive 24-hour period throughout the study. Eventually, each patient had n_i sets of EBEs, with n_i equal to the number of study days for that patient. This procedure was repeated for three information scenarios, where information observed up to the end of each 24-hour period was included in the estimation of the EBEs:

1. PK: Only plasma FVIII activity levels were included, the EBEs related to PK were estimated, and the remaining EBEs were set to zero.
2. Bleed: Only the times of bleeding events were included, the EBEs related to the bleeding hazard were estimated, and the remaining EBEs were set to zero.
3. All: Plasma FVIII activity levels and time of bleeding events as well as all covariate information were included in the estimation of all EBEs.

For Bleed and All, a right-censored observation was assumed at the end of each 24-hour period. For all scenarios, patients' dose information, predicted time-varying FVIII activity, and observed LBW were used during forecasting.

Bleeding probabilistic forecast

The individual forecasted probability of having a bleed in the upcoming 24-hour period, $P_i(bleeding)$, was calculated as follows:

$$P_i(bleeding) = 1 - S_i(t+24) = 1 - e^{-\int_t^{t+24} h_i(t)dt} \quad (2)$$

where $S_i(t)$ is the probability of not having a bleed between t and $t + 24$ derived from the time-varying hazard $h_i(t)$, and t is the end of the observation period for estimation of the EBEs.

Doses registered during the forecasted 24-hour period were considered in the calculation of $h_i(t)$ (Eq. 1) to reflect the true individual FVIII activity levels and are therefore reflected in $P_i(bleeding)$. However, doses administered between the occurrence of a bleed and the end of the corresponding 24-hour period were considered to be on-demand doses and therefore not included in the probabilistic forecast. The forecasted number of total bleeds during the study time was calculated as the sum of the $P_i(bleeding)$ values over the whole study period.

Variation of Bayesian forecasting observation period

In the default case, the EBE estimation was based on observations from the start of the study. In addition, observation periods including the past 15 days or 1, 2, 3, or 6 months for each patient were explored using the Bleed information scenario. This step was performed to assess the influence of varying the observation period for EBE estimation within each patient and learn about the trade-off between a longer period (more individual information accumulated for future predictions) and the most recent period (the most up-to-date information).

Predictive performance assessment

The predictive performance was assessed by comparing P_i(bleeding), with the observed occurrence of a bleed in the forecasted day using the following analyses with graphical illustrations. Separation plots were used to illustrate how the individual probabilistic forecast matched with the occurrence of a bleed.[31] In these plots, P_i(bleeding) was plotted in ascending order of P_i(bleeding), highlighting the P_i(bleeding) associated with the bleed occurrence in the forecasted day. In addition, we performed receiver operating characteristic (ROC) analyses to compare the predictive performance between scenarios.[32,33] An ROC curve that overlaps the diagonal line, i.e., the random-classifier baseline, corresponds to an uninformative test, equivalent to a random classifier, and the closer the ROC curve is to the upper left corner, the higher the overall performance of the test. The ROC curves were compared using the area under the curve (AUC), calculated with the trapezoidal rule, with higher AUC values indicating higher accuracy.[34] The optimal threshold with respect to P_i(bleeding) for the discrimination between bleeds and no bleeds was determined by the Youden's *J* statistics.[35] The 95% confidence intervals (CI) of ROC-related summary indices were calculated based on a stratified bootstrap (2,000 replicates). Furthermore, we performed precision-recall (PR) analyses to assess the performance of information scenarios when considering precision. The AUC calculation of the PR curve was calculated using nonlinear interpolation, as suggested by Davis and Goadrich.[36]

The individual prediction error (PE_i) for the *i*th patient was calculated as:

$$PE_i = Bleed_{forecast} - Bleed_{observed} \qquad (3)$$

where $Bleed_{forecast}$ is the number of forecasted bleeds and $Bleed_{observed}$ is the number of observed bleeds during the whole individual study period.

Software

Bayesian forecasting was performed using NONMEM version 7.4.3 (ICON Development Solutions, Ellicott City, MD), and Perl-speaks-NONMEM version 4.8.1 was used for NONMEM run control.[37,38] Data management and graphical analyses were performed using R 3.5.1 (http://www.rproject.org/).

RESULTS

Bayesian forecasting

The median forecasting period for patients aged < 12 years was 183 days (range 113–217 days), and for patients aged ≥ 12 years was 365 days (range 93–394 days). The values of η_i^{CL} and $\eta_i^{h_i(t)}$ over time, obtained through Bayesian forecasting, are represented in **Figure 1**.

For the information scenario PK, η_i^{CL} values were zero at the start of the study and changed when PK observations became available, and $\eta_i^{h_i(t)}$ values were set to zero because no information about bleeds was provided. In contrast, when bleeds were considered, η_i^{CL} values were set to zero and $\eta_i^{h_i(t)}$ values were zero at the start of the study and increased when a subject had a bleed or decreased otherwise. For All, η_i^{CL} and $\eta_i^{h_i(t)}$ were informed by covariates, PK, and bleed data from the start of the study.

