Type and intensity of FVIII exposure on inhibitor development in PUPs with haemophilia A

A patient-level meta-analysis

Maura Marcucci1,2; Maria Elisa Mancuso3; Elena Santagostino3; Gili Kenet4; Mohsen Elalfy5; Susanne Holzhauer6; Christoph Bidelmaier7; Carmen Escuriola Ettingshausen8; Alfonso Iorio9,10; Ulrike Nowak-Göttl11

1Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada; 2Department of Clinical Sciences and Community Health, University of Milan & Geriatrics, Fondazione IRCCS Ca Granda, Ospedale Maggiore Policlinico, Milan, Italy; 3Angelo Bianchi Bonomi Haemophilia and Thrombosis Centre, Fondazione IRCCS Ca’ Granda, Ospedale Maggiore Policlinico, Milan, Italy; 4The Israel National Haemophilia Centre, Sheba Medical Centre, Tel-Hashomer and the Sackler Medical School, Tel Aviv University, Israel; 5Haemophilia Centre, Ain Shams University, Egypt; 6Department of Pediatric Hematology/Oncology, Charité, Berlin, Germany; 7Department of Pediatrics, University Hospital Munich, Germany; 8HZRM- Haemophilia Centre Rhein Main, Frankfurt-Mörfelden, Germany; 9Department of Medicine, McMaster University, Hamilton, Ontario, Canada; 10Institute of Clinical Chemistry, Thrombosis and Haemostasis Treatment Centre, University Hospital Schleswig-Holstein, Kiel, Germany

Summary

The impact of treatment-related factors on inhibitor development in previously untreated patients (PUPs) with haemophilia A is still debated. We present the results of a collaborative, individual patient data meta-analytic project. Eligible data sources were published cohorts of PUPs for which patient-level data were available. The exposures of interest were factor (F)VIII type (recombinant [rFVIII] vs plasma-derived [pdFVIII]) and treatment intensity (≥ vs < 150 IU/kg/week) at first treatment. Family history of inhibitors, F8 mutations, age, treatment regimen (on-demand vs prophylaxis), secular trend and surgery were analysed as putative confounders using different statistical approaches (multivariable Cox regression, propensity score analyses, CART). Analyses accounted for the multi-centre origin of the data. We included 761 consecutive, unselected PUPs with moderate to severe haemophilia A from 10 centres in Egypt, Germany, Israel and Italy. A total of 27% of patients developed inhibitors; 40% and 22% of patients treated with rFVIII and pdFVIII (unadjusted HR 2.2, 95% CI 1.6–2.9), respectively; 51% and 24% of patients receiving high- and low-intensity treatment (unadjusted HR 2.9, 95% CI 2.0–4.2), respectively. In adjusted analyses, only treatment intensity remained an independent predictor; the effect of FVIII type was largely due to confounding, but with a significant interaction between FVIII type and treatment intensity. This patient-level meta-analysis confirms, across different statistical approaches, that high-intensity treatment is a strong risk factor for inhibitor development. The possible role of FVIII type in subgroups is suggested by the test for interactions but could not be proven because of the limited subgroups sample sizes.

Keywords
Factor VIII inhibitors, haemophilia A / B, metaanalysis, risk factors

Introduction

The development of inhibitory antibodies against factor VIII (FVIII) is the most important clinical challenge in patients with haemophilia A (HA), with relevant implications for patient outcomes and societal costs (1). The question of whether treatment modality, in terms of FVIII type (recombinant [rFVIII] vs plasma-derived [pdFVIII]) or treatment intensity, independently affects the risk of inhibitor development can be addressed using various study designs. In the specific context of rare diseases, clinicians often face a paucity of evidence from highly ranked study designs, such as meta-analyses or randomised controlled trials (2). A unique randomised controlled trial (Study of Inhibitors in Plasma-Product Exposed Toddlers, SIPPET) (3) is currently ongoing. In the meantime, a large observational cohort (RODIN study [4, 5]) found a higher risk of inhibitors associated with a second-generation rFVIII product, a result that has been extensively debated (6–8). Appropriate appraisal and synthesis of the existing evidence base is an alternative approach to address the question of the effect of treatment modality on inhibitor development.

We previously published a systematic review and meta-analysis of observational, uncontrolled cohorts of previously untreated patients (PUPs) to explore the possible differential effects of pdFVIII and rFVIII on the development of inhibitors (9). We concluded by recommending careful consideration of confounders in the analysis of observational data. We also noted the limitations of meta-
analyzing aggregate study-level data, which could be overcome by using individual patient data (IPD) (10–12).

Under the auspices of the European Association for Haemophilia and Allied Disorders (EAHAD), we conducted a collaborative project to explore and synthesise the available evidence on the role of modifiable, treatment-related risk factors for inhibitor development by pooling and meta-analysing IPD from cohorts of PUPs with moderate and severe HA. We present here the main results on the effects of type of FVIII concentrate and intensity of treatment, analysed using different statistical approaches to strengthen the degree of confidence in our results.

