Valoctocogene Roxaparvovec and Emicizumab for Hemophilia A without Inhibitors: Effectiveness and Value

Response to Public Comments on Draft Evidence Report

October 16, 2020

Table of Contents

#	Comment	Response/Integration
Patient Organizations		
National Hemophilia Federation & Hemophilia Federation America		
1.	Although described in the original PICOTS, ICER did not complete a comparison of Emicizumab directly to Valoctocogene Roxaparvovec. The fact this analysis was not completed should be clearly indicated. Combining the results of the two completed model versions (FVIII v Emicizumab, FVIII v Valoctocogene Roxaparvovec) in one blended discussion leads to confusion and potential misinterpretation. This presentation is potentially misleading and risks readers concluding the tables are intended to provide a direct comparison between Emicizumab and Valoctocogene Roxaparvovec. For example, tables such as 5.10 summarizing the results of the base case analysis present results which are derived from significantly different inputs (e.g., age, starting Pettersson score, bleed rates, dosing). The data presented for total costs, reduced joint bleeds, life-years gained and QALYs are not comparable. There is no mention of the differences in inputs when Table 5.10 appears in the report. A reader would need to recall or refer 15 pages earlier to find a description. For clarity we request the report be split between the two models	ICER did complete the comparison as described in the PICOTS, and rated that comparison "I". ICER usually does not attempt an economic comparison when the rating of comparative effectiveness is "I".
2.	For this review, we submitted data to ICER demonstrating patient benefit for several key metrics. We are disappointed that ICER did not analyze all the available direct qualitative evidence and indirect evidence reporting on core patient outcomes. Specific to Emicizumab, the report incorrectly states that ICER did not identify any studies assessing the impact of prophylaxis with Emicizumab on “Other Outcomes.” This statement may lead readers to miss or downgrade key information regarding Emicizumab’s demonstrated and clinically significant beneficial impact on patient-relevant outcomes	Thank you for your comment. We have revised the statement on 'other outcomes'.
3.	The Haem-A-QoL instrument used in the Haven 3 and 4 studies is currently the most widely used hemophilia-specific instrument for the evaluation of HRQoL in PwHA. Haem-A-Qol provides a composite score of many outcomes important to patients (physical health, feeling, view of oneself, sports and leisure, work and school, treatment, future, family planning, partnership and sexuality) as well as individual domain scores. The physical health domain is an important facet of the Haem-A-QoL that measures the high impact of hemophilia A on physical activity. Use of this instrument is a recognized strength of the HAVEN studies. The improvement seen in the physical health sub-score for patients treated with Emicizumab demonstrated clinically relevant differences between Emicizumab and FVIII prophylaxis.	Thank you for your comment. We reported the Haem-A-QOL physical health outcome in the draft report. We have revised our report to highlight the clinical relevance of the improvement seen in the physical health sub-score in the HAVEN trials, and also added data on the other available outcomes.

4.	It is also our understanding that the sponsor provided ICER with data expanding on a 2019 ISTH HAVEN study abstract (Ref 38), in anticipation of a forthcoming manuscript. The additional data demonstrate that the proportion of participants with no missed workdays increased to ≥90% with Emicizumab prophylaxis in both HAVEN 3 and HAVEN 4. (Ref 38) We request the report reflect the observations of this patient-relevant metric.	Thank you for your comment. We have revised the report to include the missed workdays observed in HAVEN 3 and HAVEN 4.
5.	The frequency of infusions required for prophylaxis regimes significantly impacts PwHA quality of life.1 We request ICER consider accounting for treatment burden in the model.	We did not find any sufficiently high-quality studies regarding disutility related to factor VIII infusions and consequently do not include such a disutility in the model. We added language to the discussion regarding this potential additional impact.
6.	Relative to gene therapy, real-world comparative data on the use of Valoctocogene Roxaparvovec is not available from clinical trial data. However, the PROBE study (referenced in our earlier submissions) provides relevant indirect evidence demonstrating a difference in outcomes for PwHA by phenotype. While PROBE does not collect head-to head data for comparison of specific drugs, submitted data responded directly to elements within the ICER Data Request (bleed disutility, impact on joint range of motion, arthropathy, utility values) for PwHA with severe disease, PwHA with severe disease on prophylaxis, and PwHA with a mild phenotype. Additional data were in-press (now published) reporting on other comorbidities, acute and chronic pain, pain occurrence and interference, pain medication frequency, impact on activities of daily living and use of mobility aids by phenotype.2 PROBE validation was described in prior submissions.	We agree that the PROBE data provide indirect evidence on the outcomes of patients with hemophilia by phenotype. We have now included the PROBE data in our discussion on the impact of valoctocogene roxaparvovec on Health-Related Quality of Life
7.	The data presented provide important contextual information for the ICER analysis and modeling to establish baselines, as well as indicate potential anticipated outcomes if individuals were to achieve a milder phenotype. It is unknown if someone born with a mild phenotype is equivalent to someone who attains a mild state through gene therapy. Nevertheless, data indicate that the phenotypic disease state does matter, not just in terms of clinical outcomes, but also for patient-important outcomes directly impacting quality of life. While some PwHA living with a milder phenotype still encounter significant negative impacts from their hemophilia, data demonstrate overall improvement across the metrics for those living with mild disease or no hemophilia. The ICER report should include a descriptive analysis clearly reflecting the value of living with a less severe phenotype.	See above.

8.	FVIII Dose Selection Not Based on Real-World Utilization We are likewise troubled that ICER did not adopt utilization data that is based on real-world experience. The goal of any cost effectiveness model should be to reflect as closely as possible real-world scenarios, so that the resulting analysis is credible. There are statements within the report (e.g., "we are uncertain of the added efficacy of these higher doses"; "especially at currently used dosages") which appear to call into question more contemporary factor utilization. If these statements remain in the final report, we ask that you indicate that such statements are outside the scope of this review and the analysis should in no way be used to assess current standards for optimal dosing strategies for factor replacement prophylaxis. There is significant clinical consensus that microbleeding, breakthrough bleeding, and other negative sequelae often result when standard prophylaxis targeted to a 1% trough level is utilized. Such suboptimal treatment does not meet WFH Guidelines, which state that "Dosing and dosing intervals should be sufficient to prevent spontaneous and breakthrough bleeding and hemarthrosis." (Ref 25) Indeed, WFH treatment guidelines recommend escalation of prophylactic doses as necessary to achieve these goals.	We have changed the base case to more closely reflect current average doses of factor VIII in practice based on ATHN data.
9.	A model built on a pharmacokinetic (PK) analysis could accommodate a real-world outcomes-based approach for utilization. Across health sectors, there is an increasing recognition of the importance of personalized medicine. Within the field of hemophilia this is often achieved through PK-guided dosing. In the absence of utilizing PK based data, it is difficult to comprehend why the base case model is not built on contemporary representative data from the real-world derived directly from the national registry (ATHN dataset).	We feel a specific PK based model is beyond the scope of our analysis, but we have changed the base case to more closely reflect current average doses of factor VIII in practice. We believe that with this change, average dosing would be similar to that seen in a PK-based model.
10.	Rather, the base case inappropriately relies on utilization data from the SPINART study (Ref 45, 46). The SPINART study was undertaken in response to an unmet need for data to establish the benefits of secondary and tertiary prophylaxis in adult PwHA, in high-resource and low-resource countries, who had at least 150 exposure days to any factor product and who had established arthropathy. To be in the study, a PwHA had to have anywhere from 6-24 bleeding episodes in the 6 months prior to enrollment and not be on prophylaxis. This is a group that essentially does not exist anymore in the United States. SPINART was a significant research study at the time. Hematologists have learned from research such as SPINART and adjusted clinical practice accordingly, precisely as we expect from a learning health care system. The older model of therapy studied in SPINART would not be considered standard of care for a PwHA receiving care through a federally supported Hemophilia Treatment Center today. The study data is now over a decade old (data collected between 2008 and 2013). It is no longer generalizable to the standard of care in 2020. Additionally, SPINART was conducted prior to the first US approval of an extended half-life FVIII product (Eloctate 2014). It also preceded foundational research such as that conducted by den Uijl et al (Ref 53) which provided critical insight into the importance of FVIII levels relative to joint bleeds. The	We now use the SPINART study and the lower doses as a scenario analysis. We do believe that it is a relevant scenario as the best comparative efficacy data come from that trial and those numbers are tied to the lower doses of factor VIII.