Bleeding probabilistic forecast

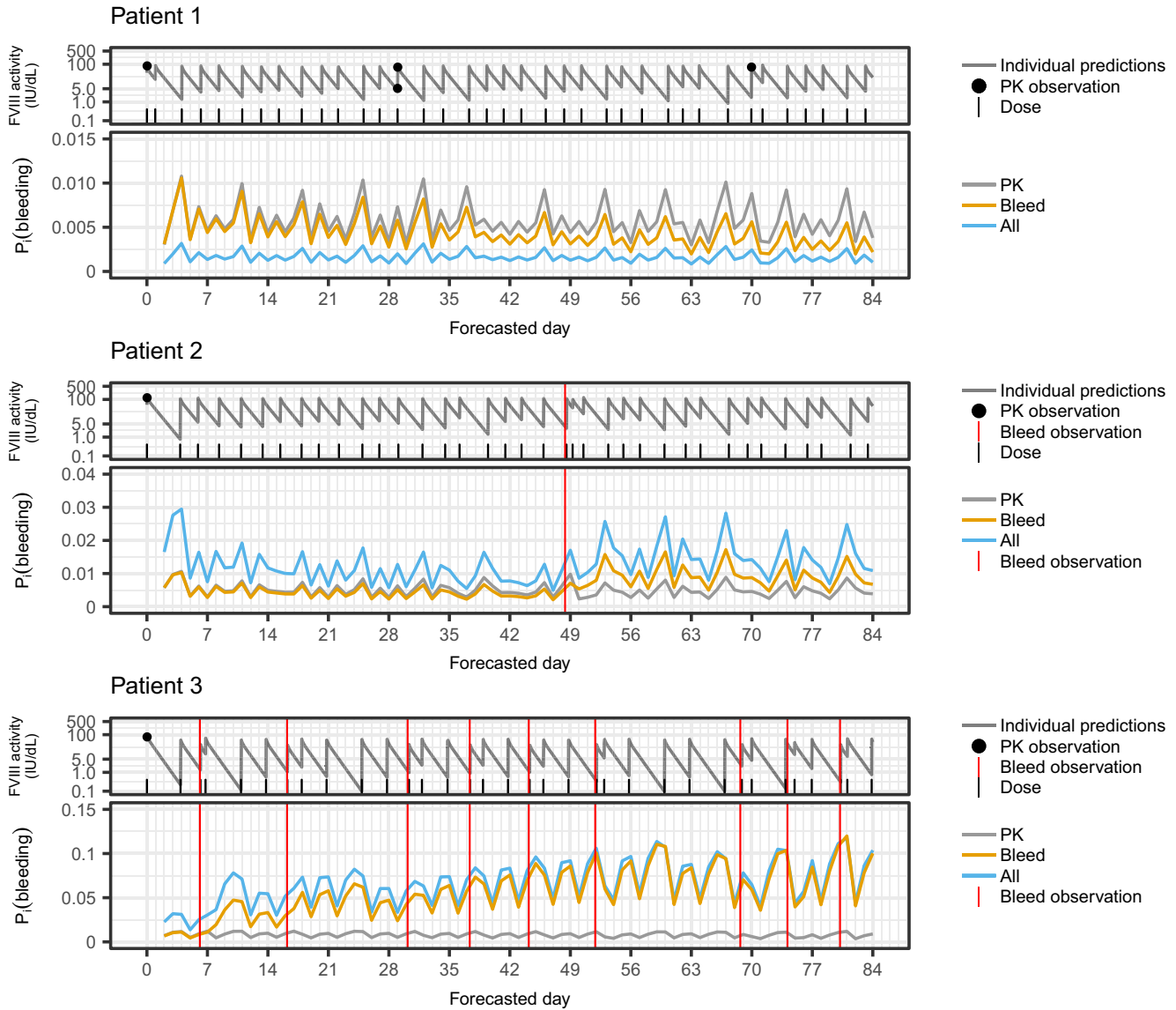The observed and model-predicted plasma FVIII activity, time of bleeds, and forecasted P_i(bleeding) for the three information scenarios for three illustrative patients throughout the first 12 weeks is shown in **Figure 2**.