Methods

The research protocol for this collaborative work, including the process for sharing and managing patient-level data, was developed by a core group of investigators (A.I., M.M and U.N-G.) and approved by all study authors before project initiation. All outcomes and analyses were established a priori, with the single exception of the investigation of differences in risk among individual concentrates, which was added as a secondary analysis after publication of the RODIN study results (4). A waiver of the need to re-consent patients to the data transfer was obtained by the Ethical Committee of the coordinating centre.

Study search and selection

Published cohort studies were considered eligible if they reported data on inhibitor development in patients with HA after their first exposure to any type of FVIII product. All studies included in our previous meta-analysis (9) on inhibitor development in PUPs were evaluated for potential inclusion. The systematic search of electronic databases (Medline, EMBASE, Web of Science, Cochrane Library) was repeated to identify studies published from 2009 to December 2012; no language restriction was imposed. The reference lists of journal articles identified, and meeting proceedings of the World Federation of Haemophilia, International Society on Thrombosis and Haemostasis, American Society of Hematology and EAHAD were manually checked to identify additional studies. The primary authors of the retrieved studies were asked to provide IPD. Mixed cohorts of PUPs and previously treated patients (PTPs) were also considered for inclusion, but only IPD for PUPs were requested.

Data collection and database generation

A standardised data collection spreadsheet was prepared to collect the patient-level variables of interest. The following data were included: patient local identifier, year of birth, severity of disease, baseline FVIII level, FVIII gene (F8) mutation, ethnicity, family history of inhibitors, date of first FVIII infusion, treatment characteristics at first exposure (on-demand or prophylaxis, dose, frequency of infusions and type of concentrate with brand name), surgery in the month before/after first FVIII infusion, number of exposure-days at study end or inhibitor development, age at inhibitor development, inhibitor peak titre, inhibitor testing frequency, and inhibitor assay characteristics. Data were transferred to a central location under the auspices of two investigators (A.I., M.M.) and merged into a single pooled database. Data were checked and queries on coding, uncertain or missing data were circulated among source authors.

Definitions

Patients

PUPs were defined as patients who were naïve to FVIII replacement therapy. Patients with previous exposure to other blood components (i.e. red blood cells or platelets) or with <5 exposure days (EDs) to fresh frozen plasma were also considered to be PUPs. Haemophilia was classified according to baseline FVIII level as severe (FVIII: C < 0.01 IU/ml) or moderate (0.01 ≤ FVIII: C ≤ 0.05 IU/ml). Patients with mild disease were not included in this analysis.

Outcomes

The main outcome was defined as the development of inhibitors with a titre >0.6 Bethesda Units per ml (BU/ml), confirmed on at least two consecutive samples. Inhibitors were coded as high-responding when the peak titre was ≥ 5.0 BU/ml.

Predictors (i.e. determinants of inhibitor risk)

• a) Exposures of interest: type of FVIII and intensity of treatment. For the main analysis, type of FVIII concentrate at first exposure was classified as pdFVIII or rFVIII. This assignment was irrespective of any subsequent switch in concentrate before the 50th ED. The intensity of treatment was defined as low or high when the patient received a dose of FVIII ≤ or > 150 IU/Kg/week, respectively, within the first 8–12 weeks of therapy, regardless of the type of regimen (prophylaxis or on-demand). A cut-off of 150 IU/Kg/week was chosen since it corresponds to a usual high-dose prophylaxis regimen and is usually reached by four or more consecutive days of treatment.

• b) Putative confounders (i.e. variables potentially associated with both the exposures and inhibitor development)

F8 mutation type

We classified F8 genotype into high- and low-risk mutations. High-risk mutations were null mutations, such as large deletions, and nonsense mutations. Low-risk mutations were small deletions and insertions, missense mutations and splice site mutations. Since there is no definitive agreement on the classification of intron 1 and 22 inversions (13, 14), such defects were first coded as high-risk mutations; then a three-level variable for F8 was used with a separate category for intron inversions.
Family history of inhibitors
A three-level categorical variable was used for family history of inhibitors: "negative" for patients with a positive family history of haemophilia but negative family history of inhibitors; "not applicable/unknown" for those with a negative family history of haemophilia and those with a positive family history of haemophilia but missing information for family history of inhibitors; and "positive" for those with a positive family history (brother, cousin, uncle or grandfather) of inhibitors.

Other treatment-related variables
Type of treatment regimen (prophylaxis or on-demand) and major surgical procedure as the reason for first treatment were considered as separate variables.

Severity of haemophilia, patient year of birth and age at the first treatment were also included as potential confounders.

Subgroup and sensitivity analyses were performed considering i) only patients with severe haemophilia, and ii) only high-responder inhibitors.

Statistical analysis
Our main goal was to investigate the role of FVIII type and treatment intensity on inhibitor development. Three different statistical approaches were used to control for putative confounding due to the observational nature of the data, and the respective results were compared among them.