	den Uijl work, along with the introduction of extended half-life products, have significantly contributed to substantial changes in treatment practice in the years since the SPINART study was conducted and published. For many PwHA, higher trough levels and thus higher utilization are required to prevent joint bleeding. Even at the time, the dosing utilized in SPINART (25 IU/kg) was considered suboptimal for many PwHA.	
11.	ICER should not build its base case analysis around data which is outdated and unrepresentative of current clinical practice. To do so disregards the objectives of hemophilia treatment today (e.g., normalizing quality of life, preserving joint function, and eliminating bleeding). We request ICER recognize the fundamental flaw in the base model assumptions and replace the utilization data of the base case with the real-world data provided by ATHN.	We are changing the base case to use doses of factor VIII consistent with average doses seen in prophylaxis use of factor VIII in ATHN data along with efficacy results from hemophilia treatment centers in the US as reported in Malec 2020. We also will include as a scenario analysis what was the base case in the draft report, and we include language in the discussion and conclusion reflecting the implications of that scenario relative to the new base case.
12.	Moreover, we are concerned that ICER's relegation of the real-world analysis to page 70-71 of the report could be interpreted as devaluing the therapeutic advances of recent years that have led to better outcomes for PwHA. The report fails to capture the clinical gains achieved with higher factor utilization. The comments on page 70-71 summarizing the model variations based on real-world (rather than outdated) data should be prominently moved to the beginning of the report and fully highlighted in the Executive Summary.	See above.
13.	NHF and HFA share ICER's concern over the availability of real-world evidence and robust patient-relevant outcomes data across the lifecycle of drug development. We appreciate that there are challenges in using real-world data for economic evaluations; however, where such (systematically collected) data exists, ICER should recognize and utilize it. We are hopeful this review serves as a framework and reminder to all stakeholders that the generation of clinically and patient-relevant outcome data should remain a high priority early in clinical development, and that the timely presentation of patient-relevant data matters for economic evaluations conducted by ICER.	We agree and have expanded our discussions of PROs based on RWE. Additionally, we expect that recommendations around better capturing PROs will come out of the public meeting.

14.	Inhibitor Development (page 39) The draft report's models omit any consideration of inhibitor development, the most significant treatment-related event today for severe PwHA. While we appreciate that ICER made this decision in order to simplify the model, this omission does not reflect real-world conditions. There is no way to predict which PwHA will or will not develop an inhibitor. Readers need to understand the resulting cost analysis omits this important, highly costly and highly impactful adverse event. Assumptions substantially differing from the real-world must be explicitly stated in the table of assumptions (Table 5.1).	This decision was not made to simplify the model but because there are not adequate data on the effects of the therapies on inhibitor development. With regard to valoctocogene roxaparvovec, the therapy was administered to adults without inhibitors. With regard to emicizumab, ICER has already found it a dominant therapy in patients with inhibitors, but we heard disagreement from experts as to whether starting emicizumab at birth would increase or decrease development of inhibitors. A clinical trial is underway to evaluate this question. It is not helpful to model an outcome for which neither the direction nor the magnitude of the therapy's effect on the outcome is known.
15.	In addition, we request ICER consider a scenario analysis modeling inhibitor development for previously untreated patients beginning prophylaxis at birth with FVIII concentrates relative to an assumed rate for those who begin with Emicizumab or gene therapy. Such an analysis would be informative for both health systems and PwHA on the total burden of hemophilia.	We received conflicting input from clinical experts as to which treatment would result in more or less inhibitor development. We add language to the discussion regarding how inhibitor development differences could impact the real world cost effectiveness of the treatments.
16.	Biological Activity of post gene therapy FVIII (page 40) The draft report notes that "The manufacturer of valoctocogene roxaparvovec has suggested that the low bleeding rate seen even as factor VIII levels decline imply that the factor VIII produced by gene therapy may be more biologically active than the factor VIII in patients with mild or moderate hemophilia". This speculative statement has no basis in fact. It has not been peer-reviewed and the sponsor has submitted no evidence to support it as a hypothesis. Thus, please omit this speculative and unsupported statement.	We try to address areas of uncertainty in our reports, and this issue was raised by prominent experts in hemophilia. We feel that we provided sufficient caveats around this.
17.	We recognize the gene therapies being developed for PwHA today may be just the first step toward a curative therapy; however, along the way, we should not legitimize a notion that low factor activity expression measured by bleeding events alone is an acceptable outcome for gene therapy. Non-severe hemophilia is not benign. The metrics for differentiating gene therapy from standard of care extend beyond bleed rates.3 In the absence of peer reviewed longitudinal data fully characterizing clinically- and patient-important outcomes including joint preservation, absence of sub-clinical joint bleeding (microhemorrhages) and impact on HRQoL from diminished factor activity levels, the inclusion of such post-hoc speculation risks uncertainty for both progress which has been made for the existing standard of care, but also the future evolution of treatment.	Although there may be a sense that the issues are conflated, the mechanistic hypothesis for observed results (above) does not imply that low factor activity is benign.

18.	ICER describes methods for estimating durability of factor activity (page 59) for both "an optimistic scenario (starting at a factor level of 89 IU/dL and using the proportional decline seen from year 3 to 4 to project) and a conservative scenario (same starting point as the base case and using a linear projection of decline)." The Cook model described on page 68 reports an average successful duration of gene therapy of roughly 11 years to 5% factor activity. It is difficult for readers to interpret and compare the various approaches modeled. Providing the duration required for cost-effectiveness at a given price (or, conversely the price for a gene therapy that would meet cost effectiveness for a known duration) would be highly insightful. Given this is a number derived from a population (e.g., mean, median), if median is used, half the patients will be below, and half above 5% and this should be taken into account in the prediction of durability. Analysis based on Factor VIII activity derived from chromogenic assays is recommended by the sponsor of Valoctocogene Roxaparvovec.4 In the future, each gene therapy manufacturer will need to demonstrate which assay (one-stage or chromogenic) provides useful predictive information for comparison of the population treated. We request ICER include a table reflecting the duration of effectiveness to 5% modeled by ICER for Valoctocogene Roxaparvovec in each of the analyses.	We have included tables illustrating our durability predictions in the base case and scenario analyses.
Patients Rising Now		
1.	The FDA approved emicizumab for use in people with hemophilia A without inhibitors on October 4, 2018. We point this out because we believe that in discussing the clinical situation for people with hemophilia A it should be made more explicitly clear that the current draft report: 1) Is addressing a different subset of patients with hemophilia A than ICER's 2018 report; 2) Is reviewing a second FDA-approved indication for emicizumab; and 3) People with hemophilia A develop inhibitors after receiving factor VIII. For example, the draft report notes that "the development of inhibitors has very important implications for management, costs, and quality of life." But simply referencing ICER's 2018 report here is inadequate. We strongly feel that the draft report should contain more extensive discussions of the differences between the two groups of people with hemophilia A, the natural course and history of how people with hemophilia develop inhibitors and what that means for their treatment and care options, costs, and lives. An updated and complete discussion of those matters is important not only for patients and clinicians to assess ICER's work, but also for policy makers and payers to be able to determine the utility of ICER's reporting for their internal health technology assessments and related practices and policies. In addition, to parallel ICER's 2018 report, we believe the report's title should include the phrase "without inhibitors," i.e., "Valoctocogene Roxaparvovec and Emicizumab for Hemophilia A Without Inhibitors."	Thank you. We agree the title was confusing given the non-overlapping population with the prior report. We have edited the title as you suggested.