The probability of having a bleed on a certain day is inversely correlated with the plasma FVIII activity predicted for that same day, as expected given the FVIII activity–
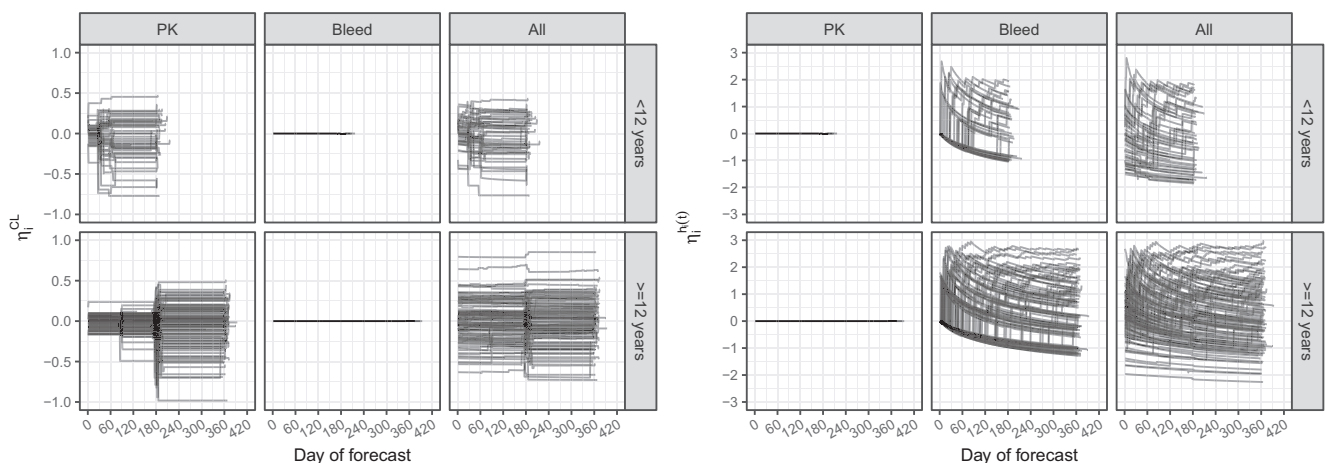
Figure 1 Longitudinal changes in η_i^{CL} (left) and $\eta_i^{h_i(t)}$ (right) when updated on a 24-hour basis based on each information scenario from the start of the study. Plots for η_i^{V1} and η_i^{ϵ} are available in **Supplemental Material Figure S1.**

Figure 2 Observed and model-predicted plasma FVIII activity (upper panels), forecasted bleeding probability (P_i(bleeding)) (lower panels), and observed time of bleeds for three illustrative patients with no bleeds (patient 1), one bleed (patient 2), and multiple bleeds (patient 3). FVIII, factor VIII.

hazard relationship defined in the Bayesian prior model. At the start of the study, the bleeding probabilities associated with PK or Bleed are similar, provided that no bleeds occurred. Throughout time, the forecasted probability given by PK remains relatively constant, regardless of the occurrence of bleeds, and is a function of dosing schedule and PK parameters, with the latter reevaluated when PK measurements are available. In contrast, the probability using only bleed data tends to decrease provided that no bleeds occurred (patient 1, patient 2 before bleed on day 49, patient 3 before bleed on day 7) or to increase if a patient has a bleed (patient 2 after day 49 and patient 3 after bleed on day 7). Moreover, the forecast based on Bleed tends to become closer to the forecast based on all information throughout time, and both tend to decrease over time, provided that no bleeds occur.

Predictive performance

The total number of observed bleeds during the evaluation period was 101 bleeds for the group aged < 12 years, and 530 bleeds for the group aged ≥ 12 years. At a population level, the forecasted number of bleeds during the same period for the group aged < 12 years was 66 (PK), 96 (Bleed), and 90 bleeds (All), and for the group aged ≥ 12 years was 218 (PK), 461 (Bleed), and 500 bleeds (All). The difference in P_i(bleeding) between days when bleeds occurred and no-bleed days was higher for Bleed and All than for PK, as represented in **Figure 3**. In addition, the occurrence of bleeds was associated with a sharper change in P_i(bleeding) for the Bleed and All scenarios when compared with PK, which was observed for both age groups.

The ROC curves indicated Bleed and All superior to PK, with the former having higher true positive rates and lower
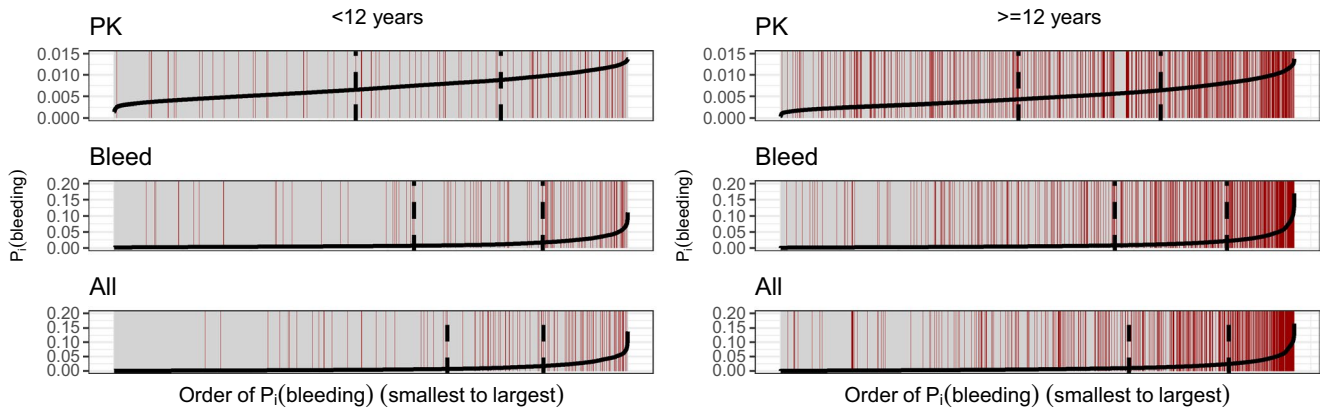
Figure 3 Separation plots for the bleeding probabilistic forecast P_i(bleeding) for the information scenarios PK, Bleed, and All (note difference in *y*-axis scale). Bleed events are represented with red vertical bars, and no-bleed events are represented with gray vertical bars. Since the occurrence of bleeds was rare, vertical bars corresponding to bleed events were emphasized. The solid black line represents the values of P_i(bleeding) associated with bleeds (red vertical bars) or no bleeds (gray vertical bars), and the black vertical dashed lines represent the 25th and 50th percentiles of the bleeding event data (red vertical bars). More red vertical bars located on the right-hand side of the plot indicates better predictive performance.

