COMMENTARY

The Rodin (Research Of Determinants of INhibitor Development among PUPs with haemophilia) study: the clinical conundrum from the perspective of haemophilia treaters

C. M. KESSLER* and A. IORIO†
*Division of Hematology-Oncology, Georgetown University Medical Center, Washington, DC, USA; and †Departments of Clinical Epidemiology and Biostatistics and Medicine, McMaster University, Hamilton, ON, Canada

A long-awaited study report

The Rodin study, recently published in the New England Journal of Medicine, has begun to provide some very important answers to several questions pertinent to the quality and safety of replacement therapy to individuals with haemophilia [1]. Probably the most serious complication of modern haemophilia care is currently the development of alloantibody neutralizing inhibitors and considerable recent literature until the Rodin trial had speculated that the recombinant factor (rFVIII) replacement products were more likely than the plasma-derived products to produce this. The Rodin trial has extended our thoughts on this issue because it is a robustly populated prospective observational study, in contrast to previously published results, which have been generated from smaller, retrospective, mostly uncontrolled heterogeneous population studies. Based on the results of 574 previously untreated severe haemophilia A patients (FVIII activity, <0.01 IU mL⁻¹), the observations in Rodin indicated that there was no difference in the incidence of alloantibody inhibitors whether patients received plasma-derived or recombinant full-length FVIII products (adjusted hazards ratio = 0.96). Furthermore, among those who developed alloantibody inhibitors while on plasma-derived FVIII concentrates, the content of von Willebrand factor (VWF) did not influence the risk of inhibitor formation (adjusted hazards ratio = 0.90). Lastly, the Rodin population study indicated that switching from a plasma-derived FVIII product (irrespective of relative VWF content) to a full-length rFVIII product did not increase the risk of inhibitor development. These conclusions are extremely cogent to patients and their physicians alike as the possibility that rFVIII products were more immunogenic, that the high content of VWF protein in plasma-derived FVIII products could protect against alloantibody formation, and that product switching would be harmful all were used as rationale to determine the choice of replacement product for previously untreated, slightly previously treated, and even those previously treated individuals with over 150 exposure days. This published report notwithstanding the possibility that it may be underpowered to conduct most of the above comparisons due to the relatively low number of patients treated with plasma-derived molecules has added to the cumulative published data, which tend to discount these concerns when determining product choice for previously untreated patients (PUPs) and others. In fact, several of the haemophilia treaters and haemophilia treatment centres (HTC), who participated in Rodin, had expressed publicly prior to this publication that these same concerns influenced their decision to initiate their PUPs considered at high risk for inhibitor development on plasma-derived products. The Rodin trial was observational, that is NOT a randomized controlled study, and allowed each HTC to determine independently how it was to treat its PUPs. This approach could have led to an imbalance in the baseline prognostic characteristics of the groups being compared in Rodin and this potential bias could have introduced a significant biostatistical flaw into the study design [2].

A sudden twist: unexpected additional results

In a post hoc analysis [1] and not apparently intended to be included in the original trial design [3], Gouw et al. compared the two generations of full-length rFVIII concentrates employed in the study for alloantibody inhibitor formation. Surprisingly, the second-generation rFVIII derived from a baby hamster kidney (BHK) cell line was declared to be more immunogenic than the third-generation rFVIII concentrate derived from a Chinese hamster ovary (CHO) cell line (adjusted
hazards ratio = 1.60). The reported P value for this difference is 0.02; however, different statistical assumptions apply when analysing post hoc-derived data, so that this P value does not prove a non-casual difference, although to the physician who is untrained in the nuances of biostatistics, the P value may appear to have the usual meaning of clinical significance [4].

There was no biologically plausible explanation for this last finding and a previous publication using the second-generation BHK-synthesized rFVIII concentrate in PUPs would refute this finding [5]. In any case, it may be a moot point since a third-generation formulation of the BHK derived full-length rFVIII concentrate is expected to be commercially available shortly. This new BHK product will match the purity, specific activity and degree of freedom from synthesis and purification in the presence of added human protein as the currently available third-generation FVIII concentrate derived from CHO cells. Nevertheless, several speculations have arisen as to the aetiology of the differential immunogenicity of the second and third-generation products. For instance, the BHK formulation may contain more FVIII protein in aggregate form [6], which could affect enhanced antigen processing by the antigen presenting cells of the immune system with subsequent peptide formation; alternatively, the two different cell lines could generate rFVIII proteins with different degrees of glycosylation and the immune system might process these two proteins differently. It should be noted, in this context, that a similar increased risk for inhibitor development, even if not reaching statistical significance, was demonstrated in PTPs in a recent published and widely discussed meta-analysis (HR for all de novo inhibitors 2.43; CI, 0.31–19.2 and HR for high-titre de novo inhibitors 1.75; CI, 0.05–65.5, for BHK vs. CHO) [7].