2.	We are again somewhat dismayed that ICER proceeded with examining valoctocogene roxaparvovec even after the FDA declined to advance its own review, which essentially pushes back the earliest approval date to late 2021 or sometime in 2022. Given that the remainder of the review concerns emicizumab, which was originally approved in November 2017, we fundamentally question the utility of the draft report: Is the purpose to review data about an approved indication that is now two years old, i.e., for hemophilia A without inhibitors? Or to review data that will be updated in about a year with additional information before any potential decision could be made by the FDA? Either of those scenarios has limited utility to patients, clinicians, payers or policy makers. Specifically, policy decisions about additional uses for emicizumab have already been established, and clinical decision makers should already be familiar with the information. Similarly, it is premature to make any clinical or policy decisions about valoctocogene roxaparvovec. Thus, the draft report's assessments and "findings" are inherently unactionable for those audiences. Similarly, the report is not actionable for payers and policy makers, who work on an annual or biannual timeframe that corresponds to benefit plan years. For them, emicizumab's second approved indication should already be incorporated into their processes, and while valoctocogene roxaparvovec should be on their radar, it is not a factor they need to consider now. In fact, they cannot evaluate it until the additional research required by the FDA has been conducted and analyzed. ICER's review seems premature at best.	As in the past, we very much appreciate the concerns Patients Rising Now has about how ICER should expend its limited resources with considerations of how to make our reports as valuable and actionable as possible.
3.	The draft report also minimizes the real patient implications of having to receive treatments or prophylaxis intravenously versus subcutaneously. This difference between emicizumab and factor VIII is important, particularly for people with hemophilia A who have transportation or mobility limitations or live in geographic areas where access to clinical facilities for intravenous treatments may require many hours of travel. Those differences in route of administration can also make the real-life benefits and utility very different from what is reported in clinical studies, particularly as they may be calculated in a meta-analysis of multiple studies. We noted that the draft report includes the sentiments of patients that support this real-world benefit, "98% of patients favored emicizumab over factor VIII prophylaxis," and "all caregivers [in the HOEHEMI trial] indicated the lower frequency of treatment and easier route of administration as the major reasons for their preference for emicizumab." And as the draft report states, "If reductions in adherence outside of trials are not similar for the two therapies [clinical trial data] could incorrectly characterize the relative benefits of the therapies in the real world. Emicizumab prophylaxis is substantially less burdensome than factor VIII prophylaxis, and so real world adherence is likely to be more similar to clinical trial adherence with emicizumab than with factor VIII." Similarly, such real-world adherence differences could translate into great benefits for people, particularly because the draft report found that "Emicizumab appears to have lower bleeding rates (of all types) compared with factor VIII." If ICER's goal is to affect real-life policies and actions rather than to provide guidance for future research studies, we urge ICER to expand its recognition and discussion of such people-centered factors.	This seems to be a misunderstanding of how emicizumab and factor VIII are used. While it is clearly an advantage for patients to have a less frequent subcutaneous prophylactic therapy, patients with hemophilia A do not generally travel hours to clinical facilities multiple times per week for prophylaxis with IV factor VIII. As noted, ICER discussed the advantage of sc prophylaxis with emicizumab in multiple portions of the report.

4.	Value Viewpoints about Cures for Serious Diseases Because the draft report includes information about valoctocogene roxaparvovec as a potential gene therapy despite the FDA's decision to defer action, we believe it is appropriate to comment on the meaning of such treatments. While it seems from the currently available data that valoctocogene roxaparvovec may have some waning of effectiveness after several years, it also needs to be viewed through the lens of how medical progress actually occurs. Similar to how biplanes were not directly or immediately replaced by jets, improvements in treatments occurs incrementally – sometimes with small steps, and sometimes in larger leaps. Clearly the development of a gene therapy that is effective for several years is a leap over injections that must be given every few weeks. However, valoctocogene roxaparvovec should also not be viewed as the finish to the race for gene therapies. Indeed, improvements will be made upon that very large step, with the ultimate goal of having reliable, stable, and permanent cures. Thus, the initial leap – in this case valoctocogene roxaparvovec – needs to be viewed in its context as part of a process of treatment advancements. For example, in viewing the advancement that valoctocogene roxaparvovec potentially represents, the variability of individual responses to treatment as depicted in Figure 4.1 in the draft report is illuminating. Such individual variability indicates that there is much still to be learned about the use of such gene therapies, and how to customize or adjust their use for individual patients, which – again – is part of the process of innovation and the advancement of scientific knowledge to improve care and outcomes. For patients, such significant leaps represent hope in concept – as well as in reality – that better treatments will be developed while they are benefiting from those that are small steps or significant leaps, but that still leave them with some impairment, limitation or dependence upon ongoing treatments. This value of hope is real for patients even if payers, policy makers, and quantitative modelers are unable or unwilling to incorporate that reality for patients into their cognition and conclusions. We hope that as ICER continues to refine its processes and practices it will be able to better include that value and similar perspectives of real patients.	We agree that medical progress is a process.