false positive rates at all cut-off values (**Figure 4**). For the group aged < 12 years, the AUC for PK was 0.67 (95% CI, 0.61–0.72), being slightly lower than for Bleed and All with AUC 0.74 (95% CI, 0.69–0.79) and 0.77 (95% CI, 0.73–0.81), respectively (**Table 2**). For the group aged ≥ 12 years, the difference between scenarios was more evident with the AUC for PK of 0.67 (95% CI, 0.65–0.69) being considerably lower than for Bleed, 0.78 (95% CI, 0.76–0.80). Bleed and All had similar performances (**Table 2**).

A similar trend was found in the PR analysis (**Figure S2**), with the information scenarios All and Bleed being superior

to PK, with PK closer to the performance of a random classifier. For the group aged < 12 years, the AUC of the PR curves were 0.023 (PK), 0.030 (Bleed), 0.033 (All), and 0.011 (random classifier). For the group aged ≥ 12 years, the AUC of the PR curves were 0.028 (PK), 0.064 (Bleed), 0.064 (All), and 0.012 (random classifier).

Variation of Bayesian forecasting observation period
The prediction error related to the forecast of bleeds for the whole study while varying the observation period is shown in **Figure 5** for the Bleed scenario (chosen given the minor
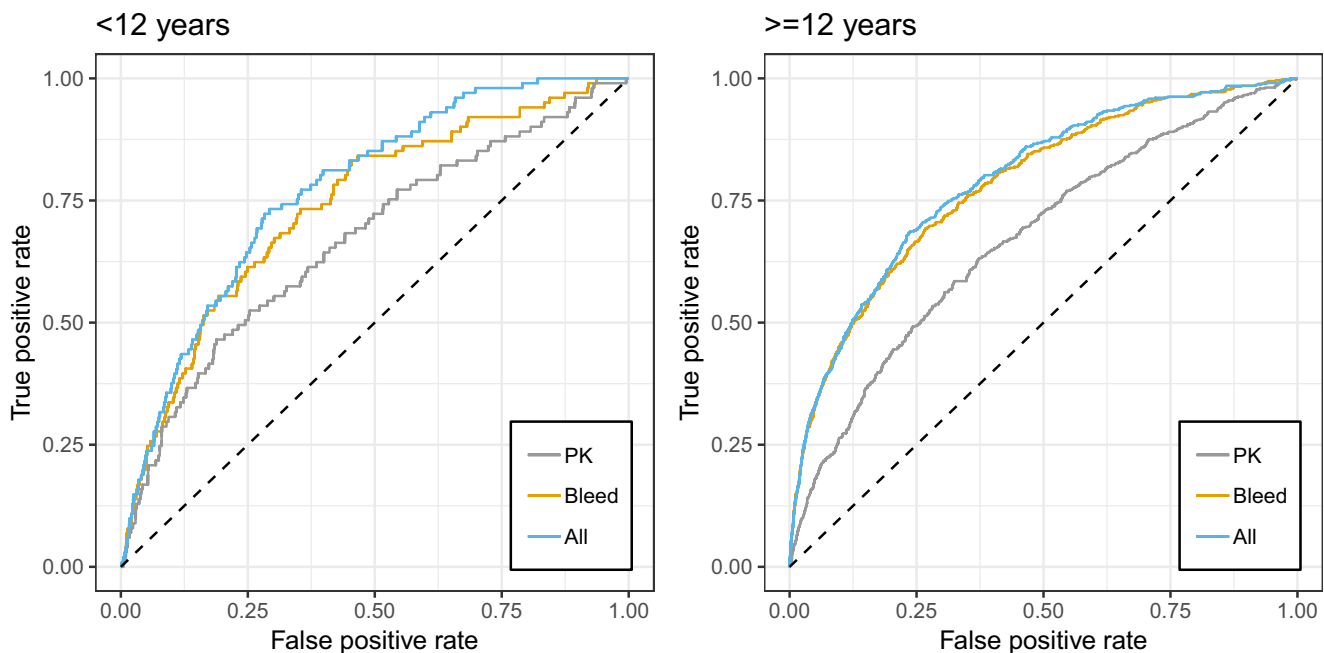

Figure 4 Receiver operating characteristic curves for the ability of each information scenario to predict the occurrence of bleeds. The dashed line is the random-classifier baseline.