**Time for reflection: appraising the Rodin study report**

Those who read this commentary understand how difficult it is to conduct randomized controlled clinical trials in the haemophilia arena. Although one of the largest and more comprehensive prospective studies to date, the Rodin study does not provide such a high level of evidence to allow a strong confidence in its results. The authors are the first ones to state that their study has important limitations. For instance, Rodin is not a fully prospective controlled study and was predominantly comprised of a lower risk ethnic population (90% Caucasians) for inhibitor development. As inferred in their published addendum, the first 4 years of their study (2000–2004) appeared to be retrospective in approach and the data generated during this period probably reflected results from the second-generation rFVIII since the third-generation rFVIII product was not commercially available in the EU during that time period. Subsequently, for the remaining 6 years of Rodin, there was a specified data collection form so that the trial was clearly prospective and involved both generation rFVIII concentrates. This article appears to have combined the data from both study periods in their biostatistical analysis rather than analysing the results separately as well as combined. It is not clear how this approach may have confounded their conclusions; however, there are currently in process several well-designed prospective studies, which may confirm or contradict Rodin’s findings.

Two initial aspects of the Rodin trial design should be examined. First, patients were allocated to the products indicated by their treaters and thus were subject to the potential ‘biases’ of their treaters and/or their Hemophilia ‘Treatment Centers’ own local guidelines, preferences, or attitudes. It is indeed possible that such treatment decisions resulted in ascertainment or selection bias. Second, although the study authors discount the possibility that centre-specific bias could have confounded their conclusions, given that the variability of prophylaxis regimens and intensity of treatment have already been adjusted for, it would have been more supportive and reassuring if alternative analytical approaches for this study design had been employed to control for the risk of bias. Such statistical techniques could have included propensity score analysis [8] and centre-stratified or adjusted Cox-regression, or an assessment of deviation from the overall mean rate of inhibitor formation in different centres. In the setting of a post hoc analysis, exploring the potential sources of variability with multiple techniques is generally useful to distinguish robust findings from chance ones.

A further methodological concern of the Rodin trial is that it relied on the Bethesda unit inhibitor levels to be measured at each individual HTC rather than performed at a central laboratory. It is not apparent whether all the HTC laboratories were standardized in their assay techniques. At first reading this might appear irrelevant for the study, which focuses on only clinically relevant inhibitors, but this is not the case, because Rodin employed a highly laboratory-dependent definition of inhibitor clinical relevance.

Of most concern in the Rodin study design is the possible deviation from the complete analysis of the entire inception cohort [9]. According to the Methods of the Rodin study, 648 ‘eligible’ patients were recruited to the study, of whom 74 were ultimately excluded from the statistical analysis. Of these, 19 in the initial cut and 30 patients in the subsequent cut were excluded for reasons related to inhibitor development/ascertainment, based on information provided in the patients’ disposition flow chart. In the third cut, two individuals had documented inhibitors, but were not included in the final statistics. Thus, of the initial 648 recruited patients, inhibitor development was calculated on only 574 patients despite the inclusive
study design. The reasons for why the dropouts were excluded from the statistics are not adequately elucidated and the size of the excluded population is such that their inclusion may have influenced the overall allocation of incidence risk of inhibitors.

In addition, the Rodin study intended to follow patients up to a 75-exposure day endpoint; however, the discussion indicates that not all the patients had reached that endpoint, that these subjects remained at risk for inhibitor development, and that they were still included in the final statistical analysis. It would be useful to know the details of how these patients were distributed among the treatment products so that a better assessment of individual inhibitor risk could be determined.

Finally, some corollary technical details in the Rodin approach to biostatistical analysis should be considered.

1. The choice of the third-generation full-length rFVIII product to be the ‘reference group’ for statistical analysis was justified in the publication with the statement ‘the product type that was used most frequently was selected as the reference category’. However, more patients were treated with second (N = 183) compared to third-generation (N = 157) rFVIII. Furthermore, the number of patients is the more appropriate denominator to calculate the rate of inhibitor development as the rate of inhibitor development decreases over time after the initial treatment period. The statistical methods of Rodin may have allowed for inadvertent selection bias and ultimately may have influenced the final results generated from the comparisons chosen for the study.

2. The results of the multivariate analysis are presented in a summary manner, without clarifying the contribution of individual risk factors and without mention of interaction terms (which are essential in understanding if an independent effect or the combination of two different risk factors is playing a role). The risk factors chosen to be included in the multivariable analysis were ethnicity, FVIII gene mutation type, family history of haemophilia with inhibitors, age at first exposure, reason for first treatment, duration between exposure days, dose of FVIII replacement, history of switching between product brands, peak treatment moments, major surgery and regular prophylaxis. Some of these are putative more than proven risk factors for inhibitor development. It is not clear how these risk factors were weighted in Rodin. Some of these potential confounders/risk factors are fixed while others are time-varying; the Rodin statistical approach employed simultaneous adjustment of all risk factors. Gouw et al. [10] recently documented in their well conducted meta-analysis that certain FVIII genotypes can raise the odds ratio of alloantibody inhibitor development by over ninefold, suggesting that the weight of FVIII genotype as a predisposing risk factor for inhibitors may outweigh the importance of all other risk factors [11] in such a manner that it might not be accounted for even in multivariate analyses. Such specific genetic data for all of the PUPs included in Rodin should be made available for each study cohort to assure the absence of unintended bias.