5.	ICER correctly determined that because valoctocogene roxaparvovec is not yet approved and more research is ongoing, the draft report should not include a Budget Impact analysis for this potential gene therapy. We applaud this decision, as such hypothetical exercises can do more harm than good. However, in the past we have expressed concern about certain technical and procedural components of ICER's Budget Impact analyses, and with the current draft report there is an additional confusing aspect. Specifically, the draft report includes a Budget Impact analysis for emicizumab, even though it is not a newly approved compound; FDA approved this medicine in November 2017. We find this inconsistent with ICER's potential Budget Impact analysis formula that includes the number of newly approved medicines as a fundamental factor. This is problematic because to anyone familiar with the reality of the U.S. health care system, off-label uses of approved medicines is both common and an expected and necessary part of quality health care, except in very rare circumstances. Thus, ICER's conducting a Budget Impact analysis on a 3-year-old medicine presents a murky analytical rationale within ICER's theoretical Budget Impact evaluation process.	Again, we appreciate Patients Rising Now highlighting the apparently confusing statement in the draft report that says, "Emicizumab already has an established presence in the market and so no potential budget impact analysis is included for emicizumab." We have not changed this for the next version of the report, however if Patients Rising Now has suggestions on how we might make this clearer, we can make edits between this revised version and the Final Report.
6.	We would appreciate ICER clarifying how it will consistently conduct potential Budget Impact analyses based upon original versus subsequently approved indications. We eagerly await ICER's insights about how it can be more consistent and coherent in this particular facet of its activities.	ICER does not usually conduct potential budget impact analyses for products with an established presence on the market. However, we would point out cases where subsequently approved indications could lead to a large expansion of the population eligible for treatment.
7.	In addition, as we discussed in our comments to ICER's 2018 draft report about the use of emicizumab, people with hemophilia are not evenly distributed among all the different payers in the U.S. Specifically, data indicates that people with hemophilia are much more likely to be insured by Medicaid, and less likely to be insured by Medicare or the Veterans Health Administration. (It is not unreasonable to postulate that they are also very unlikely to be covered by the Department of Defense's health system.) However, when a curative gene therapy for hemophilia is available, those differences may disappear. We are not advocating that ICER attempt to include such forward, evolutionary modeling into its work – since our impression is that ICER prefers to view the future as a static situation – but we believe it must be part of broader discussions concerning how budget impact should be conducted, and modeling of potential future scenarios could be constructed.	ICER's models may consider payer-specific scenarios where relevant and as data allow, for the incorporation of alternative scenarios as more data become available.
8.	Lastly, we would be remiss if we did not point out that ICER's style of global budget impact assessments don't account for the patient perspective: what matters to patients and their families is their actual costs, not some aggregate for the entire country. And further, regarding health system or payor budgets and spending, people with serious and chronic conditions have intense concerns about how any budget or access restrictions will impair innovations that could help treat or cure their health problems, and improve or prolong their lives – real-world implication that are generally missing from ICER's work and activities.	ICER's potential budget impact analyses are intended to provide an "access and affordability alert" so that planning can take place to maintain patient access despite budgetary pressures that may arise from payments for innovative treatments. We believe that patients are concerned not only about their own out-of-pocket costs, but also about rising aggregate costs of health care and health insurance.

9.	We are confused by the lack of inclusion of Serious Adverse Events in the Long-Term Cost Effectiveness Model inputs since in ICER's 2018 report SAEs were included in the model at a rate of 3%. If there is a difference in clinically observed SAEs in people with and without inhibitors then this certainly should be presented and discussed by ICER.	This distinction is mentioned in the section comparing emicizumab with factor VIII.
10.	ICER's SST framework that arbitrarily picks caps of $150,000/patient/year or 50% over a lifetime for the amount of "cost savings" that a company might receive from a new treatment in this category continues to be puzzling. We are particularly concerned about treatments – such as gene therapies – that could be very expensive to produce and administer, and as such if either the $150,000 number or a 50% threshold of "cost savings" were somehow implemented, it could result in net losses for the company, leading to the discontinuation of the treatment or service.	ICER's value framework acknowledges that any given level of sharing of cost offsets may be considered arbitrary (as is assuming that all of the cost savings should be captured in the price). However, we believe that these scenarios provide useful information to stimulate broader discussion on the use of cost effectiveness to guide value-based pricing for SSTs and similar health care interventions.
11.	Reference #13 should be updated to the link for the most current label for emicizumab since the text refers to both the initial approval for hemophilia with inhibitors and the subsequent approval for hemophilia without inhibitors. The year for the World Federation of Hemophilia Guidelines should be 2020 not 2012. There is reference in the text to an economic model in a 2017 ICER report, but there is no citation or footnote.	Thank you for these comments. We have updated the citations and text accordingly.
Partnership to Improve Patient Care		
1.	ICER neglects to use available real-world data In most cases, using real world data on healthcare utilization as an alternative to data from a clinical trial has a very strong justification and will be more reliable. This has been outlined quite clearly by ISPOR and many other leading experts in the field of health economics. Populations in clinical trials tend to be on average much healthier than those in the real world, as many grounds for exclusion from clinical trials such as existing co-morbidities, age, ongoing current treatment failure can mean exclusion at registration and other factors can lead to withdrawal from the trial once it begins . Additionally, individuals participating in clinical trial settings tend to receive a much more intensive form of healthcare regime due to study requirements and close monitoring of patients, ultimately resulting in better overall control of the studied disease. This can also result in a reduction in acute or adverse events associated with their respective condition that would not otherwise be replicated outside of the trial setting.	The report uses real world data in multiple sections.
2.	ICER's draft evidence report relies on RCT data for its base case but also contains a discussion that indicates both therapies are cost saving when "real world usage" of Factor VIII is incorporated into the model rather the use measured from RCT populations. It is our opinion that the real-world data source for use of Factor VIII should be used in the base case. If this had been done, ICER's results could look very different with more reflective utility inputs, incorporating mortality effects, better long-run data on effectiveness and assuming Factor VIII costs that are more reflective of the real world.	We have changed the base case to more closely reflect current average doses of factor VIII in practice. We have also changed the source of the efficacy data in the base case from the trial based NMA to a survey based estimate of bleeds from patients in US Hemophilia Treatment Centers.

3.	ICER's use of the quality-adjusted life year is inappropriate As we have noted in many previous comment letters, the use of the quality-adjusted life year (QALY) in ICER's evaluations is inappropriate and discriminatory, as it is inherently biased against people with disabilities and patients suffering from chronic illnesses, like hemophilia. The use of the QALY in this assessment is particularly concerning given the health state utility valuations (HSUV) used. It is widely acknowledged that people experiencing chronic illnesses and disabilities regularly adapt to their conditions. This leads to what is known as disease state adaptation or hedonic adaptation—when patients and people with disabilities to overestimate their own quality of life. The result of this is that assessments relying on these HSUVs will undervalue treatments being assessed. Hedonic adaptation is a well-documented phenomenon among hemophilia patients. In fact, a paper written by some of the same authors as the paper from which ICER sources its utility values puts this point forward and makes clear that it should be considered when conducting economic evaluations for hemophilia.	As we have noted in response to your previous comment letters, ICER follows common academic and health technology assessment standards in using cost per QALY gained, but also presents cost per life year gained and cost per evLYG. A recent legal analysis found that the QALY does not disadvantage patients who have a disability or a chronic condition that is not curable: http://icerreview.org/wpcontent/uploads/2020/03/ICER-Analysesand-Payer-Use-of-Cost-effectivenessResults-Based-on-the-QALY-and-evLYGAre-Consistent-With-ADA-Protections-forIndividuals-With-Disabilities.pdf. We are aware of the issue of hedonic adaptation, but believe this needs to be balanced with the desire to reflect the experiences of actual patients.
4.	ICER omits outcomes that matter to patients Organizations representing patients with hemophilia expressed to ICER early in the review process that, although annualized bleed rates are important, there are other factors that matter deeply to patients, including chronic pain and mental health. ICER neglected to incorporate these patient-reported outcomes in its model, and we encourage ICER to do so prior to the release of the final report. Furthermore, the National Hemophilia Foundation and Hemophilia Foundation of America requested in their previous letter that ICER include Factor Activity Level, Chronic Pain, and Mental Health in its assessment outcomes. They noted that, " Where data are not available for the outcomes of interest (those listed in the Draft Scoping Document and the additional outcomes noted above) or a metric is not yet established, ICER should nevertheless recognize the full set of outcomes within its valuation as they are of importance to patients. Such recognition will guide future clinical trials and the planned real-world evaluations noted in the 2020-2023 ICER Value Assessment Framework." We agree with and echo this request. It is seminally important that value assessments move in the direction of capturing data on outcomes that actually matter to the ultimate user, the patient. As a leader in the HTA field, it is ICER's responsibility to move assessments in a direction rooted in patient-centeredness and accounts for their feedback.	ICER's model focused on patient-important outcomes related to joint pain, joint deterioration, and the need for joint surgery. It did not focus on annualized bleed rates except to the extent that a given bleed caused pain, required treatment, and hastened joint damage.