Table 2 Summary statistics of the receiver operating characteristic analyses and the optimal thresholds

	< 12 years old			≥ 12 years old		
	PK	**Bleed**	**All**	**PK**	**Bleed**	**All**
Threshold (95% CI)	0.0095 (0.0064–0.010)	0.0079 (0.0061–0.017)	0.0086 (0.0059–0.012)	0.0058 (0.0051–0.0070)	0.011 (0.0079–0.015)	0.014 (0.010–0.015)
AUC (95% CI)	0.67 (0.61–0.72)	0.74 (0.69–0.79)	0.77 (0.73–0.81)	0.67 (0.65–0.69)	0.78 (0.76–0.80)	0.79 (0.77–0.81)
Sensitivity (95% CI)	0.47 (0.39–0.81)	0.73 (0.53–0.90)	0.73 (0.64–0.86)	0.58 (0.47–0.68)	0.69 (0.62–0.78)	0.69 (0.65–0.77)
Specificity (95% CI)	0.81 (0.46–0.87)	0.65 (0.53–0.84)	0.71 (0.60–0.76)	0.68 (0.59–0.80)	0.73 (0.65–0.81)	0.76 (0.69–0.78)
J	0.28	0.38	0.44	0.26	0.43	0.45

AUC, area under the receiver operating characteristic curve; CI, confidence interval; J, Youden's J statistics.
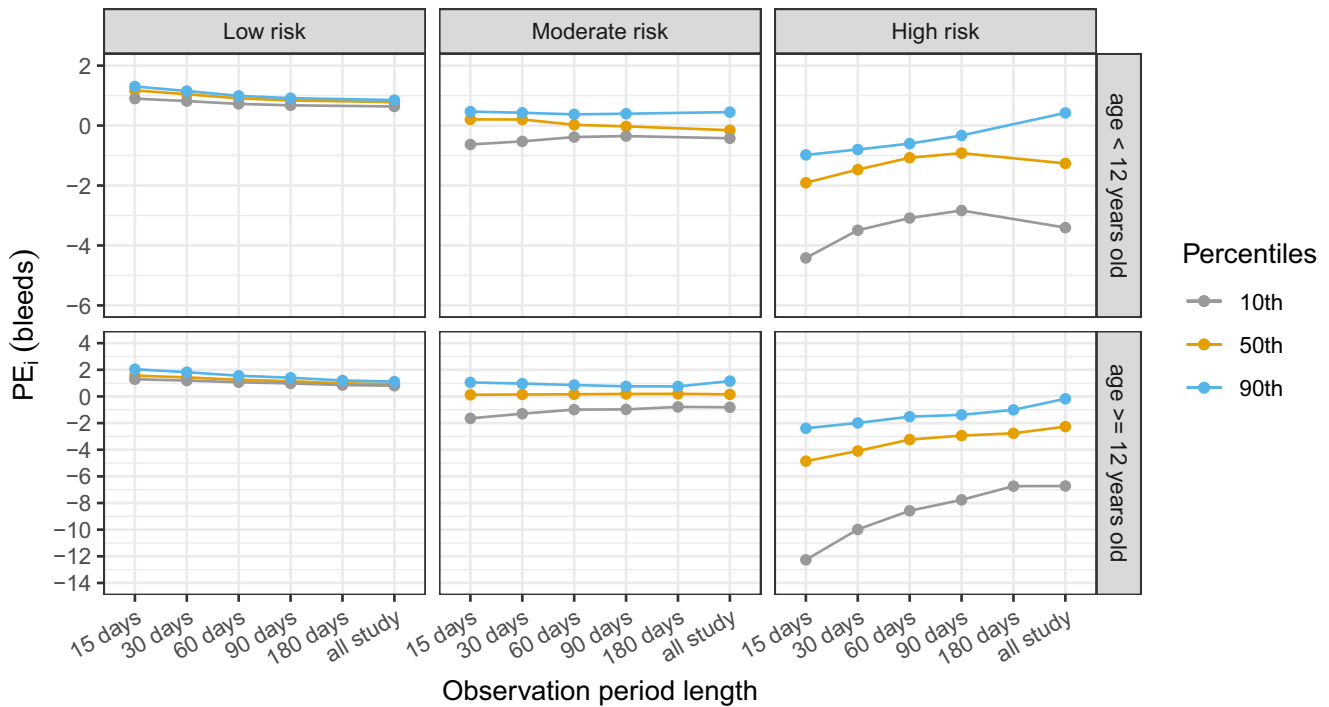
Figure 5 Percentiles of the prediction error (PE_i) while varying the Bayesian forecasting observation period length, stratified by bleeding risk and age group for the scenario Bleed. Bleeding risk was categorized as low risk if patients did not bleed during the study (< 12 years, $N = 23$ patients; ≥ 12 years, $N = 33$); moderate risk if 1–2 bleeds were observed in the group aged < 12 years ($N = 11$) or 1–4 bleeds in the group aged ≥ 12 years ($N = 49$), and high risk if ≥ 3 bleeds were observed in the group aged < 12 years ($N = 17$) or ≥ 5 bleeds in the group aged ≥ 12 years ($N = 39$).

performance differences when compared with All). The separation plots and results of the ROC analyses can be found in **Figures S3 and S4** and **Table S1**.