3. Despite differences between second and third-generation rFVIII concentrates in table 4 of the supplementary material (which reports the risks of inhibitor development according to type of FVIII product in patients who did not participate in PUP studies) and in fig. 2 of the article (PUP study participants) the hazards ratios, both adjusted and unadjusted, are not significantly different for second-generation and plasma-derived factor concentrates.

4. In the Rodin study, a small number of patients received a variety of plasma-derived FVIII products. Monoclonal plasma-derived highly purified FVIII products were conglomerated with all other plasma-derived products although they are more equivalent to recombinant products in terms of their von Willebrand factor protein content. The article did not describe the different plasma FVIII products used, but interestingly 2/7 on ‘low VWF content’ concentrates developed inhibitors with a low number of exposure days. More clinical detail on these individuals and the other PUPs on plasma-derived FVIII would be helpful.

All these small details taken together and in absence of a pre-specified hypothesis make the odds of a chance finding very likely.

Building our take home message: confidence in various results and exploring clinical implications

The question arises as to whether or not one can accept the validity of the first two conclusions presented in the Rodin study: [1] equivalent allo-FVIII antibody inhibitor rates between recombinant and plasma-derived concentrates, and [2] no increased inhibitor development associated with switching from plasma-derived to recombinant concentrates, but reject the comparison among different brands of rFVIII concentrates based on statistical bases. The strongest reason for doing so is that the first two objectives were a priori specified as study objectives, while the third one was an unexpected finding of a post hoc analysis. The second reason is the weak biological rationale for a difference between subsequent ‘generations’ of full-length rFVIII. Data from well-designed PUP studies for each of the full-length rFVIII products employed in Rodin are available. For the BHK second-genera-
tion product, the maximum inhibitor incidence was 18% [5]. The CHO derived third-generation product was employed in the prospective Early Prophylaxis Immunologic Challenge (EPIC) Study (ClinicalTrials.gov Identifier: NCT01376700) of PUPs and this trial was recently terminated because of a surprisingly high incidence of alloantibody inhibitor development, attributed to ‘protocol deviations’ (personal communication). Thus, when studying inhibitors in haemophilia populations, trial design is critical, and, due to limited population availability, particularly pertinent to PUPs, statistical analysis may yield unexpected results. Such unexpected results may lead to publication of provocative although counterintuitive conclusions with unpredictable consequences on clinical practice and decision-making.

It is critical to the clinician that the conclusions of the Rodin study be placed in perspective so that wise treatment and regulatory decisions can be made. This study primarily set out to prove (or otherwise) that there was no reduced inhibitor risk with plasma-derived products when compared to rFVIII products. In addition, the study was designed to determine whether product switching would increase the risk of inhibitor development in PUPs. These important objectives were convincingly achieved and will certainly influence the standard of care for PUPs. On the other hand, the Rodin study also suggested that a second-generation rFVIII concentrate may increase inhibitor risk and this conclusion has promulgated a loud danger signal. How this finding will be interpreted by government agencies, patient consumers, and physician prescribers may adversely affect patient, treater and health care authorities’ acceptance of such products. Provoked by the results of the Rodin study, the European Medicines Agency has recently initiated a review of the safety of the second-generation rFVIII used in the trial and intends to determine whether the marketing authorization of the product should be maintained, varied, suspended, or withdrawn across the EU [12]. A similar fate also could potentially befall the third-generation rFVIII used in Rodin, related to the higher than anticipated PUP inhibitor incidence in the EPIC study. This commentary offers an opportunity for open discussion of the results of the Rodin trial, the appropriate biostatistical approach to study design for future clinical research efforts in this field, and the relative value of a prospective/retrospective observational study vs. prospective, randomized controlled trials.

Disclosures

CMK has received research funding from Baxter, Bayer, Grifols, Octapharma, NovoNordisk and Pfizer. He has also served on advisory boards for Baxter, Bayer, Biogen, CSL, Grifols, Octapharma, NovoNordisk, and has consulted for all mentioned companies. He is not on any speakers bureaus but has received honoraria from all mentioned companies for providing educational programmes and participating in CME generating symposia. AI has received research funding from Baxter, Bayer, Pfizer and NovoNordisk. He has also served on advisory boards for Baxter, Bayer, Pfizer and NovoNordisk and has consulted for Bayer and NovoNordisk. He received honoraria from all mentioned companies for providing educational programmes and for participating in CME generating symposia.

References