5.	ICER makes the questionable assumption of no mortality effects ICER noted that it expects no mortality effects to result from treatment with either therapy. Specifically, ICER implied that as prophylaxis for hemophilia A in patients without inhibitors has not been demonstrated to decrease mortality, then there should be no mortality effect as a result of the treatments being studied. This is an oversimplification and appears to contradict other studies. The same paper ICER references to make this justification suggests that the severity of hemophilia has a strong impact on relative mortality. It states that the mean life expectancy of someone with severe hemophilia is 63 compared to someone with mild or moderate hemophilia of 75. If the severe form of the disease can remove 12 years from a patient's life expectancy, then it is highly likely that treatment to alleviate the root cause of the disease and its consequences will result in a lower mortality	For the reasons outlined in the report, we think there are inadequate data to conclude that the new treatments improve survival compared with factor VIII prophylaxis, and also that prior studies looking at survival in hemophilia likely do not apply to today's patients. That said, it remains possible that an effective gene therapy will extend life, and that would clearly increase the value of such a therapy. We hope that future studies explore this issue.
6.	A recent Swedish cohort study based on a long-standing hemophilia registry showed that the hazard ratio for all-cause mortality for those with Hemophilia A compared with controls was 1.7, $P < 0.001$ when patients with HIV and/or viral hepatitis were excluded. The corresponding figures for the severe hemophilia subgroup were 6.6, $P < 0.001$. This was despite the fact that those with hemophilia were 57% less likely to die from ischemic heart disease than controls. There is also evidence of reduced inhibitors in those treated with emicizumab, which is a known risk factor for morbidity. Even though findings were mixed in smaller studies due to their role in mortality, recent larger studies also suggest that they are an important factor and can increase risk of death by up to 70%. Additionally, a similar study concluded mortality rates were five times higher in Hemophilia A patients with inhibitors, than those without. By assuming no mortality effects from treatments deemed to be highly effective in treating a disease known to have higher rates of premature death, ICER is underestimating the true "value" of these therapies.	The most recent period in this study was from 2001-2008 and calculated a mean age of death based on 19 patients. As above, it is possible that hemophilia shortens life expectancy but even if it does, it would need to be proven that different modes of prophylaxis affect this differently. Assuming full life expectancy is a favorable assumption for therapies that improve quality of life since they improve that quality of life over more years.
7.	ICER oversimplifies health states. Transition probabilities between Pettersson score (PS) health states are based on expected annual joint bleed rates and a literature-based assumption that on average 36.52 joint bleeds result in an increase of the PS by 1 for patients under 25 and 6.52 joint bleeds result in the same shift for patients aged 25 years or more.	We acknowledge that our model simplifies across total possible health states. However, the model incorporates the best available structure for projecting costs and utilities related to bleeds in the short run and joint deterioration and subsequent needs for surgery in the long term. The relationship between bleeds and PS scores comes from a recently published large sample size analysis of bleeds and subsequent changes in PS scores from the POTTER trial.
8.	There is a significant difference between these rates, and the use of only these two buckets oversimplifies the patient experience. The result of this oversimplification may be to over- or under- estimate the true value of therapies dependent on the age at which the treatment starts. As such, it would make far more sense to use a regression model to determine the relationship between number of joint bleeds and PS score, than it would to have just two categories covering a wide level of heterogeneity across patients.	See above. The results incorporated in the model were in fact based on regression analyses of patient data in the POTTER trial.

Manufacturers		
Bayer Healthcare		
1.	The ICER Evidence Rating of "B+" for emicizumab versus factor VIII prophylaxis should be lowered Concerns regarding the comparison of HAVEN 3 and SPINART The draft evidence report concluded with "high certainty that there is at least a small net benefit of emicizumab compared with factor VIII prophylaxis, and moderate certainty of a substantial net health benefit." As such, in patients with severe hemophilia A without inhibitors, the evidence for emicizumab compared with factor VIII prophylaxis was rated as "incremental or better" (B+; Rind 2020).	In this revised Evidence Report, we decided the correct comparator for emicizumab is factor VIII prophylaxis at the doses that are typically used in the US today. For reasons presented in the report, this has led us to downgrade the rating for the comparison to C++. Given that we have made this change, we are choosing not to address every individual issue raised below, but we believe that the B+ rating is correct when comparing emicizumab with the doses of factor VIII used in SPINART.
2.	While it is unclear how much of the overall rating is based on the results of the network meta-analysis (NMA), as the report does not describe how the evidence considered was qualitatively or quantitatively synthesized to reach a net comparative rating, the evidence from the NMA likely contributed substantially to the overall rating. However, we believe the current rating greatly overstates the level of certainty of the comparative net health benefit of emicizumab that could have been obtained from the NMA	See above.
3.	Level of Bias: How much risk of bias is there in the study designs that comprise the entire evidence base? There is substantial risk of bias due to differences in the 2 trials that were employed in the NMA (Rind 2020). First, the patient populations were substantially different in terms of age. In the HAVEN 3 trial, the median age was 40 years (range: 16-77; Mahlangu 2018). Conversely, the median age (incorrectly reported as 31 years in the ICER report; Rind 2020) in the SPINART trial was 29 years (range: 15-50; Manco-Johnson 2013; Manco-Johnson 2017). It is unclear if this difference may modify the treatment effects. More problematic is the large discrepancy in the treatment duration between the 2 trials (24 weeks in HAVEN 3 vs 3 years in SPINART) included in the NMA (Rind 2020). While the ICER report concluded that "this was not expected to affect NMAs of bleeding rates, as these outcomes were annualized," the results from the SPINART trial do not support this conclusion. Results from the SPINART trial were reported with a treatment duration of 2 years (Manco-Johnson 2013) and then again with a treatment duration of 3 years (Manco-Johnson 2017). These results clearly show the annualized bleeding rates can change based on the treatment duration. For example, the mean treated bleeding rate per year in the on-demand treatment arm (no prophylaxis) changed from 30.5 at 2 years to 37.2 at 3 years (approximately a 22% change), and for the prophylaxis group, mean treated bleeding rate per year changed from 2.0 at 2 years to 2.5 at 3 years (a 25% change). Similar changes were also observed for the annualized treated joint bleeds.	The results from the SPINART trial are consistent with our conclusion. The data you are referencing as the results from the SPINART trial have been subsequently corrected by the authors in a corrigendum. According to the corrigendum, the mean treated bleed at 2 years in the on-demand arm was 36.9 (versus 37.2 in year 3). And the mean treated bleed at 2 years in the prophylaxis group was 2.2 (versus 2.5) at year 3. For treated joint bleed, it was 29.2 versus 28.7 in the on-demand arm; and 1.9 versus 1.9 in the prophylaxis arm. Overall, the rate ratio for both treated bleed (0.06) and treated joint bleed (0.07) stayed consistent in year 2 and year 3[Manco-Johnson Corrigendum 2013]