Bleeds in patients with a low bleeding risk were overpredicted, improving marginally when longer observation periods were used to estimate the EBEs. In contrast, bleeds in patients with a high bleeding risk were underpredicted and longer observation periods were associated with improved forecasts. In these patients, the estimation of the EBEs was informed mostly by observations in the previous 60–90 days. Concerning the bias observed, it is important to recognize that any good model would underpredict bleeding values higher than the median number of bleeds and overpredict values lower than the median as a result of η-shrinkage. The

prediction of bleeds in patients with a moderate bleeding risk was unbiased, and an observation period longer than 60 days did not contribute substantially with information to improve the bleeding forecast.

DISCUSSION

Pharmacokinetic-based dosing with sparse plasma FVIII activity observations has been proposed as a tool to optimize prophylactic dosing regimens in hemophilia A. In this study, we assessed the use of a PK-RTTE model for Bayesian forecasting of bleeding events and compared the ability to predict future bleeds using different sources of patient information, namely, PK observations, bleeds, or all information

available. The PK data, time of bleeds, and covariates were those observed in children and adults during the LEOPOLD studies. The Bayesian forecasting of bleeds based on individual daily bleeding information was found to perform better than forecasting based on individual PK information only.

The forecasted number of bleeds during the study was lower for PK than for the other approaches because no bleeding information was used. The distribution of the number of observed bleeds was positively skewed (median 2 bleeds, mean 3.7 bleeds), and the bleeding hazard was modeled using a log-normal distribution. Therefore, in the absence of bleeding information, the typical hazard was translated into the median number of bleeds, which resulted in an underprediction of the number of bleeds in patients with a high bleeding risk. Separation plots can be used to compare predictors of events without having to consider probability thresholds. In these plots, the information scenario Bleed had a similar performance when compared with All, and both were superior to PK. The same conclusion can be drawn from the ROC analysis with AUC values for Bleed being substantially higher than PK and only slightly worse than All. We also performed a PR analysis because the ROC analysis could be misleading in the presence of an imbalanced class distribution, as the number of no bleeds outweighs considerably the number of bleeds (bleeds occurred during only ~ 1% of the forecasted days).[39] The PR analysis was in agreement with the ROC analysis with respect to the ranking of the approaches. However, the PR analysis, which specifically focus on precision, revealed a low precision for all information scenarios. This is expected because the occurrence of a bleed has a sustained effect on P_i(bleeding), which remains above the optimal probability thresholds, thus leading to many false positive values.

Overall, the minor improvement of including all information, compared with only including bleeds, confirms the previous observation that covariates are weak predictors of the occurrence of bleeds.[24] Past observed bleeding information driving the forecast of future bleeds agrees with the theoretical mathematical expectations because the unexplained interindividual variability on the bleeding hazard is large (136% CV). In addition, it agrees with clinical expectations because the clinical phenotype between patients with severe hemophilia A is known to be variable, and patients may respond differently to identical plasma FVIII activity levels.[12,18–21] Information about individual PK is relevant to understand intraindividual fluctuations in the bleeding hazard, which might be useful to adapt dosing regimens to the patient's needs (e.g., during a period of more intense physical exercise). According to this evaluation, bleed information may be more informative than PK to characterize the interpatient bleeding risk. However, both individual PK and bleeding risk may be necessary to decide a suitable course of action with respect to the future individualization of dosing regimens.

According to our results, the longer the observation period for EBEs estimation the better, with no advantage to use only the most up-to-date information. Patients with a high bleeding risk were found to require shorter observation periods to inform the EBEs, namely, between 60 and 90 days prior to the EBEs estimation. In contrast, patient

with no bleeds had a much slower learning curve and required longer observation periods. However, these results should be interpreted with caution because in the group aged < 12 years the information was limited (lower number of bleeds and shorter observation periods) for the assessment of the relative merits of the different observation periods.

One limitation of this study is that the individual observed PK, bleed, and covariate data used for Bayesian forecasting were those included in the development of the prior model used in the forecast. The variability found in a real-world scenario might be higher than in the LEOPOLD clinical trials. However, using the same set of data decreases bias related to an inappropriate structural or stochastic prior model and provides a realistic comparison of the bleed prediction between information scenarios. The methodology applied in our study, using maximum *a posteriori* Bayesian estimation, to estimate the individual PK parameters using the information scenario PK was similar to the existing tools available for PK-based FVIII dose individualization employing parametric methods. However, a limitation of our evaluation is that the number of PK samples and sampling times available during the LEOPOLD studies did not correspond necessarily to the 2–3 samples recommended for FVIII individual PK profiling used in PK-based prophylaxis.[5,9] On some occasions, only one sample was available, whereas on others more samples were available. In addition, the treatment with FVIII in the LEOPOLD studies was efficient, and patients were overall well protected against bleeds by allowing the dose to vary between patients where the individual PK (known from other products) and bleeding phenotype were indirectly considered. Therefore, an extrapolation of our findings to PK-based prophylaxis or when considering other dosing regimens has to be done with caution.