• Univariate and multivariable Cox regressions, the latter adjusting the association between the exposures of interest and inhibitor development for the effect of the abovementioned putative confounders. The number of EDs at inhibitor development was set as time-to-event.

• Propensity score (PS) analyses (15, 16), adjusting the association between the exposures of interest and inhibitor development for different chances (i.e. measured by the PS) of being treated in a certain way (confounding by indication). Two independent PSs (one for FVIII type and one for intensity) were calculated by two separate logistic models including the putative confounders (as expected to influence how the patient was treated). The two PSs were subsequently used a) as covariates in a multivariable Cox regression, and b) to calculate the "average effect of treatment on treated" (ATT). In brief, the ATT corresponds to the difference in the risk of inhibitors between pdFVIII-treated and rFVIII-treated patients, and between patients who received high- and low-intensity treatment, as estimated after matching or stratifying patients by PS. To explore the reciprocal-effect modification, the ATT related to FVIII type was calculated in subgroups defined by treatment intensity, and the ATT related to treatment intensity in subgroups defined by FVIII type.

• Classification and Regression Tree (CART) analysis. CART analysis is a semi-automatic non-parametric technique used to hierarchically select among a large number of variables those that better classify the population into patients with or without the outcome of interest (17). It allows for simultaneous exploration of different degrees of interaction (effect modification) between exposures and confounders, while adjusting for confounding.

Preliminary testing of the existence of a different distribution of FVIII type across subgroups defined by level of intensity, and vice versa, and of putative confounders across subgroups defined by FVIII type or by treatment intensity, was done using multilevel models to account for the clustering effect of study centre. All the subsequent analyses were stratified by centre. Discrimination of logistic and Cox models was measured by c-statistics (area under the receiver-operating characteristic curve [AUC]), which is a way of assessing discrimination. The higher the AUC, the better discrimination is. Discrimination refers to the ability of a putative predictor or a set of putative predictors to distinguish patients who developed from those who did not develop inhibitors. In Cox regression, missing data were handled by multiple imputation, impact of which was tested in a sensitivity analysis. STATA (version 12.0) was used for statistical analyses. The user-written commands psorescore and att* were used for PS (18).

Secondary analysis on type of FVIII products
The differential impact of FVIII molecules on inhibitor development was explored by univariate and multivariable Cox regressions. The following product categories were considered: purity of pdFVIII concentrates (high- or low-purity, defined as ≤0.01 or...
> 0.01 IU of Von Willebrand factor antigen per IU of FVIII antigen; rFVIII product formulation and molecule length (first-generation full-length, second-generation full-length, second-generation B-domain-deleted, or third-generation full-length products); rFVIII product cell line of development (Chinese hamster ovary cells or baby hamster kidney cells).

**Results**

**Study inclusion**

In addition to the 24 references retrieved for the aggregate data meta-analysis (9), five full articles (19–23) published after 2009 were judged to be eligible for this IPD meta-analysis. The authors of the 29 references were contacted, and patient data were provided for one Egyptian cohort (24) (100 PUPs), one German multi-centre cohort (25–28) (184 PUPs), one Israeli cohort (27–29) (100 PUPs) and one Italian multi-centre cohort (21) (377 PUPs), for a total of 10 different centres.

**Patient characteristics**

Seven hundred sixty-one (761) PUPs, born between 1935 and 2011 (median birth-year 1988) were included in the meta-analysis. Most patients (86.3 %) had severe HA. At first exposure to treatment, 229 patients (30.1 %) received rFVIII; among 700 patients with available data on initial weekly dose, 104 (14.9 %) received high-intensity treatment according to our definition (Table 2).

**Table 2: Patient characteristics in the total population and in by treatment-defined groups.**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total</th>
<th>Missing data, n (%)</th>
<th>rFVIII</th>
<th>pdfFVIII</th>
<th>P-value*</th>
<th>High intensity</th>
<th>Low intensity</th>
<th>P-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severe haemophilia, n (%)</td>
<td>657 (86.3)</td>
<td>0</td>
<td>191 (83.4)</td>
<td>466 (87.6)</td>
<td>0.187</td>
<td>87 (83.6)</td>
<td>514 (86.2)</td>
<td>0.810</td>
</tr>
<tr>
<td>High-risk mutation, n (%)</td>
<td>473 (73.2)</td>
<td>115 (15.1)</td>
<td>165 (77.8)</td>
<td>308 (71.0)</td>
<td>0.022</td>
<td>76 (80.8)</td>
<td>368 (73.0)</td>
<td>0.062</td>
</tr>
<tr>
<td>Family history of inhibitor, n (%):</td>
<td>192 (25.2)</td>
<td>514 (67.5)</td>
<td>55 (7.2)</td>
<td>50 (21.8)</td>
<td>164 (71.6)</td>
<td>15 (6.6)</td>
<td>142 (26.7)</td>
<td>350 (67.8)</td>
</tr>
<tr>
<td>No</td>
<td>0</td>
<td>12 (1.5)</td>
<td>165 (77.8)</td>
<td>308 (71.0)</td>
<td>0.022</td>
<td>76 (80.8)</td>
<td>368 (73.0)</td>
<td>0.062</td>
</tr>
<tr>
<td>NA/Unknown#</td>
<td>192 (25.2)</td>
<td>514 (67.5)</td>
<td>55 (7.2)</td>
<td>50 (21.8)</td>
<td>164 (71.6)</td>
<td>15 (6.6)</td>
<td>142 (26.7)</td>
<td>350 (67.8)</td>
</tr>
<tr>
<td>Median age at first exposure, y (Q1, Q3)</td>
<td>1.3 (0.6, 2.7)</td>
<td>28 (3.7)</td>
<td>0.9 (0.1–1.4)</td>
<td>1.6 (0.8–3.2)</td>
<td>&lt;0.001</td>
<td>0.9 (0.4–1.3)</td>
<td>1.4 (0.6–2.8)</td>
<td>0.001</td>
</tr>
<tr>
<td>Surgery at first exposure, n (%)</td>
<td>55 (7.2)</td>
<td>0</td>
<td>14 (6.1)</td>
<td>41 (7.7)</td>
<td>0.433</td>
<td>12 (11.5)</td>
<td>41 (6.9)</td>
<td>0.002</td>
</tr>
</tbody>
</table>

