Citation: Clin Transl Sci (2020) 13, 1127–1136; doi:10.1111/cts.12794

ARTICLE

Annual Bleeding Rates: Pitfalls of Clinical Trial Outcomes in Hemophilia Patients

Christine Keipert[1,*,†], Mirco Müller-Olling[1,†], Franca Gauly[1], Cornelia Arras-Reiter[1] and Anneliese Hilger[1]

Emerging treatment options for hemophilia, including gene therapy, modified factor products, antibody-based products, and other nonreplacement therapies, are in development or on their way to marketing authorization. For proof of efficacy, annual bleeding rates (ABRs) have become an increasingly important endpoint in hemophilia trials. We hypothesized that ABR analyses differ substantially between and within medicinal product classes and that the ABR observation period constitutes a major bias. For ABR characterization, an internal factor VIII (FVIII) treatment database has been built based on confidential clinical trial data submitted to the Paul-Ehrlich-Institut (PEI). Furthermore, anonymized data from 46 trial protocols submitted for review to the PEI were analyzed (FVIII replacement, *n* = 27; antibody-based, *n* = 12; and gene therapy, *n* = 7) for methodology. Definitions of bleeding episodes and ABR observational periods differed substantially in clinical trials. In the initial observation phase, individual ABRs of patients, treated prophylactically for 1 year, vary by about 40% (*P* < 0.001), which finally led to a significant reduction of the ABR group mean by 20% (*P* < 0.05). Furthermore, the high variance in ABRs constitutes a major challenge in statistical analyses. In conclusion, considerable heterogeneity and bias in the ABR estimation in clinical trials was identified, which makes it substantially more difficult to compare the efficacy of different treatment regimens and products. Thus, awareness of the important pitfalls when using ABR as a clinical outcome is needed in the evaluation of hemophilia therapies for patients, physicians, regulators, and health technology assessment agencies.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
☑ Annual bleeding rates (ABRs) become increasingly important as comparative clinical endpoints in replacement, nonreplacement, and gene therapy trials in patients with hemophilia A and B. Methodological and clinical use of this measure is debated controversially.

WHAT QUESTION DID THIS STUDY ADDRESS?
☑ What is the significance of ABRs in hemophilia clinical trials?

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?
☑ Differences in observational period and bleeding definitions are the most critical parameters for ABR estimation and compromise the direct comparison of different treatment regimens or products.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?
☑ Qualified evaluations of hemophilia treatment strategies are of tremendous importance for patients, physicians, regulators, and health technology assessment agencies. For future CTs the following is recommended: use of agreed-upon bleeding definitions and physician-based monitoring of bleedings, observational periods should cover a minimum of 12 months, appropriate use of statistical models, eligible patients should have a clinically severe phenotype, and use of imaging techniques where appropriate.

Hemophilia is an X-linked rare bleeding disorder that is characterized by a deficiency of functional coagulation factor VIII (FVIII) or IX and can be categorized based on endogenous factor activity levels as severe (< 1% activity), moderate (1–5% activity), and mild (> 5–40% activity). Individuals with severe hemophilia experience frequent bleeding episodes (BEs) either spontaneously or following minor trauma, which can be acutely life-threatening or lead to debilitating long-term complications. For example, joint, muscle, mucosal, and gastrointestinal tract bleeding, and most severely, intracranial hemorrhage can result in disability and death. Current treatment of severe hemophilia mainly relies on replacement therapy with plasma-derived or modified recombinant factor concentrates.

New hemophilia treatment options are in development or have been approved recently, including gene therapy, bispecific monoclonal antibodies, anti-tissue factor pathway inhibitor antibodies, and other nonreplacement

†These authors contributed equally to this work.
[1]Hematology and Transfusion Medicine, Paul-Ehrlich-Institut, Federal Institute for Vaccines and Biomedicines, Langen, Germany. *Correspondence: Christine Keipert (christine.keipert@pei.de)
Received: March 10, 2020; accepted: March 13, 2020. doi:10.1111/cts.12794

therapies.[1–8] This is reflected by a large number of ongoing clinical trials (CTs) in this field. In fact, a search in the ClinicalTrials.gov database in June 2019 of phase I–III-declared studies in congenital hemophilia yielded a total of 69 CTs comprising factor-based (*n* = 26), gene therapy-based (*n* = 23, including one trial referring to genome editing), antibody-based (*n* = 12), RNAi-based (*n* = 6), and stem cell-based (*n* = 2) products. Importantly, these approaches intervene in different parts of the coagulation cascade and solely coagulation factor levels do not necessarily reflect therapeutic efficacy.

Estimation of the annualized bleeding rate, also referred to as annual bleeding rate (ABR), has been introduced early as an efficacy variable for prophylactic replacement therapies in order to complement measures of FVIII or FIX trough levels. However, in contemporary CTs, ABRs are increasingly used as comparative and main outcome parameters.

Estimation of bleeding rates has intricate challenges and depends on numerous patient-related and external factors, including individual clotting factor level, pharmacokinetic profile and pain perception, the subject's age, health status, activity level, dosing regimen, BE definition, time to follow-up, and number of patients analyzed. ABR estimation is prone to subjective assessment, as patients as well as treating physicians have to define each bleed. This issue was also demonstrated in a musculoskeletal ultrasound study, which showed that pain perception as well as swelling and warmth is unreliable for bleed detection, resulting in substantial false-positive and false-negative bleeding rates.[9]

Typically, mean total ABRs are in the low to mid-single-digit range, whereas specific ABRs, such as the annual joint bleed rate, are in the low single-digit range.[10] It has been demonstrated that there is a substantial range of bleeding frequencies among patients with similar clotting factor levels, confirming the ABR as a more personalized parameter. In addition, there is ongoing discussion about the optimal outcome measure and suitability of ABR as an efficacy measure in patients with hemophilia with and without inhibitors.[11–15] In the European Medicine Agency (EMA) guidelines on core summary of product characteristics for human plasma derived and recombinant coagulation factor FVIII and FIX products, it is stated that ABR is not comparable between different factor concentrates and between different clinical studies.[16,17] This statement has been introduced empirically based on the long-standing experience in the regulation of hemophilia therapeutics, however, there is lack of supportive and published evidence.

We hypothesize that ABR analyses in CTs differ substantially and that the ABR observation period constitutes a major bias. For this approach, we constituted an internal database of confidential FVIII CT data at the Paul-Ehrlich-Institut (PEI) to determine basic characteristics of the ABR endpoint. In addition, we analyzed study protocols from contemporary hemophilia CTs comprising replacement and nonreplacement products as well as gene therapies to characterize differences in the methodology of ABR estimation. The results of this study should facilitate guidance on the minimum standards for bleeding rate estimation in CTs of rare bleeding disorders.