4.	The difference in the annualized bleed rate at 2 and 3 years within the SPINART trial also suggests that there were in fact differences in the baseline risk between the HAVEN 3 and SPINART trials. While the ICER report stated that the 2 trials had similar mean annualized bleed rates for the no prophylaxis arm (38.2 vs 37.2), this was based on a comparison of the annualized bleed rate at 24 weeks vs 3 years. When the comparison is made at 24 weeks vs 2 years, the baseline risk for annualized treated bleeds is 38.2 vs 30.5 (over a 25% difference). Given the clear impact of the treatment duration on the baseline risk for the annualized bleed rate, the difference at 24 weeks was likely even greater. The 2 trials did not report the bleed rates at baseline in a consistent manner that would lend itself to a clear comparison of the populations, so unfortunately, it is unclear if the difference in bleed rates existed at the outset. Ultimately, the bleed rate is highly likely to act as a treatment effect modifier for the 2 outcomes investigated. Given the differences in trial durations, combined with changes in bleed rates over time, it is likely that a substantial amount of bias exists in the NMA results.	See above.
5.	Applicability: How generalizable are the results to real-world populations and conditions? The trials both employed relatively small sample sizes (89 in HAVEN 3 and 84 in SPINART) and provided relatively little background information on the trial populations (Mahlangu 2018; Manco-Johnson 2013; Manco-Johnson 2018). Thus, it is unclear how representative these results are of the general population of individuals with severe hemophilia A without inhibitors, and consequently, the generalizability of the results is low to moderate, at best.	We acknowledge the limitations of the small sample sizes. Background information on the two trials are presented in Table 4.2 in our evidence report.
6.	Consistency: Do the studies produce similar treatment effects, or do they conflict in some ways? As mentioned under the section on bias, the baseline risk in bleeding rates was likely different for the 2 trials. This has the potential to introduce heterogeneity and inconsistency into the NMA. Unfortunately, since there is only a single trial for each direct comparison, neither heterogeneity nor inconsistency could be quantitively assessed.	The baseline risk for the two trials were similar (38.2 in HAVEN 3 vs. 37.2 in SPINART)
7.	Directness: Are direct or indirect comparisons of therapies available, and/or are direct patient outcomes measured or only surrogate outcomes, and if surrogate outcomes only, how validated are these measures? Only an indirect comparison of emicizumab vs factor VIII is available, with a single small trial for each treatment compared to no prophylaxis. While the results of the indirect comparison are highly questionable, the results within each trial clearly demonstrate that prophylaxis with either emicizumab or factor VIII is far superior to no prophylaxis in patients with severe hemophilia A. Further, both treatments appear to completely eliminate treated bleeding events for over half of the patients included in the trial (58% at 24 weeks for patients on emicizumab and 52% at 2 years for patients on factor VIII from SPINART; Mahlangu 2018; Manco-Johnson 2013). This result suggests that more clinically meaningful endpoint definitions beyond annualized bleed rate are needed to compare these treatments.	See above.

8.	Another concern regarding the comparison of HAVEN3 and SPINART is the difference in comparator arms between trials. In SPINART, the comparator arm is on-demand treatment with Kogenate (Rind 2020); whereas, in HAVEN3, the "no prophylaxis" comparator arm does not describe which on-demand treatment is used (Mahlangu 2018).	See above.
9.	Precision: Does the overall database include enough robust data to provide precise estimates of benefits and harms, or are estimates/confidence intervals quite broad? There is a high level of uncertainty in both endpoints analyzed in the NMA. For annualized treated bleeds, the rate ratio 95% credible interval extends from 0.22 to 1.47, clearly spanning all 4 magnitude of effect categories (substantial net benefit to negative net benefit; Rind 2020). Similarly, for annualized treated joint bleeds, the rate ratio 95% credible interval extends from 0.2 to 1.39, again clearly spanning all 4 categories. Additionally, it should be noted that the exposure time in person-years (a very important data input for the NMA) was approximated for the SPINART trial, which introduces further uncertainty into these estimates	See above.
10.	Furthermore, results from the NMA showed a non-significant difference between emicizumab and factor VIII prophylaxis on bleeding outcomes of treated bleeds and treated joint bleeds (Rind 2020). Although the NMA results showed lower rates of treated bleeds and treated joint bleeds for emicizumab compared to factor VIII prophylaxis, the lack of a significant result on both clinical outcomes from the NMA indicate that there is no evidence to show that emicizumab has a greater clinical benefit than factor VIII prophylaxis.	See above.
11.	Heterogeneity of factor VIII products The heterogeneity of factor VIII products should also be considered in the assessment as data from SPINART may not be generalizable to all factor VIII products. Currently, a basket of over 15 factor VIII products have been studied and approved for prophylaxis. Additionally, Kogenate is a legacy standard half-life agent (studied from 2008 to 2011), and many newer agents have been approved since that allow for reduced treatment burden and have improved pharmacokinetic parameters that reduce clearance and ensure higher exposure, as demonstrated in area under the curve (AUC) studies (Shah 2019).	See above.

12.	Safety of emicizumab All risk data for emicizumab should be considered. Although published clinical trial data for emicizumab from HAVEN3, HAVEN4, and HOHEMI reported no thrombotic adverse events or deaths (Rind 2020), long-term safety data is lacking. The median treatment duration from HAVEN3 is 29 weeks, from HAVEN4 is 25.6 weeks and from HOHOEMI is 39.1 and 32.1 weeks in the "3 mg/kg every 2 weeks cohort" and the "6 mg/kg every 4 weeks" cohort, respectively (Rind 2020; Shima 2019). In the pooled HAVEN clinical trials, 3.5% of patients tested positive for anti-emicizumab antibodies and 1% of patients developed anti-emicizumab antibodies with neutralizing potential based on declining pharmacokinetics (HEMLIBRA 2020). One patient from HAVEN 2 developed an anti-emicizumab neutralizing antibody and experienced loss of efficacy after 5 weeks of treatment. Additionally, recent emicizumab data from clinical trials, expanded access, and compassionate use, showed 1 case of a serious thrombotic event from HAVEN 3, and recent emicizumab data from post-FDA approval showed 19 cases of thrombotic events in people treated with emicizumab, all of which were not in cases that met criteria for the boxed warning (excess on average of a cumulative amount of >100 U/kg/24 hours of activated prothrombin complex concentrate for ≥24 hours; Genentech 2020a). Unfortunately, the report on thrombotic events does not specify whether these events occurred in patients with or without inhibitors. In contrast, thrombotic events have not been reported in patients receiving factor VIII products such as Jivi (JIVI 2018). Recent emicizumab data from clinical trials, expanded access, compassionate use, and the postmarketing setting also showed deaths in 13 patients with congenital hemophilia A without factor VIII inhibitors, 3 patients with hemophilia A and whose inhibitor status was not reported, and 3 patients whose indication was not reported (Genentech 2020b). Taken together, there are safety concerns of emicizumab that need to be further assessed.	We added a discussion of uncertainties around safety to the report.
13.	Therefore, when considering the 5 domains related to the strength of evidence (level of bias, applicability, consistency, directness, and precision), the heterogeneity of factor VIII products, and concerns regarding the safety of emicizumab, we believe the current evidence rating of "B+" overstates the level of certainty of the comparative net health benefit of emicizumab compared to factor VIII prophylaxis and suggest the evidence rating be revised to "I" to indicate insufficient evidence of a net health benefit.	As above, we have concluded that we do not have high certainty of at least a small net health benefit. This has led to a rating of "C++", not "I".
14.	The dosing of factor VIII prophylaxis used in the scenario analysis is not representative of real-world dosing, and thus should not be used to make conclusions regarding cost-effectiveness In the ICER draft evidence report, scenario analyses were conducted using doses of factor VIII that were intended to be more representative of doses currently used in the US (Rind 2020). However, the doses of 118.2 IU/kg for Advate and 111.2 IU/kg for Eloctate were based on unpublished data that is not publicly available, and the methodology for how this data was collected is not described and therefore may be subject to bias and limitations. Specifically, the scenario analysis dosing represents a 48% increase from the base-case dose of 80 IU/kg for Advate and a 43% increase from 78 IU/kg for Eloctate.	The doses come from an average of initial prophylactic treatment regimens in the ATHN data, which we will describe more clearly in the report. We maintain that these doses are valid and representative of actual prophylactic doses of factor VIII in patients receiving care in US Hemophilia Treatment centers.