rFVIII, recombinant FVIII concentrates; pdfFVIII, plasma-derived FVIII concentrates. Q1, first quartile. Q3, third quartile. y, years. *p value for testing the existence of a difference in the distribution of patient characteristics between patients treated with pdfFVIII and those treated with rFVIII, or between patients treated at high and those treated at low intensity. #Not applicable (NA)/Unknown category includes both patients with a negative family history of haemophilia and patients with missing information for family history of inhibitors.

**Effect of FVIII type, treatment intensity and confounders on inhibitor development**

Overall, patients were followed up for a median of 150 EDs (first quartile [Q1], third quartile [Q3]: 65, 200). At a median exposure of 20 EDs (Q1, Q3: 12, 40), 208 patients developed inhibitors (27.3 %); 172 of the 740 patients with available data on peak titre developed a high-responding inhibitor (22.6 %). In patients first treated with rFVIII and those first treated with pdFVIII, 92 (40.2 %) and 116 (21.8 %), respectively, developed an inhibitor; 53 (51.0 %) and 143 (24.0 %) patients who received high-intensity and low-intensity treatment, respectively, developed an inhibitor. The same estimates are reported by country in Suppl. Table 1 (available online at www.thrombosis-online.com).

The results of the Cox regression analyses (analyses 1 and 2a) are reported in Table 3. In univariate and bivariable analyses, stratified by centre, both rFVIII and a high-intensity regimen significantly increased risk of inhibitor development.

In multivariable Cox regression, intensity of treatment, but not FVIII type, was confirmed as an independent risk factor for inhibitor development (Table 3). However, the type-by-intensity interaction term was statistically significant, suggesting that in patients receiving low-intensity treatment, rFVIII use was associated with a higher risk of inhibitor development; this effect was not
seen, or even reversed, in patients receiving high-intensity treatment (►Table 3).

A more recent birth year, positive family history of inhibitors and high-risk F8 mutation, together with treatment intensity, were associated with a higher inhibitor risk in the multivariable Cox model. The AUCs for the model including only FVIII type, only treatment intensity, or both, were 0.588, 0.550 and 0.621, respectively. The AUC for the model including as predictors all the putative confounders, but not FVIII type and treatment intensity, was 0.749 (0.665 when only family history of inhibitors was included). The AUC for the full model including all variables was 0.802.

Cox analyses without multiple imputation of missing data produced similar results.

Adjusting the Cox model for the PS (analysis 2a) produced overlapping results (►Table 3). The calculation of the ATT (analysis 2b) led to slightly different results in terms of statistical significance (►Table 4). When all patients were considered, the effect of treatment intensity was stably significant across four different matching/stratifying approaches; the effect of FVIII type was statistically significant in one out of four approaches. The higher risk of inhibitors with rFVIII vs pdFVIII in patients receiving low-intensity treatment, and lower risk with rFVIII vs pdFVIII in those receiving high-intensity treatment, was confirmed in PS analyses (►Table 4).

The results for CART analysis (analysis 3) are graphically displayed in ►Figure 1 and confirmed the pivotal role of treatment intensity, which resulted in the most efficient criteria for splitting the population according to risk of inhibitors. The CART analysis also identified a secular trend in the risk of inhibitors, which was significantly higher for patients treated more recently.

The effects of FVIII type and treatment intensity among severe patients and/or considering only high-responding inhibitors were similar to what was found in the main analyses.

Effect of different types of FVIII products

See ►Table 5 for results.