METHODS

PEI-database analysis

CT data, generated for marketing authorization of FVIII products, was submitted by the marketing authorization holders to the PEI as concerned regulatory authority. This confidential CT data is entered by the PEI into an internal FVIII treatment database (PEI-DB) encompassing > 1,300 patients with almost 85,000 FVIII concentrate infusions. For every single infusion, the following parameters were extracted: infusion date, reason of infusion, dose in international units (IU)/kg bodyweight and total IU. Reasons for infusion were categorized as bleed, follow-up, prophylaxis, study related, surgery, or other. The infused patients participated in 30 different CTs performed between 1986 and 2012. Eight of these CTs were performed in previously untreated patients and the remaining 22 in previously treated patients. CTs were included in the PEI-DB in the order of their first identification in the archives. For analysis, patients on prophylactic treatment were defined as those who received at least 2 prophylaxis infusions per week for a minimum of 6 months (prophylaxis group), and on-demand patients were defined as patients who were treated prophylactically no more than 2 days per month for a minimum of 6 months (on-demand group). Patients not matching the criteria of the prophylaxis or on-demand group were included in the diverse group (data collection of < 6 months, surgical trials, or CTs where patients were switched between on-demand and prophylactic treatments). For analysis of the ABR observational periods, all patients on prophylactic treatment that had documented infusions for a period of 1 year or more were selected (1-year prophylaxis group (1yP group)). For illustration of ABR changes, grouping of 10 patients per panel was considered the maximum in order to depict the complete dataset, and patients with ABR values in the low single-digit range were evenly distributed among the panels to minimize clustering. For seasonal variance analysis, all bleedings observed during 1 year (*n* = 571) were grouped by season (spring: March 21 to June 20; summer: June 21 to September 22; fall: September 23 to December 21; and winter: December 22 to March 20).

Study protocol analysis

A total of 46 CT protocols submitted to PEI for review was analyzed. Study protocols from CTs evaluating the efficacy and safety of FVIII replacement therapy in hemophilia A (HA), submitted for CT approval from 2008 to 2018, and declared as phase II/III or phase III studies (*n* = 27), were analyzed with regard to specifications of ABR and the underlying definitions for BEs. In comparison, protocols of phase II (excluding phase IIa trials), phase III and non-interventional studies related to antibody-based products (*n* = 12, 2015–2018) and protocols of phase I/II and phase III trials evaluating gene therapy-based products (*n* = 7, 2014–2019) in the HA and hemophilia B (HB) populations were analyzed. Four of 12 study protocols concerning antibody-based products were submitted to PEI for scientific advice procedure. The following dataset was collected by two independent reviewers throughout all study protocols: year of approval, overall study duration, duration of subject participation, enrollment number, ABR (availability, end

point, observation period, and subsets), definition of BE (location, cause, severity, treatment requirement, new bleed, single/multiple bleeds, pain, aura/sensation, and swelling/warmth), planned statistical comparison, and use of statistical models.

Computational and statistical analysis

Data analysis and graph plotting were performed using Microsoft Excel 2010 (Microsoft, Redmond, WA), GraphPad Prism version 7.04 (GraphPad Software) and nQuery Sample Size Software 8.2 (Statsols).

RESULTS

Characteristics of the FVIII PEI-DB

There were 1,309 patients included in the PEI-DB with 84,342 documented FVIII-product infusions. Age at study start was known for 1,212 patients: mean and median ages of the patients were 23 years each, maximum age was 73.1 years, and minimum age < 1 month. Severity of HA was documented for 1,291 patients: 92.6% were classified as severe, 4.8% as moderate, and 2.5% as mild. The prophylaxis group included a total of 204 patients with severe HA and the mean age was 26.7 years (median age 25 years; range 0.04–65 years). Ninety patients were included in the PEI-DB 1yP group. Sixty of these patients were already identified by the methods described above. Thirty patients were selected by hand from the pool of patients who had received on-demand as well as prophylactic treatment while they participated in CTs and had documented prophylactic treatment for more than 1 year. All 90 patients were classified as patients with severe HA, with a mean age of 14.5 years (median 1.5 years; range 0.02–53 years). The on-demand group included a total of 413 patients, mean age was 13.4 years (median 1.5; range 0–64 years). The patient distribution within the analysis groups is depicted in **Figure 1a**.

ABR – Analysis of distribution, observation period, and statistical comparison

The ABR range and relative frequency distribution of the prophylactic and on-demand groups are shown in **Figure 1b,c**. ABR data are generally characterized by count data having a right-skewed distribution with a high proportion of zeros and a median smaller than the mean, in line with our dataset. Both treatment groups have a considerable dispersion with SDs surpassing their mean. Mean ABR (range) was 4.2 (0–28) for the prophylaxis group (n = 204) and 18.7 (0–160) for the on-demand group (n = 413).

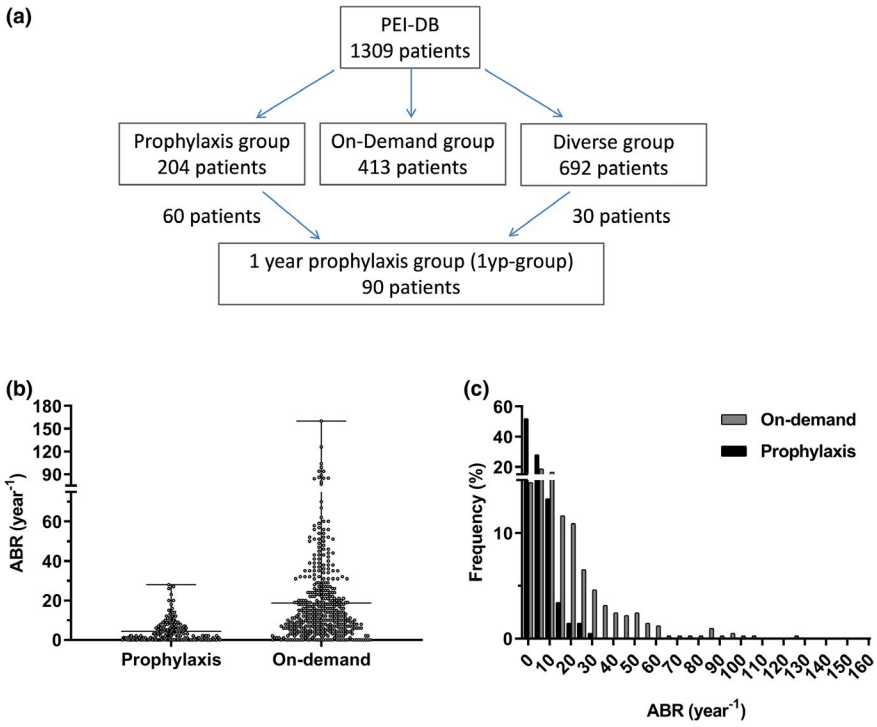

Figure 1 The Paul-Ehrlich-Institut internal FVIII treatment database (PEI-DB). Analysis groups and annual bleeding rate (ABR) distribution. (**a**) There were 1,309 patients included in the PEI-DB, of which 204 patients fulfilled the criteria for definition as patients on prophylactic treatment (minimum of 2 infusions/week for > 6 month) and 413 were classified as patients with on-demand treatment (maximum of 2 prophylactic infusions/month for > 6 months). A majority of 692 patients was categorized in the diverse group (data collection of < 6 months, surgical trials, or clinical trials where patients were switched between on-demand and prophylactic treatments). Of those, 30 patients were identified who switched regimens between on-demand and prophylactic treatment but finally were on prophylactic treatment for > 1 year. Together with 60 patients from the prophylactic group, they were joined to form the group of 90 patients who were on prophylactic treatment for > 1 year (1yP group), (**b**) ABR mean and range of the prophylaxis (n = 204) and on-demand group (n = 413), (c) relative frequency distribution. Segmented axes were used in (**b**) and (**c**) to plot the prophylaxis and on-demand groups on the same scale.