15.	While we appreciate ICER's willingness to incorporate real-world evidence into assessments, data used in the scenario analyses must be carefully selected and contextualized to ensure accuracy. An analysis of the Cost of Hemophilia: A Socioeconomic Survey (CHESS) dataset in the US using chart review and physician survey data demonstrated that extended half-life products have lower utilization compared to standard half-life products in the real-world setting. Patients receiving extended half-life products were found to use 9.0 IU/kg per week less than their standard half-life counterparts (mean of 70.2 IU/kg and 79.2 IU/kg, respectively; Curtis 2019).	We realize there is variance in dosing of these products. We are using evidence based doses in the base case.
16.	Additionally, based on an internal claims analysis of real-world specialty pharmacy data from Symphony Health from July 1, 2018 to June 20, 2019 (among patients with hemophilia A, ≥6 fills, aged ≥12 years, with weight data available), real-world dosing of extended half-life products (Jivi, Eloctate and Adynovate) was consistent with the label midpoint (real-world dosing ranged from 97%-101% of label midpoint; Bayer 2020). Conversely, the real-world dosing of the standard half-life product (Advate) was 139% higher than the label midpoint (Bayer 2020). These data confirm that newer extended half-life products demonstrate narrow dosing ranges in the real-world setting that are consistent with the label dosing which allows for more predictability in utilization	Again, we realize there is variance in dosing of these products. We are using evidence based doses in the base case. We also account for variance in dosing in the sensitivity and scenario analyses.
17.	These studies show that there is variability in real-world factor utilization due to differences in study methodologies. While it is not clear how data provided by the American Thrombosis and Hemostasis Networks (ATHN) was collected, it is likely based on information from clinicians. Results from the CHESS US study, which was based on data collected from physicians, showed notably different utilization rates for both extended half-life (CHESS US: 70.2 IU/kg, ATHN: 111.2 IU/kg) and standard half-life (CHESS US: 79.2 IU/kg, ATHN: 118.2 IU/kg) products. While chart review data provide important information around what the physician ordered, claims data more accurately represent patient utilization and therefore are more appropriate to use for a cost-effectiveness model. Results from the specialty pharmacy claims data also showed notably different utilization rates from the ATHN data for extended half-life products (SP: 71.5-90.7 IU/kg, ATHN: 111.2 IU/kg). Given the large impact that factor VIII utilization can have on report conclusions, data used in the scenario analyses should be carefully selected, with consideration of the data source and methodology.	We are now using the ATHN based doses in the base case analyses. These doses reflect average dosing in US Hemophilia Treatment Centers.
18.	Lastly, the base case results showed that emicizumab had an incremental cost-effectiveness ratio of $10,393,000 per QALY relative to factor VIII, and the probabilistic sensitivity analysis results showed that emicizumab was found to be cost-effective at thresholds from $50,000 to $250,000 per QALY in only 14.0-14.5% of the simulations, which included scenarios with higher factor utilization. These results should be emphasized more in the conclusions on long-term cost-effectiveness to provide the reader with a complete picture of the long-terms costs of emicizumab compared to factor VIII prophylaxis.	We will be changing the base case and the scenarios. We will include a variety of sensitivity analyses for the new base case and will discuss the implications of those findings. We will also discuss the implications of what will now be a scenario analysis based on the trial doses and efficacy that were the old base case.

BioMarin		
1.	Given the US Food & Drug Administration (FDA)'s issuance on August 19, 2020 of a "Complete Response Letter" in response to BioMarin's biologics license application for valoctocogene roxaparvovec, BioMarin agrees with ICER's statement that the results in this report should be considered "highly preliminary". Since valoctocogene roxaparvovec will not be available to patients in the near term and more clinical data will be available for future consideration of marketing approval by the FDA1, it would be premature to conduct and publish clinical and cost-effectiveness review of valoctocogene roxaparvovec or use the current assessment for any recommendations at this time	ICER would not have initiated this review had the conclusion of the FDA review been known at the outset. However, we believe that publication of our results will be helpful to multiple stakeholders in thinking about evaluating gene therapies for hemophilia and that input on this review will improve future reviews. As noted, conclusions around valoctocogene roxaparvovec are preliminary.
2.	Under clinical effectiveness review, ICER rates the evidence as "promising but inconclusive" (P/I) comparing valoctocogene roxaparvovec to FVIII prophylaxis (FVIII) and the evidence as 'insufficient' (I) comparing valoctocogene roxaparvovec to emicizumab. In cost-effectiveness review, ICER concludes valoctocogene roxaparvovec is not cost effective compared to FVIII in the base case. However, ICER also notes the results are highly sensitive to assumptions made in the model (e.g., valoctocogene roxaparvovec had >43% chance of being cost-effective in probabilistic sensitivity analyses at all listed willingness to pay thresholds). ICER highlights that the model drivers in the valoctocogene roxaparvovec model are costs of emicizumab and costs of prophylactic FVIII, parameters that are not related to the efficacy or cost of valoctocogene roxaparvovec. More importantly, the analysis shows that when doses of prophylactic FVIII reflecting current practice in the US are used, the model results are completely reversed, with valoctocogene roxaparvovec dominating prophylactic FVIII.	We have changed the base case to reflect dosing more consistent with current use as seen in ATHN data. We have also changed the efficacy of factor VIII to be consistent with results from patients using the current dosing levels of factor VIII in the US.
3.	The cost-effectiveness model is inconsistent in assuming that follow-on long-term prophylaxis after valoctocogene roxaparvovec loss of efficacy will be with emicizumab rather than FVIII, while the comparator arm continues with life-long FVIII prophylaxis; such an approach is incongruent with appropriate economic assessment generally as well as the objectives and scope of this particular review.	Since emicizumab is cost saving with equal efficacy using the revised dosing of factor VIII, it is a preferred strategy once a switch is required. In fact, it is an issue in the analysis overall that factor VIII at doses larger than those seen in earlier trials is a cost-inefficient strategy.
4.	The draft model assumes that after treatment with valoctocogene roxaparvovec, once a patient's FVIII activity level dropped below 5 IU/dL (after 16 cycles or 8 years) and below 1 IU/dL (after 25 cycles or 12.5 years), 5% and the remaining 95%, respectively, of patients in the valoctocogene roxaparvovec cohort would switch back to prophylactic therapy. The model assumes that the prophylactic therapy patients would switch to is emicizumab, and all patients would remain on emicizumab for the remainder of their life. Given the lifetime horizon of the model, patients will spend more time on subsequent therapy, which then drives the results, and the model becomes a comparison between emicizumab and FVIII. The decision to use emicizumab as the follow-on therapy alongside the reference scenario of life-long FVIII prophylaxis generates a comparison that is difficult to interpret and of limited applicability for decision makers, as the results are largely driven by the costs of emicizumab vs. FVIII, rather than the costs or the benefits of valoctocogene roxaparvovec	See above. The results are clear for decision makers relative to factor VIII at current doses.
5.	BioMarin recommendations for ICER's updated model and future evaluation of hemophilia A:	See above.