- In univariate analysis, all types of rFVIII were significantly associated with a higher risk of inhibitors compared to pdFVIII. The increase in risk disappeared after adjusting for confounders.
- pdFVIIIs with different degrees of purity (based on content of Von Willebrand factor) did not differ for association with inhibitors in both univariate and multivariable analyses.

| Table 3: Inhibitor development according to FVIII type and treatment intensity: by-centre stratified Cox regression with covariate and propensity score adjustment. |
|---------------------------------|-----------------|----------|-----------------|----------|
| **Cox regression**              | rFVIII vs pdFVIII | **P-value** | High-intensity vs low-intensity | **P-value** |
| Univariate                      | 2.2 (1.6–2.9)   | <0.001   | 2.9 (2.0–4.2)   | <0.001   |
| Bivariable*                     | 1.9 (1.4–2.6)   | <0.001   | 2.4 (1.6–3.5)   | <0.001   |
| Bivariable* with interaction | If high-intensity: 1.0 (0.6–1.7) | 0.970 | If rFVIII: 1.6 (1.0–2.6) | 0.048 |
|                                  | If low-intensity: 2.5 (1.8–3.6) | <0.001 | If pdFVIII: 4.1 (2.4–6.9) | <0.001 |
| Multivariable (covariate-adjusted) | 1.3 (0.8–2.0) | 0.268 | 1.9 (1.3–2.8) | 0.002 |
| Multivariable (covariate-adjusted) | If high-intensity: 0.7 (0.4–1.3) | 0.264 | If rFVIII: 1.3 (0.8–2.0) | 0.349 |
|                                  | If low-intensity: 1.8 (1.1–3.0) | 0.028 | If pdFVIII: 3.3 (1.9–5.7) | <0.001 |
| Multivariable (PS-adjusted‡)     | 1.1 (0.6–1.9) | 0.813 | 2.1 (1.4–3.3) | 0.001 |
| Multivariable (PS-adjusted‡) with interaction | If high-intensity: 0.5 (0.2–1.0) | 0.049 | If rFVIII: 1.4 (0.8–2.3) | 0.213 |
|                                  | If low-intensity: 1.7 (0.9–3.1) | 0.098 | If pdFVIII: 4.7 (2.6–8.5) | <0.001 |

rFVIII, recombinant FVIII concentrates. pdFVIII, plasma-derived FVIII concentrates. HR, Hazard Ratio. CI, Confidence Interval. PS, propensity score. * Including only the variables for FVIII type and treatment intensity. ‡ Interaction term between FVIII type and treatment intensity. † Variables included in the two propensity scores (for FVIII type and intensity) were the same used as covariates in multivariable Cox regressions (birth year, disease severity, gene mutation, familiar history of inhibitor, age at the first treatment, on-demand/prophylaxis regimen, surgery, country) plus intensity in the PS for FVIII source, and vice versa. C statistics for propensity score for FVIII type was 0.952 and 0.847 for propensity score for treatment intensity.
Marcucci et al. Treatment-related risk of inhibitors in haemophilia

- In patients receiving rFVIII, univariate analysis showed no significant difference among different types of rFVIII; multivariable analysis showed that third-generation full-length rFVIII was associated with a lower risk of inhibitors than first-generation full-length and second-generation B-domain–deleted rFVIII; no difference was found between third- and second-generation full-length rFVIII.
- No effect of type of cell line was found.

Discussion

The main result of our analysis is that unadjusted assessment of treatment-related-risk factors for inhibitor development is subject to significant confounding. Our conclusion is based on robust evidence generated by comparing multiple complementary statistical approaches applied to a large IPD dataset of PUPs from 10 centres in four different countries. Specifically, the intensity of treatment was consistently found to be an independent risk factor across all the analyses performed, whereas the effect of type of FVIII appeared to be confounded by other patient-related risk factors. Furthermore, calculation of the AUC underlined that variables other than type of concentrate and intensity, together, distinguished between patients at high and low risk of inhibitors better than type of concentrate or intensity alone. When adjusting for all other patient-level factors, treatment-related factors added only a little in terms of discrimination.

The increase in the risk of inhibitors with high-intensity treatment was larger in patients treated with pdFVIII, while rFVIII showed a higher risk of inhibitors only in the absence of high-intensity treatment. Unlike most published evidence, rFVIII appeared to be associated with a lower risk than pdFVIII in patients treated with intensive regimens. As with any other non-replicated research finding, this last finding might be “real” (and based on some still unexplained biological mechanism) or it might be due to residual confounding.

The analysis of the difference in inhibitor risk between FVIII brands was not prespecified and was added after publication of three recent published reports (4, 30, 31). In our study population, no differences were found between pdFVIII and different categories of rFVIII, pdFVIII with different Von Willebrand factor content, or different cell lines used to produce rFVIII. Interestingly, we confirmed a low inhibitor risk in patients treated with third-generation full-length rFVIII; ironically, the risk associated with third-generation full-length rFVIII was significantly lower compared to all other molecules, except for second-generation full-length rFVIII, which had borderline statistical significance. We recommend cautious appraisal of these findings. Indeed, the