Inconsistencies in the observation period pose difficulties for frequency analyses. Indeed, the variance of ABR observation periods in CTs is large. To analyze the potential impact of different observation periods on the ABR estimate, an analysis of the individual ABRs of the PEI-DB 1yP group (n = 90; **Figure 2**) was carried out. The mean relative individual ABR change, calculated from absolute values of the individual changes irrespective of whether it is a positive or negative ABR shift, was 43.5% when the observation period was extended from 3 to 6 months (**Figure 3a**). This variance decreases significantly for the 9-month and 12-month ABRs ($P < 0.001$), with averaged individual ABR changes of 30.2% and 20.8%, respectively. Quantification of the overall effect of individual changes on the ABR mean of the 1yP group (**Figure 3a**) showed that the 6-month ABR mean was significantly decreased when compared with the 3-month ABR (–17%; $P < 0.05$). This was calculated from the signed values of the individual changes and, thus, takes into account the direction of the individual ABR change, which could be negative or positive. Further ABR group mean reduction was barely visible for the 9-month and 12-month ABR means. ABR means (range) in the 1yP group were as follows: $ABR_{3\text{-months}}$ = 4.0 (0–24), $ABR_{6\text{-months}}$ = 3.3 (0–30), $ABR_{9\text{-months}}$ = 3.1 (0–26), and $ABR_{12\text{-months}}$ = 3.2 (0–23) per year, respectively.

In-depth analysis of the 3-month to 6-month ABR reduction (**Figure 3b**) showed that decreasing ABRs occurred in the majority of patients with a generally higher 3-month ABR. When grouped by their 3-month ABR, the proportion of patients showing an $ABR_{3\text{-6-month}}$ reduction was 88.9%, 71.4%, and 83.3% in the ABR groups 5–10 years, 10–15 years, and > 15/years, respectively, which was significantly more when compared with the patients group with an ABR < 5 per year ($P < 0.001$).

To assess the contribution that differences in the treatment regimen may exert on the ABRs in the 1yP group, the FVIII consumption was analyzed. In result, the annual FVIII consumption (IU/kg) was similar for patients with ABR = 0/year (median: 4,084; range 1,630–7,857), ABR = 1–3/year (median: 4,380; range 1,492–12,889), ABR = 4–6/year (median: 4,455; range 2,470–9,918), and ABR > 6/year (median: 3,868; range 2,607–11,651). We further analyzed the extent

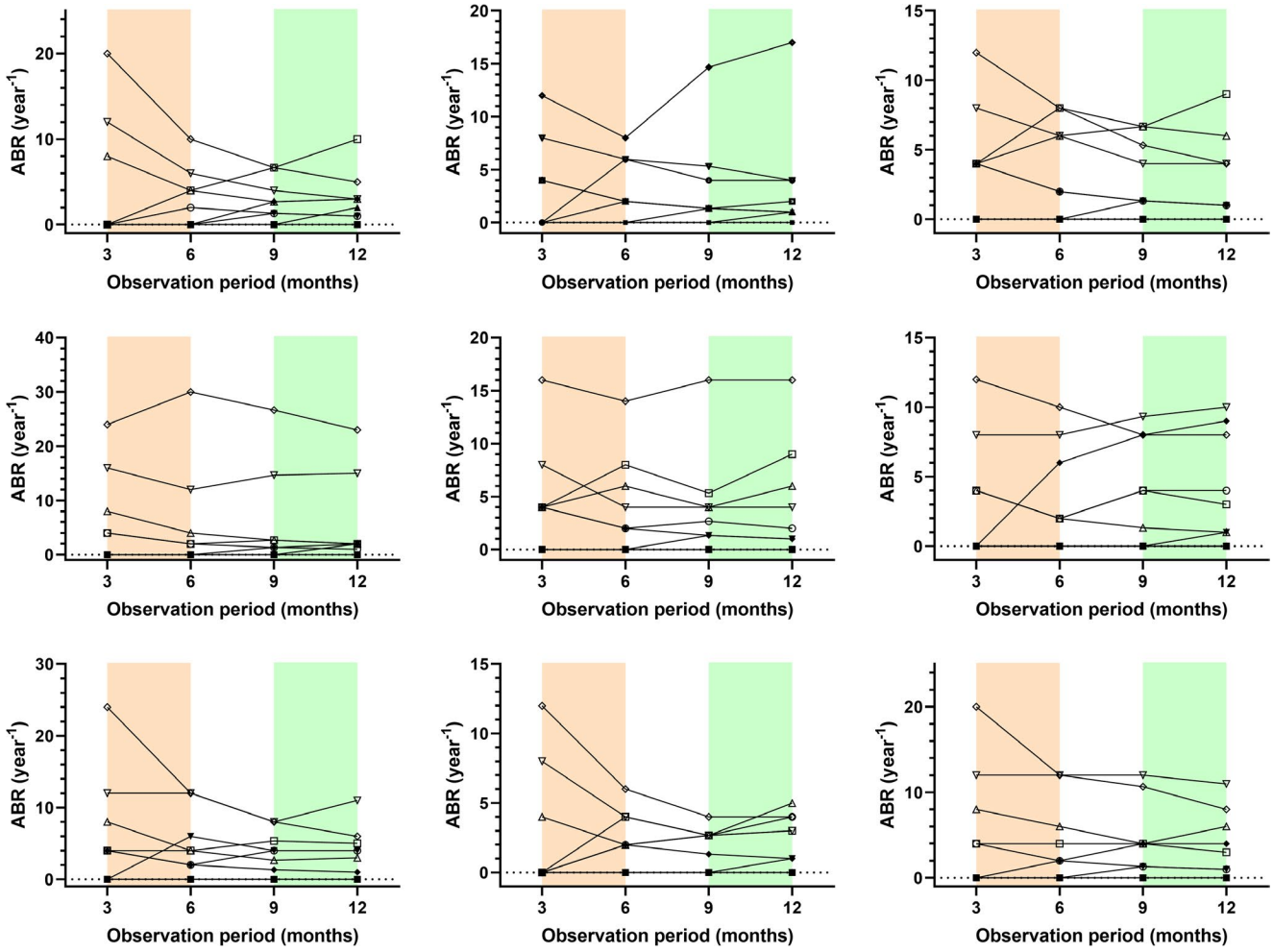

Figure 2 Individual annual bleeding rates (ABRs) from the Paul-Ehrlich-Institut internal FVIII treatment database (PEI-DB) 1-year-prophylaxis group. Each graph shows individual ABRs from 10 different patients. These ABRs were calculated for observational periods of 3, 6, 9, and 12 months from all 90 patients receiving prophylactic replacement therapy. Overlap in the illustration of low ABR patients occurs due to their relatively high proportion and often similar rates.