Several aspects related to our evaluation deserve further study. This work focuses on forecasting bleeds. The translation of the findings to dose optimization through model-based bleed-guided and PK-guided prophylaxis in the clinic requires further exploration. The magnitude of the individual dose change in the case of dose adaption needs to be defined, and the identification of the threshold value of the ROC curves for individual dose adaptation can be expanded to include an efficiency criterion, accounting for the false-positive and false-negative cost.[40] In addition, information about covariates may have the potential to improve the predictive performance of the Bleed approach, in particular, when a patient is observed for the first time, and deserves further exploration. Finally, the model used in this study assumed a bleeding hazard that decreases over time, likely because of a treatment or trial effect, and how the information strategies perform in the presence of a time-independent hazard merits further analysis.

In this study, we evaluated if an integrated PK-RTTE model could be used to predict the occurrence of future bleeds using past observed data from clinical trials. Bayesian forecasting using such a model can be applied to other studies involving hemophilia A patients to adapt dosing regimens and to further assess the importance of different sources of patient information to include during Bayesian forecasting. In addition, the approach proposed in this study, through the

calculation of the individual bleeding probabilistic forecast based on the individual's information about bleeds, could be implemented in a platform providing patients and clinics with the prospective time-varying risk of having a bleed. Furthermore, this platform could alert the patient to consult the treating physician for a possible change in the dosing regimen in case the forecasted bleeding probability is above a certain probability threshold for a certain time period.

In conclusion, we developed an integrated PK-RTTE approach for bleeding forecast considering the individual historical bleeding pattern under prophylactic treatment with a recombinant FVIII product. In addition, this work presents promising results suggesting that individual bleeding information adds value to the optimization of prophylactic dosing regimens in severe hemophilia A. Further steps to optimize the proposed tool for FVIII dose adaptation in the clinic are required.