<table>
<thead>
<tr>
<th>Type of matching#</th>
<th>rFVIII vs pdFVIII ATT (standard error)</th>
<th>High-intensity vs Low-intensity ATT (standard error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS matching (nearest neighbour method)</td>
<td>All: 0.020 (0.147)</td>
<td>All: 0.233 (0.095)**</td>
</tr>
<tr>
<td></td>
<td>High-int: –0.358 (0.299)</td>
<td>rFVIII: 0.192 (0.114)</td>
</tr>
<tr>
<td></td>
<td>Low-int: 0.304 (0.125)**</td>
<td>pdFVIII: 0.500 (0.133)**</td>
</tr>
<tr>
<td>PS matching (radius method)</td>
<td>All: 0.169 (0.049)**</td>
<td>All: 0.296 (0.057)**</td>
</tr>
<tr>
<td></td>
<td>High-int: –0.201 (0.178)</td>
<td>rFVIII: 0.116 (0.082)</td>
</tr>
<tr>
<td></td>
<td>Low-int: 0.212 (0.052)**</td>
<td>pdFVIII: 0.385 (0.086)**</td>
</tr>
<tr>
<td>PS matching (kern-based method)</td>
<td>All: 0.091 (0.118)</td>
<td>All: 0.302 (0.066)**</td>
</tr>
<tr>
<td></td>
<td>High-int: –0.364 (0.156)***</td>
<td>rFVIII: 0.197 (0.102)#</td>
</tr>
<tr>
<td></td>
<td>Low-int: 0.274 (0.055)**</td>
<td>pdFVIII: 0.486 (0.089)**</td>
</tr>
<tr>
<td>PS stratification</td>
<td>All: 0.088 (0.143)</td>
<td>All: 0.272 (0.075)**</td>
</tr>
<tr>
<td></td>
<td>High-int: –0.341 (0.204)</td>
<td>rFVIII: 0.145 (0.090)**</td>
</tr>
<tr>
<td></td>
<td>Low-int: 0.288 (0.048)**</td>
<td>pdFVIII: 0.482 (0.108)**</td>
</tr>
</tbody>
</table>

rFVIII, recombinant FVIII concentrates. pdFVIII, plasma-derived FVIII concentrates. High-int, high-intensity. Low-int, low-intensity. PS, Propensity Score. ATT, Average effect of Treatment on Treated (calculated as difference in inhibitor development between rFVIII and pdFVIII, or between high- and low-intensity treatment; i.e. a negative ATT corresponds to, respectively, a higher risk with pdFVIII and with a low-intensity treatment). *Difference between rFVIII and pdFVIII, or between high- and low-intensity treatment, significantly different from 0. #The nearest neighbor is the basic PS matching technique. Variants of this basic method were developed to improve the quality of matching, although reducing the number of subjects used to obtain the matched groups or strata (18). **Difference between rFVIII and pdFVIII, or between high- and low-intensity treatment.
small size of the samples in our and others’ (4, 30, 31) subgroup analyses, as well as the partial overlap and overall lack of consistency between the findings of the four studies, suggests a possible effect of chance and the need for further analyses before the issue can be considered clarified.

The value of the present analysis relies in the sample size of the full cohort, which is one of the largest ever analysed. Moreover, only a limited proportion of patients in our cohort had been included in the RODIN study (4) (49 of 761, i.e. 6% of patients), minimizing the overlap of data and enhancing the novelty of our results. A strength of our methods is the adjustment for the clustering effect of centres and the consistency of the results across three different analytical approaches based on different complementary assumptions. In particular, in addition to the more common multivariable Cox regression, we used for the first time PS (15, 16) and CART (17) analyses.

PS analysis aims at maximizing the comparison of “like” groups, differing “only” for the exposure of interest, when this is not guaranteed by randomisation. In our analysis, for example, previous knowledge of the role of non-modifiable risk factors and the dispute regarding the possible effect of type of FVIII concentrate on inhibitor development might have influenced physician selection of a pdFVIII or a rFVIII product for patients judged to be at different risk of inhibitor development. PS analysis aims to quantify the probability of receiving a certain treatment (expressed as a scalar summary with values from 0 to 1), conditional on a set of known pre-treatment variables, and uses it to adjust the analysis. Adjusting a regression model (e.g. a Cox model) for the PS rather than adjusting for every single putative confounder has the advantage of saving degrees of freedom. Another way to adjust for PS is to match patients on (or stratify by) the PS. The ATT routine allowed us to use this type of adjustment and strengthen our confidence in our results, even though the ATT is an estimate of the as-

Figure 1. Classification and Regression Tree (CART) analysis including all the variables. The CART algorithm selects the variables that most significantly differ between patients who developed or not inhibitors. They are selected in a hierarchically way and choosing, for continuous, discrete and categorical variables, the most discriminative cut-point or category. The selected variables are displayed as splitting knots from which one branch springs up for each value/category of the variable associated with a different probability of developing inhibitors; for each terminal branch (a branch not followed by any knot) a measure of the relative risk of inhibitors is provided. In this way CART generates subgroups at different risk of inhibitors. In particular, the CART figure suggests that while among high risk patients receiving a high-intensity treatment the type of concentrate was not associated with any differential effect on risk, among patients treated at low intensity, rFVIII was associated with a higher inhibitor risk than pdFVIII (shifting the patients from being at low to being at moderate risk of inhibitors) but only among the subgroup (127 patients) born (and likely treated) earlier in secular time, with a negative family history for inhibitors, treated on demand and with a high risk F8 mutation. rFVIII, recombinant FVIII concentrates; pdFVIII, plasma-derived FVIII concentrates; N, number; F, failure; RHR, Relative Hazard Ratio (where a RHR=1 represents a hypothetical point of indifference for inhibitor development/equipoise and corresponds to the HR for a status where all the predictors take on their mean value)
sociation between exposures and inhibitors that does not take into
time to event.