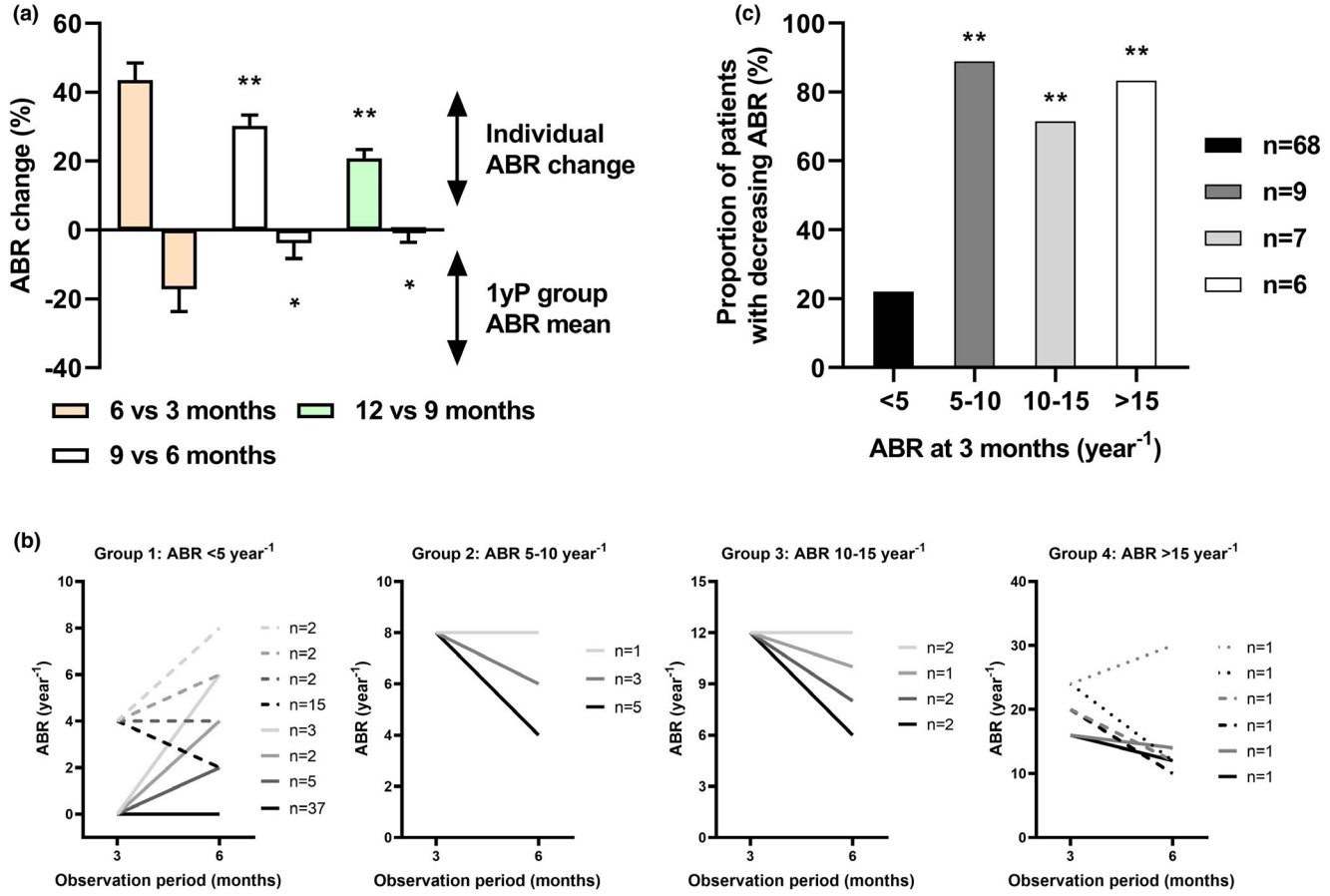

Figure 3 Impact of the observational period on the annual bleeding rate (ABR). (**a**) Relative individual ABR change was calculated from absolute values of the individual changes of the patients in the 1-year-prophylaxis (1yP) group (n = 90) irrespective of whether it is a positive or negative ABR shift for the shown observation intervals. Changes of the 1yP group ABR mean were calculated using the valued signs (direction) of the individual ABR shifts. Means are presented with 95% confidence intervals. (**b**) Patients were grouped by ABRs measured at 3 months (< 5, 5–10, 10–15, and > 15/year) and changes to the ABRs at 6 months are shown (n = 90). (**c**) Proportion of patients with decreasing ABR in the groups defined in (**b**). *P < 0.05; **P < 0.001.

of seasonal variation of bleedings in the 1yP group because of inconsistencies in the available literature.[18] In the 1yP group, a slightly larger portion of bleeds occurred during the spring (28.0%) and summer (25.2%) seasons when compared with fall (23.8%) and winter (23.0%). No statistical significance has been found neither for FVIII consumption nor seasonal variance.

The statistical properties of the ABR may pose a challenge for CT planning and indeed a statistical comparison was also planned for 52% of the analyzed CT protocols. We performed sample size model calculations to identify possible issues by using the prophylaxis group dataset from the PEI-DB (n = 204) with the ABR event rate λ_1 as reference and compared it with a hypothetical treatment group with an expected ABR event rate λ_2.

Superiority analyses showed that, for example, a total of ~ 200 subjects would be required to demonstrate a reduction of the ABR by half (**Figure 4a**) when using the Wilcoxon-Mann-Whitney test. Transforming data, such as by using the square root transformation ($ABR' = \sqrt{(ABR+0.5)}$) is another approach sometimes used in CTs to reduce data right-skewness and to enable parametric testing.[19] However, this method did not render the ABR data normally distributed (P < 0.001, D'Agostino-Pearson normality test) and led to an increase of the calculated sample size (**Figure 4a**).

Noninferiority sample size calculations were performed using a negative binomial model, which also was the most frequently used model among all CT protocols (44.2%). For noninferiority calculations, the following variables have been tested: dispersion φ, noninferiority margin δ—also referred to as the noninferiority ratio R_0—and the event ratio of the treatment and control group λ_2/λ_1. Choosing a noninferiority ratio of R_0 = 1.2 (i.e., $ABR_{treatment}/ABR_{comparator}$) and an expected event rate for the treatment group similar to that of the control group (i.e., λ_2/λ_1 = 1), sample size calculations showed a linear dependency on the dispersion parameter within the range φ = 0.3–3 (**Figure 4b**). For further analyses, we did not use the dispersion of the PEI-DB data in order to exclude CT heterogeneity of the PEI-DB. A dispersion parameter of φ = 1.5 previously has been suggested as a conservative estimate,[20] which was, therefore, used. Notably, values up to φ = 7 have been used in some CT protocols for statistical analyses. Using an event ratio of λ_2/λ_1 = 1 and φ = 1.5, sample sizes were subsequently calculated for the variation of