1. Onwuzurike, N., Warrier, I. & Lusher, J.M. Types of bleeding seen during the first 30 months of life in children with severe haemophilia A and B. *Haemophilia* **2**, 137–140 (1996).
2. Collins, P.W. *et al.* Break-through bleeding in relation to predicted factor VIII levels in patients receiving prophylactic treatment for severe hemophilia A. *J. Thromb. Haemost.* **7**, 413–420 (2009).
3. Ahlberg, A. Haemophilia in Sweden. VII. Incidence, treatment and prophylaxis of arthropathy and other musculo-skeletal manifestations of haemophilia A and B. *Acta Orthop. Scand.* **77** (suppl.), 73–132 (1965).
4. Berntorp, E. If you know you will also see: population pharmacokinetics is the way to personalize and optimize prophylaxis in hemophilia. *J. Thromb. Haemost.* **15**, 1103–1105 (2017).
5. Iorio, A. *et al.* Estimating and interpreting the pharmacokinetic profiles of individual patients with hemophilia A or B using a population pharmacokinetic approach: communication from the SSC of the ISTH. *J. Thromb. Haemost.* **15**, 2461–2465 (2017).
6. Abrantes, J.A., Nielsen, E.I., Korth-Bradley, J., Harnisch, L. & Jonsson, S. Elucidation of factor VIII activity pharmacokinetics: a pooled population analysis in patients with hemophilia a treated with moroctocog alfa. *Clin. Pharmacol. Ther.* **102**, 977–988 (2017).
7. Bjorkman, S. *et al.* Population pharmacokinetics of recombinant factor VIII: the relationships of pharmacokinetics to age and body weight. *Blood* **119**, 612–618 (2012).
8. Garmann, D. *et al.* Population pharmacokinetic characterization of BAY 81-8973, a full-length recombinant factor VIII: lessons learned—importance of including samples with factor VIII levels below the quantitation limit. *Haemophilia* **23**, 528–537 (2017).
9. Bjorkman, S. & Collins, P. Measurement of factor VIII pharmacokinetics in routine clinical practice. *J. Thromb. Haemost.* **11**, 180–182 (2013).
10. Iorio, A. *et al.* Development of a web-accessible population pharmacokinetic service-hemophilia (WAPPS-Hemo): study protocol. *JMIR Res. Protoc.* **5**, e239 (2016).
11. Alvarez-Roman, M.T. *et al.* Experience of tailoring prophylaxis using factor VIII pharmacokinetic parameters estimated with myPKFiT((R)) in patients with severe haemophilia A without inhibitors. *Haemophilia* **23**, e50−e54 (2017).
12. Valentino, L.A. *et al.* A randomized comparison of two prophylaxis regimens and a paired comparison of on-demand and prophylaxis treatments in hemophilia A management. *J. Thromb. Haemost.* **10**, 359–367 (2012).
13. Iannazzo, S. *et al.* Cost-effectiveness analysis of pharmacokinetic-driven prophylaxis vs. standard prophylaxis in patients with severe haemophilia A. *Blood Coagul. Fibrinolysis* **28**, 425–430 (2017).
14. Pasca, S., Milan, M., Sarolo, L. & Zanon, E. PK-driven prophylaxis versus standard prophylaxis: When a tailored treatment may be a real and achievable cost-saving approach in children with severe haemophilia A. *Thromb. Res.* **157**, 58–63 (2017).
15. Lissitchkov, T. *et al.* PK-guided personalized prophylaxis with Nuwiq((R)) (human-cl rhFVIII) in adults with severe haemophilia A. *Haemophilia* **23**, 697–704 (2017).
16. Megias-Vericat, J.E. *et al.* Bayesian pharmacokinetic-guided prophylaxis with recombinant factor VIII in severe or moderate haemophilia A. *Thromb. Res.* **174**, 151–162 (2019).
17. Nagao, A., Yeung, C.H.T., Germini, F. & Suzuki, T. Clinical outcomes in hemophilia A patients undergoing tailoring of prophylaxis based on population-based pharmacokinetic dosing. *Thromb. Res.* **173**, 79–84 (2019).
18. Bjorkman, S. Prophylactic dosing of factor VIII and factor IX from a clinical pharmacokinetic perspective. *Haemophilia* **9** (suppl. 1), 101–108 (2003); discussion 109–110.
19. Ahnstrom, J., Berntorp, E., Lindvall, K. & Bjorkman, S. A 6-year follow-up of dosing, coagulation factor levels and bleedings in relation to joint status in the prophylactic treatment of haemophilia. *Haemophilia* **10**, 689–697 (2004).
20. Aledort, L.M., Haschmeyer, R.H. & Pettersson, H. A longitudinal study of orthopaedic outcomes for severe factor-VIII-deficient haemophiliacs. The Orthopaedic Outcome Study Group. *J. Intern. Med.* **236**, 391–399 (1994).
21. Petrini, P. What factors should influence the dosage and interval of prophylactic treatment in patients with severe haemophilia A and B? *Haemophilia* **7**, 99–102 (2001).
22. Valentino, L.A. *et al.* Association of peak factor VIII levels and area under the curve with bleeding in patients with haemophilia A on every third day pharmacokinetic-guided prophylaxis. *Haemophilia* **22**, 514–520 (2016).
23. Iorio, A. *et al.* Target plasma factor levels for personalized treatment in haemophilia: a Delphi consensus statement. *Haemophilia* **23**, e170–e179 (2017).
24. Abrantes, J.A. *et al.* Relationship between factor VIII activity, bleeds and individual characteristics in severe haemophilia A patients. *Haematologica.* 2019 https://doi.org/10.3324/haematol.2019.217133. [e-pub ahead of print].
25. Saxena, K. *et al.* Efficacy and safety of BAY 81-8973, a full-length recombinant factor VIII: results from the LEOPOLD I trial. *Haemophilia* **22**, 706–712 (2016).
26. Kavakli, K. *et al.* Prophylaxis vs. on-demand treatment with BAY 81-8973, a full-length plasma protein-free recombinant factor VIII product: results from a randomized trial (LEOPOLD II). *J. Thromb. Haemost.* **13**, 360–369 (2015).
27. Ljung, R. *et al.* BAY 81-8973 safety and efficacy for prophylaxis and treatment of bleeds in previously treated children with severe haemophilia A: results of the LEOPOLD Kids Trial. *Haemophilia* **22**, 354–360 (2016).
28. Yngman, G., Nyberg, J., Jonsson, N. & Karlsson, M.O. Practical considerations for using the full random effects modeling (FREM) approach to covariate modeling. Abstracts of the Annual Meeting of the Population Approach Group in Europe, Budapest, Hungary, June 6–9, 2017. Abstract 7365. < https://www.page-meeting.org/default.asp?abstract=7365 > (2017).
29. Karlsson, M.O. A full model approach based on the covariance matrix of parameters and covariates. Abstracts of the Annual Meeting of the Population Approach Group in Europe, Venice, Italy, June 5–8, 2012. Abstract 2455. < https://www.page-meeting.org/default.asp?abstract=2455 > (2012).
30. Sheiner, L.B., Beal, S., Rosenberg, B. & Marathe, V.V. Forecasting individual pharmacokinetics. *Clin. Pharmacol. Ther.* **26**, 294–305 (1979).
31. Greenhill, B., Ward, M.D. & Sacks, A. The separation plot: a new visual method for evaluating the fit of binary models. *Am. J. Political Sci.* **55**, 991–1002 (2011).
32. Fawcett, T. An introduction to ROC analysis. *Pattern Recogn. Lett.* **27**, 861–874 (2006).
33. Lusted, L.B. Decision-making studies in patient management. *N. Engl. J. Med.* **284**, 416–424 (1971).
34. Hanley, J.A. & McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29–36 (1982).
35. Youden, W.J. Index for rating diagnostic tests. *Cancer* **3**, 32–35 (1950).
36. Davis, J. & Goadrich, M. The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning.* 233–240 (ACM, New York, 2006).

37. Beal, S.L., Sheiner, L.B., Boeckmann, A.J. & Bauer, R.J. (Eds). NONMEM 7.4 Users Guides (ICON plc, Gaithersburg, MD, 1989–2018).
38. Keizer, R.J., Karlsson, M.O. & Hooker, A. Modeling and simulation workbench for NONMEM: tutorial on Pirana, PsN, and Xpose. *CPT Pharmacometrics Syst. Pharmacol.* **2**, e50 (2013).
39. Saito, T. & Rehmsmeier, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE* **10**, e0118432 (2015).
40. Clements, J.D. *et al.* Receiver operating characteristic analysis and clinical trial simulation to inform dose titration decisions. *CPT Pharmacometrics Syst. Pharmacol.* **7**, 771–779 (2018).