CART analysis is a best-split iterative method widely used in
data mining for large databases (17). CART has an interesting ex-
ploratory potential in that it is almost completely computer-based
and automatic, and thus free from researcher-driven biases. En-
compassing both covariate adjustment and interactions, it ac-
counts for confounding and effect modification in a hierarchical
and “economic” way, respectively. In brief, CART findings would
have been equivalent to the findings of Cox or PS analyses in terms
of confounding; i.e. when accounting for other patient-level vari-
ables, type of concentrate was not an independent risk factor of in-
hibitors. The added value is the inclusion of every interaction be-
tween putative predictors (not only between the two exposures of
interest), something that would have made the Cox or ATT ana-
yses difficult to interpret or technically unfeasible. For type of
concentrate, the inclusion of every interaction enables us to see if
the type of concentrate could play a secondary role in any of the
subgroups defined by the other putative predictors, and presented
according to a hierarchy of their importance in terms of distin-
guishing between high and low risk (another advantage in terms of
immediacy of message of CART and its graphical output). In this
way, CART analysis suggests which groups to examine. The limi-
tation is that CART analysis works best with large data sets. Des-
pite our large sample size, CART generated very small subgroups;
thus, caution is needed when interpreting and transferring these
subgroup findings into clinical practice. However, for researchers
trying to interpret the paucity of evidence in order to drive further

Table 5: Unadjusted and adjusted Cox regression for the effect on inhibitor development of different categories of FVIII concentrates.†.

<table>
<thead>
<tr>
<th></th>
<th>n patients (n inhibitors)</th>
<th>Unadjusted Hazard Ratio (95 % CI)</th>
<th>P-value</th>
<th>Adjusted Hazard Ratio (95 % CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Different rFVIII vs pdFVIII</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pdFVIII (reference)</td>
<td>392 (65)</td>
<td>‐</td>
<td>‐</td>
<td>‐</td>
<td>‐</td>
</tr>
<tr>
<td>rFVIII – first generation, full length*</td>
<td>115 (37)</td>
<td>1.7 (1.1–2.6)</td>
<td>0.009</td>
<td>1.0 (0.6–1.7)</td>
<td>0.914</td>
</tr>
<tr>
<td>rFVIII – second generation, full length**</td>
<td>60 (28)</td>
<td>2.6 (1.6–4.2)</td>
<td>&lt;0.001</td>
<td>1.4 (0.7–2.7)</td>
<td>0.298</td>
</tr>
<tr>
<td>rFVIII – second generation, B domain deleted***</td>
<td>32 (17)</td>
<td>3.2 (1.9–5.6)</td>
<td>&lt;0.001</td>
<td>1.7 (0.9–3.3)</td>
<td>0.107</td>
</tr>
<tr>
<td>rFVIII – third generation, full length****</td>
<td>20 (8)</td>
<td>2.4 (1.2–5.1)</td>
<td>0.018</td>
<td>0.6 (0.2–1.6)</td>
<td>0.341</td>
</tr>
<tr>
<td>Different pdFVIII (purity)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pdFVIII – high purity (reference)</td>
<td>142 (28)</td>
<td>‐</td>
<td>‐</td>
<td>‐</td>
<td>‐</td>
</tr>
<tr>
<td>pdFVIII – low purity</td>
<td>250 (47)</td>
<td>0.8 (0.5–1.5)</td>
<td>0.591</td>
<td>0.9 (0.5–1.9)</td>
<td>0.894</td>
</tr>
<tr>
<td>Different rFVIII (formulation and gene length)</td>
<td>227 (90)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rFVIII – first generation, full length* (reference)</td>
<td>115 (37)</td>
<td>‐</td>
<td>‐</td>
<td>‐</td>
<td>‐</td>
</tr>
<tr>
<td>rFVIII – second generation, full length**</td>
<td>60 (28)</td>
<td>1.3 (0.7–2.3)</td>
<td>0.443</td>
<td>0.5 (0.2–1.2)</td>
<td>0.120</td>
</tr>
<tr>
<td>rFVIII – second generation, B domain deleted***</td>
<td>32 (17)</td>
<td>1.6 (0.9–2.9)</td>
<td>0.135</td>
<td>1.2 (0.6–2.2)</td>
<td>0.656</td>
</tr>
<tr>
<td>rFVIII – third generation, full length****</td>
<td>20 (8)</td>
<td>1.2 (0.6–2.7)</td>
<td>0.573</td>
<td>0.3 (0.1–0.7)</td>
<td>0.008</td>
</tr>
<tr>
<td>Different rFVIII (cell lines)</td>
<td>227 (90)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rFVIII – baby hamster kidney cells (reference)</td>
<td>104 (44)</td>
<td>‐</td>
<td>‐</td>
<td>‐</td>
<td>‐</td>
</tr>
<tr>
<td>rFVIII – Chinese hamster ovary cells</td>
<td>123 (46)</td>
<td>1.2 (0.7–1.9)</td>
<td>0.440</td>
<td>0.7 (0.3–1.5)</td>
<td>0.361</td>
</tr>
</tbody>
</table>