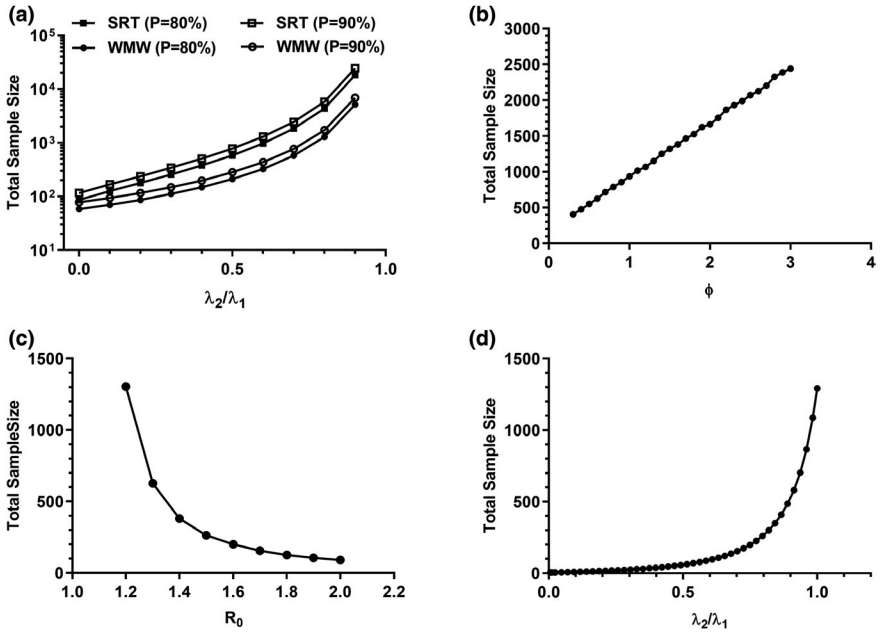

Figure 4 Sample size model calculations. The prophylaxis group from the Paul-Ehrlich-Institut internal FVIII treatment database (PEI-DB; $n = 204$) with the annual bleeding rate (ABR) event rate λ_1 was used as a control and compared with a hypothetical treatment group with an expected ABR event rate λ_2. (**a**) Calculations based on the Wilcoxon-Mann-Whitney test (WMW) or *t*-test after square root transformation (SRT) of ABR values are shown for the significance level of $\alpha = 0.05$ and powers of 80% and 90%. (**b–d**) Sample size determinations for noninferiority analysis. (**b**) Sample size dependency on the dispersion parameter φ. (**c**) Sample size dependency on the noninferiority ratio R_0. (**d**) Sample size dependency on the event ratio λ_2/λ_1.

the noninferiority ratio R_0 (**Figure 4c**). As depicted, there is large variation of the sample size up to $R_0 \leq 1.7$ and less than a 20% change with $R_0 > 1.7$. Next, the impact of different event ratios $R_1 = \lambda_2/\lambda_1$ was tested by varying the expected event rate of the hypothetical treatment group (λ_2), whereas the dispersion and noninferiority ratio were kept constant at $\varphi = 1.5$ and $R_0 = 1.2$, respectively. As a result, the estimated sample size strongly increases as λ_2 approaches λ_1 (**Figure 4d**). The major implication is that $\lambda_2 << \lambda_1$ to keep the sample size within a reasonable range. For example, an event ratio of $\lambda_2/\lambda_1 \leq 0.6$ (or $ABR_{treatment} \leq 0.6 \times ABR_{comparator}$) would be required to keep the sample size below $n = 100$.

Overall, by means of the PEI-DB analysis, we showed that ABRs have a high variance in prophylactically treated patients and that the ABR observation period is a very critical parameter in CTs.

Comparative analysis of clinical hemophilia trials and bleeding definitions

To substantiate the importance of ABR as a clinical endpoint and to analyze potential methodological differences in the estimation, a grouped analysis of anonymized data from 46 CT protocols submitted to the PEI for review was conducted. This analysis includes 27 study protocols from confirmatory CTs conducted with various FVIII products in severe HA subjects, 12 protocols from phase II/III trials conducted with antibody-based products or from noninterventional trials, and 7 protocols from gene therapy-based CTs (study phases I/II and phase III). Characteristics of the CTs are shown in **Figure 5a–d**, main summary statistics are provided in **Table 1**.

For factor-replacement products, the overall mean study duration was 40 months, and a mean of 121 subjects were planned for enrollment. The planned mean minimum duration of subject participation was 22 months or 71 exposure days (EDs). The ABR was evaluated in 74.1% of these trials and was listed as a primary and/or secondary outcome measure in 33.3% and 63.0%, respectively. Notably, in 25.9% of the trials, bleeding rates other than the ABR were estimated, such as bleeds per month or unspecified bleeding frequencies. The mean ABR observation period was 16.3 months or 66.9 EDs (median: 50; range 46–105). For factor-replacement products, the ABR was calculated in 40.7% of the CTs from on-demand or episodically treated subjects and in 44.4% from a specific subset including spontaneous ABR, traumatic ABR, joint ABR, ABR for treatment-requiring bleeds and bleeds that required no treatment, and age group-specific ABR.

In comparison, CTs investigating antibody-based products had a considerably shorter study duration and subject participation time (25.4 vs. 40.1 months and 12 vs. 22 months, respectively). The enrollment numbers were similar; however, it must be noted that some trials with antibody-based products enroll both HA and HB patients, patients with and without inhibitors, or patients with severe as well as moderate hemophilia in contrast to CTs with FVIII replacement or gene therapy trials. Therefore, nonexistent differences in study size appear in a different light. For gene therapy-based products, the planned study duration and subject participation period were longer when compared with factor-based and antibody-based products, whereas the planned number of enrolled patients was reduced (**Figure 5a** and **Table 1**).

Table 1 Summary statistics of 46 hemophilia clinical trial protocols grouped by product class

	Factor products (*n* = 27)	Antibody products (*n* = 12)	Gene therapy products (*n* = 7)
Mean study duration in months [median, range]	40.1 [30.0, 3–99]	25.4 [25.0, 18–36]	64.8 [62.0, 60–75]
Mean subject participation time in months [median, range]	22.0 [15.0, 3–72]	12.0 [10.0, 5.5–27]	31.0 [15.0, 9.2–66]
Mean subject participation time in minimum EDs [median, range]	71 [51, 50–105]	–	–
Mean planned enrollment number [median, range]	121 [125, 25–250]	128 [122, 20–272]	40 [18, 10–130]
Mean ABR observation period in months [median, range]	16.3 [8.0, 6–60]	8.4 [5.5, 5.5–24]	9.1 [10.8, 6–12]
ABR estimated (%)	74.1	100	100
Primary EP (%)	33.3	91.7	0
Secondary EP (%)	63.0	75.0	85.7
Exploratory EP (%)	3.7	8.3	14.3

ABR, annualized bleeding rate; ED, exposure day; EP, endpoint (double entries possible).

Generally, the ABR was highly frequently measured in all product classes. It was most often a primary endpoint in trials evaluating antibody-based products and was used as a secondary endpoint in 63–85% of all studies (**Figure 5b,c** and **Table 1**). ABR within a prophylactic treatment regimen was also highly frequently measured among all product types, whereas heterogeneity was observed in the estimated ABR subsets (**Figure 5c**). ABR observation periods differed notably within and between all product classes (**Figure 5d** and **Table 1**). Noteworthy, the ABR observation period was defined by the number of EDs for replacement therapies only (15 of 27 trials).

Further CT protocol analyses were focused on the definition of BEs (**Figure 6**). For FVIII replacement products, the parameters bleeding cause, severity, location, and pain were most frequently used for defining BEs in CTs. Distinction of single and multiple bleeds was less frequently taken into account (81.4%) as well as the definition of a new bleed (74.0%). Whether a BE will only be used for the calculation if treatment is required is critical for the ABR estimate. However, this criterion was applied only in two of three factor-based CT protocols (66.6%). Another important characteristic for the assessment of a BE is its association with symptoms of swelling, warmth, unusual sensation, and aura, which particularly refer to muscle and joint bleeds. Only 59.2% of the FVIII study protocols referred to swelling or warmth and even less (14.8%) to aura or unusual sensations. As shown in **Figure 6**, the frequency of use of appropriate BE definitions varies substantially between and within product classes.

The Scientific and Standardization Committee of the International Society on Thrombosis and Haemostasis (ISTH) has published a guidance to foster the consistent use of definitions in hemophilia.[21] Published definitions for joint bleeds, target joints, muscle bleeds, and new bleed were compared with those used in the respective CT protocols (**Supplementary Figure S1**). In the majority of CT protocols of factor-based products (*n* = 27), the definitions of joint bleeds, target joints, muscle bleeds, and new bleeds did not or only partially match the definitions proposed by the ISTH, consensus in the respective categories was found in 11.1%, 33.3%, 7.4%, and 44.4% of the protocols. For example, the ISTH definition suggested the occurrence of an unusual sensation or aura in the joint in combination with any other symptom, including increased swelling, warmth, pain, and loss of range of motion, for the definition of a joint bleed. In our analyses, to match this definition ≥ 2 of the symptoms had to be named, and aura or sensation had to be included. Furthermore, it was proposed to define a new bleed as a bleed that occurs > 72 hours after stopping treatment for the original bleed for which treatment was initiated. If rebleedings had been defined in the study protocol, time frames of 24 and 48 hours were evaluated as only partially matching. A subanalysis (data not shown) revealed that CTs evaluating primarily factor-based products and for which study protocols have been submitted to the PEI in 2015 and later, bleed definitions were more consistent with ISTH recommendations being published in 2014, however, the overall number of these CTs was small. In contrast to factor-based products, the majority of bleed and target joint definitions for CTs related to antibody-based products (*n* = 8/12) did match with the ISTH proposal, except for the definitions of muscle bleeds. For gene therapy-based products (*n* = 7), in approximately one third of the cases, matching definitions were used for the respective categories, whereas the absence of any definition was more common. Notably, among the four evaluated categories, muscle bleeds were most poorly defined throughout all products.

In sum, the comparative CT analysis reflects a considerable heterogeneity in the ABR estimation methodology within and between product classes. In particular, differences in the observation period and in the use of bleeding definitions have been identified as main issues.

DISCUSSION

New treatment options for hemophilia are pushing for the market. With emicizumab, the first monoclonal antibody has received approval in Europe in 2018, and gene therapy

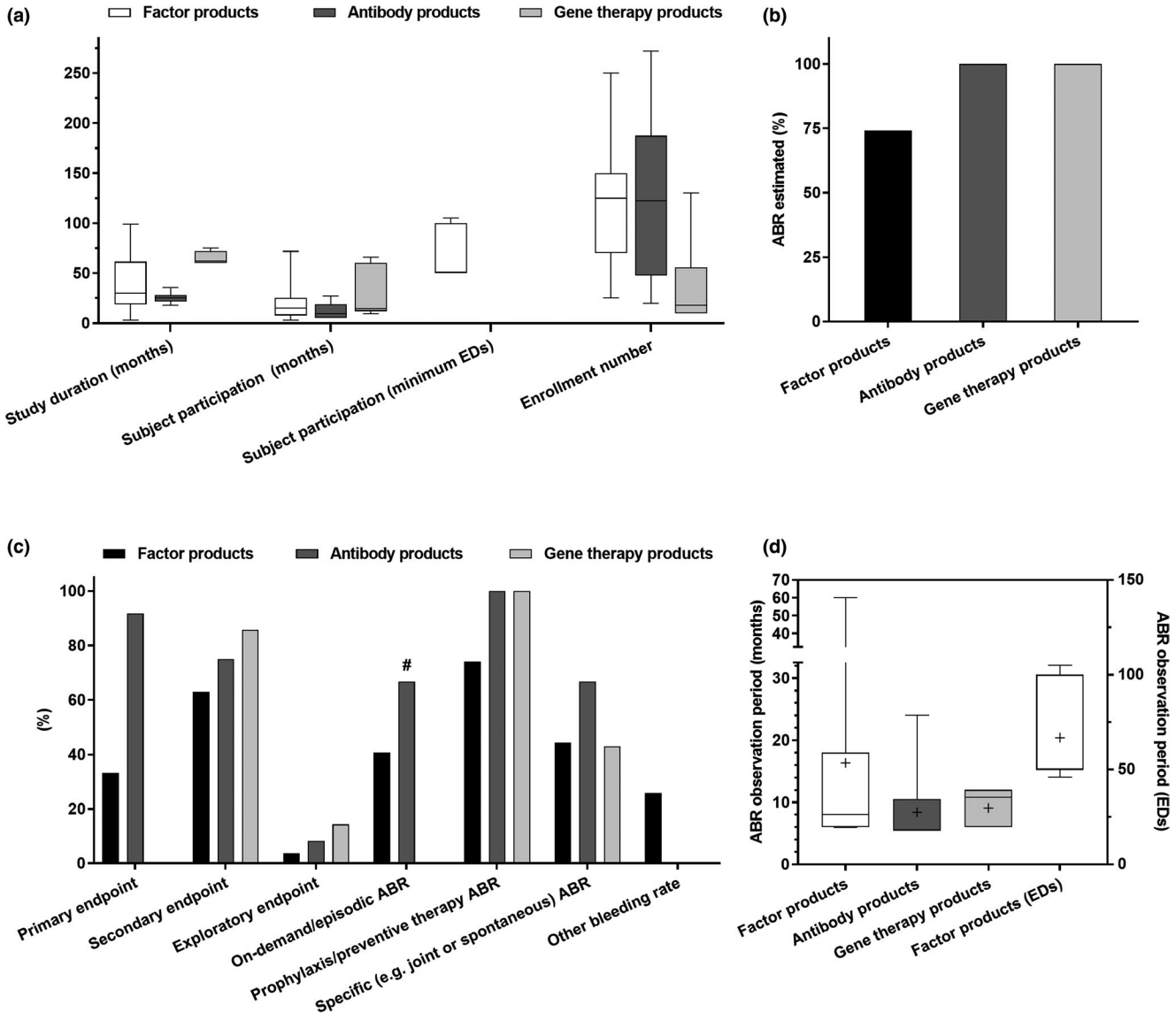

Figure 5 Comparison of 46 hemophilia clinical trial (CT) protocols. Basic characteristics of confirmatory CTs evaluating the efficacy of prophylactic substitution therapy in subjects with hemophilia A (n = 27) are shown in comparison to trial data of antibody-based (n = 12) and gene therapy-based (n = 7) products, including study duration, subject participation, and enrollment data (**a**). The frequencies of overall and specific ABR estimation are shown in (**b**) and (**c**), respectively. The ABR observation periods are depicted as box-and-whisker plots with min to max whiskers and means marked as "+" (**d**). EDs, exposure days. #, For CTs primarily evaluating the efficacy of antibody-based products, ABRs were also counted when assessed from a comparative on-demand therapy using factor-based products.