n, number. CI, confidence interval. rFVIII, recombinant FVIII concentrates. pdFVIII, plasma-derived FVIII concentrates. † Missing data for variables included in the multivariable analyses, other than the one defining the FVIII product classification of interest, were managed by a multiple imputation technique. * Kogenate, Helixate, Recombinate. ** Kogenate FS, Helixate NextGen. *** Refacto. **** Advate. ‡ When the rFVIII-third generation (full length) was taken as reference, no significant differences were found in the non-adjusted analysis, while in the multivariable analysis, the third generation was associated with a lower risk compared to all the other categories (except the rFVIII-second generation full length) in a statistically significant way. Compared which the rFVIII-second generation full length, the rFVIII-third generation tended to be associated with a lower risk but without reaching a statistical significance (p=1.34).
research, our synthesis provides some interesting hints about the role of FVIII type in some relevant subgroups that might be investigated in the design and analysis of future studies.

We also recognise some weaknesses of our study, including the inability of exploring the effect of unknown confounders or known confounders for which data were completely missing; absence of a centralised testing laboratory; inability of separately analysing inhibitors with titre s between 0.6 and 1 BU; absence of a standardised protocol for treatment and for inhibitor testing; wide timeframe of the included data; inclusion of analyses that were not defined a priori; low power for subgroup analyses; and lack of data on predictors and treatment that vary over time, including switching to a different concentrate. The impact of this last assumption was mitigated by the overall reluctance of physicians to switch before the 50th ED (32). Furthermore, the unavailability of individual data on the frequency of inhibitor testing prevented us from investigating whether the important confounding effect of the birth year could be explained by lower frequency and accuracy in seeking inhibitors in patients treated in less recent time periods. Another aspect of our analysis that could be seen as a limitation is the choice of defining treatment intensity as a threshold value in terms of units/kg/week, without accounting for whether days of treatment were consecutive or not. Our decision was initially based on the lack of explicit information on consecutiveness of treatment across the source databases. It became intentional, with the support of clinical and laboratory immunologists, in an effort to move away from the raw concept of EDs toward the more flexible one of concentration-over-time area. Despite these limitations, the large sample size and thorough methodological approach that tested across different assumptions led to consistent results and allowed us to be confident in our findings. Our open acknowledgement and discussion of the limitations, some of which are not unique to this study (e.g. unknown confounding can affect any observational research), aims at the accurate interpretation and clinical use of the present evidence. We also hope to foster improvements in observational research (e.g. improving the completeness of data collection and consistency of standards and protocols) in the context of conditions such as haemophilia, in which observational research plays a fundamental role, and systematic reviews and meta-analyses can expand our analytical power and understanding. Finally, we are aware that this individual patient data meta-analysis did not include all relevant studies since we only pooled data that were made available to us. Although we hope to increase the number of patients included in our analysis, we felt it was important to communicate the results of the analysis of the first 761 patients to contribute to decision making, promote collaboration between haemophilia researchers, and encourage authors to share individual patient data on PUPs.

Until then, the message for clinicians is clear: unlike the finding of an increased risk of inhibitor development associated with higher-intensity treatment, evidence for an increased risk associated with type of concentrate is too weak to prompt any unequivocal clinical decisions. Type of concentrate might have a more relevant role in some subgroups, but consistent findings on such a role are not yet available. In this light, the efforts of designing and performing a controlled randomized clinical trial (i.e. the SIPPET trial) will be invaluable in refuting or confirming equipoise between different F VIII types.

Acknowledgements
The project was endorsed by the European Association for Haemophilia and Allied Disorders (EAHAD) and was coordinated by a working party of the EAHAD chaired by the two senior authors, and was partly supported by EAHAD operating funds.

Authorship contributions
M.M., M.E.M., E.S., A.L., U.N-G., and G.K. were responsible for the study concept and design. M.M. and A.I. performed the statistical analyses and wrote the initial manuscript. G.K., M.E., C.B., C.E.E., U.N-G. and S.H. were responsible for source patient recruitment. All authors contributed to the interpretation of data and critically revised the report for important intellectual content.

Conflicts of interest
None declared.

References
Marcucci et al. Treatment-related risk of inhibitors in haemophilia