products have recently been approved eligible for the PRIority MEdicines (PRIME) program by the EMA. In the regulatory practice of FVIII replacement products in the European Union, efficacy mainly relies on FVIII kinetics, on hemostatic response in the on-demand treatment and in surgery, and on factor consumption. For prophylactic treatment, these data are to be complemented by an assessment of bleeding episodes, bleeding intervals, and number of treatments in the long-term use, whereas an ABR is not requested.[22] However, ABRs are now increasingly used in the evaluation of replacement, nonreplacement, and gene therapeutic treatments.

The analysis presented here of 46 confirmatory hemophilia CTs and the PEI-internal database encompassing ~ 1,300 previously treated patients and previously untreated patients resulting from 30 different CTs of replacement products showed considerable methodological differences in ABR analyses.

Analyses of the PEI-DB showed that differences in the length of the observation period are very critical for the ABR estimation and indeed considerable heterogeneity was also present in CT protocols with observation periods ranging from 5.5 to 60 months. Notably, although having important implications for efficacy and safety evaluation in clinical hemophilia trials,[22,23] EDs were used to define an observation period for 15 of 27 CTs analyzing factor substitution products but not in any CT dealing with nonreplacement products. The PEI-DB analysis significantly demonstrated advantage

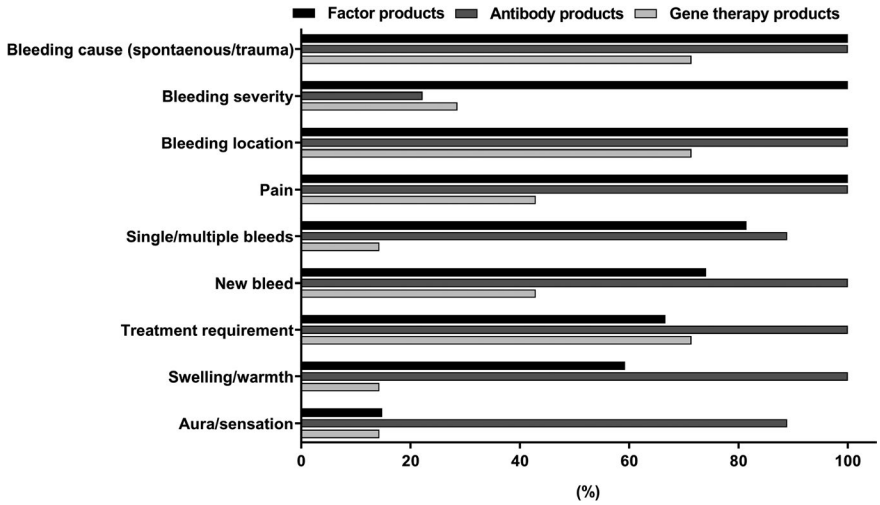

Figure 6 Definition of bleeding episodes in hemophilia clinical trials (CTs). Frequencies of the main characteristics used in CT protocols for the definition of bleeding episodes are shown for factor-based, antibody-based, and gene therapy-based products.

of an observation period of 12 months over shorter durations. Shorter periods were shown to pose major challenges due to the high ABR variance of prophylactically treated patients and the reduction of the ABR group mean within the first 6 months of observation. With regard to seasonal variation, the analysis showed a 5% difference between spring and winter. This minor finding is in line with published data showing that bleeding rates are higher in the summer and lower in the winter.[24]

In statistical analyses, the negative binomial approach is a standard parametric model for skewed count data[25] and it was also the most frequently used model in the here analyzed CTs. However, selection of the most appropriate model may differ for a particular patient cohort or treatment intervention, and more accurate estimates may be provided such as by use of zero-inflated models.[26] Our calculations showed that, depending on the event rates in the control and treatment groups, the superiority of a treatment measured against a well-established standard prophylaxis requires considerably large sample sizes. In noninferiority studies, the event and the noninferiority ratio along with the dispersion parameter have a tremendous impact on the estimated sample sizes. A meticulous choice of these parameters must be made for valid comparative analyses.

The data presented here further revealed that considerable differences in the use of bleeding definition, such as for new bleeds or joint bleeds, compromise ABR comparisons within and between the product classes. Notably, a greater number of bleeding parameters, in accordance with ISTH definitions,[21] were defined in CTs dealing with antibody-based products, whereas CTs evaluating gene therapy products showed a general lack in the definition of BEs. In addition to bleeding definitions and recommendations made by the ISTH and World Federation of Hemophilia,[21,27] subclinical bleedings could be considered in CTs as these potentially confound the efficacy evaluation. Notably, for intra-articular bleeds, the importance of imaging techniques, such as musculoskeletal ultrasound, in the assessment of BEs has been demonstrated in recent studies.[9,28] In fact, one of these studies showed that less than half of the patient-reported BEs could be confirmed by point-of-care imaging. Accordingly, cerebral micro-bleedings, which are also common in patients with hemophilia,[29,30] may be considered.

Another critical factor in the ABR analysis is the selection of patients. Residual FVIII-levels or FIX-levels define the severity of hemophilia and are generally used as inclusion criteria for CTs in support of marketing authorization. However, the "bleeding phenotype" depends considerably on additional individual factors, including genetic background, age, education, lifestyle, physical activity, or level of previous medical care.[10,31,32] Attempts to correlate residual factor levels with clinical bleeding severity, as recently published, show low specificity and reproducible and quantifiable clinical tools may be needed to better identify clinically severe patients for CTs.[33]

This study was focused on CTs in patients with HA, however, the ABR is a clinical endpoint that is also used in a large number of CTs on treatment of HB, von Willebrand disease, factor X deficiency, and other rare bleeding disorders. Although our results may not apply to every particular investigational medicinal product, they are largely transferable to these studies, which is important as the number of study participants is often smaller, thus further challenging the ABR estimation.

Overall, the following recommendations should be considered for future CT concepts: use of agreed-upon bleeding definitions and physician-based monitoring of BEs, an observational period should cover a minimum of 12 months, careful consideration of statistical models, eligible patients should have a clinically severe phenotype, and use of imaging techniques where appropriate.

Supporting Information. Supplementary information accompanies this paper on the *Clinical and Translational Science* website (www.cts-journal.com).

Acknowledgments. The authors thank Joachim Messer, Kerstin Heimann, and Ramya Raju for assistance in the constitution of the PEI-DB.

Funding. No funding was received for this work.

Conflict of Interests. The authors declared no conflict of interest.

Author Contributions. C.K. and M.M.-O. wrote the manuscript. C.K., M.M.-O. and A.H. designed the research. All authors performed the research.

1. Peters, R. & Harris, T. Advances and innovations in haemophilia treatment. *Nat. Rev. Drug Discovery* **17**, 493–508 (2018).
2. Weyand, A.C. & Pipe, S.W. New therapies for hemophilia. *Blood* **133**, 389–398 (2019).
3. Yamaguti-Hayakawa, G.G. & Ozelo, M.C. Gene therapy: paving new roads in the treatment of hemophilia. *Semin. Thromb. Hemost.* **45**, 743–750 (2019).
4. Perrin, G.Q., Herzog, R.W. & Markusic, D.M. Update on clinical gene therapy for hemophilia. *Blood* **133**, 407–414 (2019).
5. Mahlangu, J. *et al.* Emicizumab prophylaxis in patients who have hemophilia a without inhibitors. *N. Engl. J. Med.* **379**, 811–822 (2018).
6. Lillicrap, D. Bispecific antibody therapy in hemophilia. *N. Engl. J. Med.* **377**, 884–885 (2017).
7. George, L.A. *et al.* Hemophilia B gene therapy with a high-specific-activity factor IX variant. *N. Engl. J. Med.* **377**, 2215–2227 (2017).
8. Rangarajan, S. *et al.* AAV5-factor VIII gene transfer in severe hemophilia A. *N. Engl. J. Med.* **377**, 2519–2530 (2017).
9. Ceponis, A., Wong-Sefidan, I., Glass, C.S. & von Drygalski, A. Rapid musculoskeletal ultrasound for painful episodes in adult haemophilia patients. *Haemophilia* **19**, 790–798 (2013).
10. Oldenburg, J. Optimal treatment strategies for hemophilia. Achievements and limitations of current prophylactic regimens. *Blood* **125**, 2038–2044 (2015).
11. Spotts, G. *et al.* Modeling minimally-effective FVIII trough levels in hemophilia a patients on PK-guided prophylaxis. *Blood* **124**, 689 (2014).
12. Pierce, G.F. *et al.* Establishing the appropriate primary endpoint in haemophilia gene therapy pivotal studies. *Haemophilia* **23**, 643–644 (2017).
13. Chai-Adisaksopha, C., Hillis, C., Thabane, L. & Iorio, A. A systematic review of definitions and reporting of bleeding outcome measures in haemophilia. *Haemophilia* **21**, 731–735 (2015).
14. Boehlen, F., Graf, L. & Berntorp, E. Outcome measures in haemophilia: a systematic review. *Eur. J. Haematol. Suppl.* **76**, 2–15 (2014).
15. Hermans, C. *et al.* Outcome measures for adult and pediatric hemophilia patients with inhibitors. *Eur. J. Haematol.* **99**, 103–111 (2017).
16. European Medicines Agency Core summary of product characteristics for human plasma-derived and recombinant coagulation factor VIII products. EMA/CHMP/BPWP/1619/1999 Rev. 3 (2018). <https://www.ema.europa.eu/en/core-summary-product-characteristics-human-plasma-derived-recombinant-coagulation-factor-viii>.
17. European Medicines Agency Core summary of product characteristics for human plasma-derived and recombinant coagulation factor IX products. EMA/CHMP/BPWP/1625/1999 rev. 3 (2018). <https://www.ema.europa.eu/en/core-summary-product-characteristics-human-plasma-derived-recombinant-coagulation-factor-ix-products>.
18. Shafer, F., Smith, L., Vendetti, N., Rendo, P. & Carr, M. Lack of seasonal variation in bleeding and patient-assessed pain patterns in patients with haemophilia B receiving on-demand therapy. *Haemophilia* **20**, 349–353 (2014).
19. Antunes, S.V. *et al.* Randomized comparison of prophylaxis and on-demand regimens with FEIBA NF in the treatment of haemophilia A and B with inhibitors. *Haemophilia* **20**, 65–72 (2014).
20. Ragni, M.V. Rationale for a randomized controlled trial comparing two prophylaxis regimens in adults with severe hemophilia A: the Hemophilia Adult Prophylaxis Trial. *Expert Rev. Hematol.* **4**, 495–507 (2011).
21. Blanchette, V.S., Key, N.S., Ljung, L.R., Manco-Johnson, M.J., van den Berg, H.M. & Srivastava, A. Definitions in hemophilia. Communication from the SSC of the ISTH. *J. Thromb. Haemost.* **12**, 1935–1939 (2014).
22. European Medicines Agency Guideline on the clinical investigation of recombinant and human plasma-derived factor VIII products. EMA/CHMP/BPWP/144533/2009 Rev. 2 (2018). <https://www.ema.europa.eu/en/clinical-investigation-recombinant-human-plasma-derived-factor-viii-products>.
23. European Medicines Agency Guideline on Clinical investigation of recombinant and human plasma-derived factor IX products. EMA/CHMP/BPWP/144552/2009 rev. 2 Corr. 1 (2018). <https://www.ema.europa.eu/en/clinical-investigation-recombinant-human-plasma-derived-factor-ix-products>.
24. Fischer, K. *et al.* Internal and external determinants of bleeding patterns on prophylaxis for severe haemophilia A. *Blood* **110**, 1145 (2007).
25. Cameron, A.C. & Trivedi, P.K. Regression analysis of count data, Vol. **53** (Cambridge University Press, West Nyack, NY, 2013).
26. Hu, M.-C., Pavlicova, M. & Nunes, E.V. Zero-inflated and hurdle models of count data with extra zeros: examples from an HIV-risk reduction intervention trial. *Am. J. Drug Alcohol Abuse* **37**, 367–375 (2011).
27. Srivastava, A. *et al.* Guidelines for the management of hemophilia. *Haemophilia* **19**, e1–e47 (2013).
28. Nguyen, S., Lu, X., Ma, Y., Du, J., Chang, E.Y. & von Drygalski, A. Musculoskeletal ultrasound for intra-articular bleed detection: a highly sensitive imaging modality compared with conventional magnetic resonance imaging. *J. Thromb. Haemost.* **16**, 490–499 (2018).
29. Zanon, E. *et al.* Cognitive dysfunctions and cerebral microbleeds in adult patients with haemophilia A: a clinical and MRI pilot-study. *Thromb. Res.* **134**, 851–855 (2014).
30. Husseinzadeh, H. *et al.* Prevalence of and risk factors for cerebral microbleeds among adult patients with haemophilia A or B. *Haemophilia* **24**, 271–277 (2018).
31. Broderick, C.R. *et al.* Association between physical activity and risk of bleeding in children with hemophilia. *JAMA* **308**, 1452–1459 (2012).
32. Manco-Johnson, M.J., Soucie, J.M. & Gill, J.C. Prophylaxis usage, bleeding rates, and joint outcomes of hemophilia, 1999 to 2010: a surveillance project. *Blood* **129**, 2368–2374 (2017).
33. Mancuso, M.E., Bidlingmaier, C., Mahlangu, J.N., Carcao, M. & Tosetto, A. The predictive value of factor VIII/factor IX levels to define the severity of hemophilia: communication from the SSC of ISTH. *J. Thromb. Haemost.* **16**, 2106–2110 (2018).