	• To minimize confounding of the analysis results by choice of follow-on treatment, the model should assume that if a patient switches back to continuous prophylactic therapy following valoctocogene roxaparvovec loss of efficacy, this would be to the same prophylactic treatment as in the comparator arm of analysis (i.e., if the comparator is FVIII, the follow-on treatment should be FVIII; and if the comparator is emicizumab, the follow-on treatment should be emicizumab). • In future models, to better reflect the clinical reality, ICER could consider valoctocogene roxaparvovec follow-on therapy to be a mix of prophylactic therapies representing the market basket. In the comparator arm, treatment switches due to non-adherence issues should also be incorporated in the model.	
6.	The analysis would be most useful and relevant to the stakeholders when the doses and costs of prophylactic FVIII modelled reflect current clinical practice. For the base case analysis, prophylactic FVIII doses used in the model were based on clinical trials (Advate for standard half-life treatment: 80 IU/kg/week; Eloctate for extended half-life [EHL] treatment: 78 IU/kg/week). However, the report also acknowledges that real-world dosing is often 40–50% higher than those used in clinical trials (Advate: 118.2 IU/kg/week; Eloctate: 111.2 IU/kg/week, based on recent market data from the American Thrombosis and Hemostasis Networks [ATHN]). More importantly, ICER's own summary concludes that the cost-effectiveness results were completely reversed when the model incorporated FVIII doses that are more representative of current use in the US, with valoctocogene roxaparvovec dominating prophylactic FVIII.	We agree with this and have changed the base case to reflect higher doses seen in the ATHN data and efficacy for factor VIII more plausibly related to the higher doses.
7.	The SPINART trial from which the Advate dose was based is a global trial completed over a decade ago and does not reflect the current practices in the US as observed in the ATHN data set. Although there are limited efficacy data at these real-world doses, observations from the non-interventional study conducted prior to the HAVEN 3 emicizumab trial and a recent analysis of EHL products indicate that continued bleeding events occur at real world doses, including doses similar to those observed in the ATHNdataset.	The base case will be changed to reflect higher doses seen currently in practice.
8.	In addition, the analysis used a one-time Wholesale Acquisition Cost (WAC) price of $2.5M as a placeholder price for valoctocogene roxaparvovec. However, for FVIII therapies and emicizumab, ICER attempted to estimate net revenue to the manufacturer of current chronic therapies based on Centers for Medicare & Medicaid Services (CMS) published Average Sales Price (ASP) limits, by removing the furnishing fees as well as add-on administration costs. By doing so, the analysis grossly underestimates the true health system impact of these therapies. Beyond Medicare, most people with Hemophilia A are covered under commercial payers who may cover hemophilia therapies under either the medical or pharmacy benefit; reimbursement is more likely to be based on WAC than ASP, and will include additional administrative costs (e.g., furnishing fee) as well.	The report acknowledges that: "As valoctocogene roxaparvovec has not been approved, no WAC or net price estimates are available. We therefore conducted the base-case analysis using a placeholder price of $2,500,000, based on statements from the manufacturer indicating consideration of prices of around $2 million to $3 million per treatment. In the absence of data on usual discounts for gene therapy, we assumed no discounting and used this placeholder for the net price of this treatment." We used ASP minus furnishing fees after discussion with payers and other stakeholders indicated that would be most representative across all payers.
9.	BioMarin recommendations for ICER's updated model and future evaluation of hemophilia A:	We have changed the base case as described above. In addition, we use estimated net

	• To reflect true health system impact and make the report more actionable, BioMarin requests that ICER models "real-world" FVIII usage with potential improved efficacy in the base case, and presents the clinical trial-based FVIII usage in scenario analysis. • In addition, BioMarin requests ICER to use WAC for all therapies in this review to be better aligned with the costing methodology under ICER's value framework and make comparisons across a level playing field.	prices rather than WAC to better reflect actual payments in the health care system.
10.	The cost-effectiveness model is an over-simplification of the clinical course and does not adequately capture the impact of disease and treatment on patients and families. In the current model, patient health-related quality of life (HRQoL)/utility was tied only to discrete bleeding events and surgery. Despite the report acknowledging the substantial burden and adherence issues associated with chronic prophylactic therapies, the analysis did not factor in the impact to patients of novel treatments with respect to reduced treatment burden and improved bleeding. ICER acknowledges that only 50–70% of patients with severe hemophilia A are adherent to prophylactic FVIII regimens due to substantial treatment burden , and valoctocogene roxaparvovec could provide patients with many years of treatment freedom and eliminate the adherence issue entirely. Patients could benefit from such freedom following treatment with gene therapy, which could provide profound patient benefits in terms of career/education choices, recreational activities, anxiety/depression, and overall well-being. While phase 3 data are not yet mature, patients treated with valoctocogene roxaparvovec demonstrated an increase from baseline in total scores—above and beyond the defined clinically important difference of 5.5 points—in the Haemophilia-specific Quality of Life Questionnaire for Adults (Haemo-QoL-A) throughout 4 years in the phase 1/2 clinical trial. , Increase in score is mainly driven by changes in physical function, role function, worry and consequences of bleeding domains, and these results substantiate the broader patient benefits associated with one-time administration with valoctocogene roxaparvovec.	Adherence in the real world may itself be related to underlying severity. It may also lead to switching to Emicizumab and, overall, it would not change the basic conclusions of the model. Nonetheless, we have added language to the discussion.
11.	The analysis also fails to differentiate across worsening joint health states by assuming that the same utility value applies to all Petterson Score (PS) health states. PS is not a patient-relevant endpoint and is not a sensitive proxy for joint health morbidity. ICER considered joint health as a binary outcome (i.e., YES=PS 1–27, or NO=PS 0). This crude assumption fails to differentiate across worsening joint health states by assuming that the same utility value applies to all PS health states, substantially underestimate the benefits of slowing joint health morbidity, which is a bias in favor of treatments with worse bleeding outcomes.	We based the utility scores on a recent, rigorous, peer reviewed published set of results. That study reported differences found across ages but not across the scores between a PS of 1-27. The sensitivity analyses provide further info on this, but the number of bleeds and disutility of bleeds are much more primary drivers of the model results.

12.	BioMarin recommendations for ICER's updated model and future evaluation of hemophilia A: • The model should incorporate a disutility associated with the frequent and burdensome infusions required for prophylaxis with FVIII of -0.0004 per infusion. • The model should differentiate utility values across PS health states to reflect progressive impact of accumulating bleeds on patient HRQoL. , , • More broadly, ICER should include the HRQoL benefit associated with a one-time treatment in ongoing SST value framework adaption. This would be a compelling opportunity for collaboration with stakeholders to develop novel methods and guide future evidence generation in this important area.	We did not find adequate literature based disutility estimates associated with factor VIII infusions. However, we do now report the discounted total number of infusions so that a reader/user of the model could incorporate that if they wished.
13.	Finally, BioMarin cautions ICER to be aware of inherent assumptions made when mapping FVIII activity levels achieved with valoctocogene roxaparvovec to annualized bleeding rates (ABRs) seen in other studies of FVIII prophylaxis in order to attribute ABRs with gene therapy. Some investigators for valoctocogene roxaparvovec have suggested that the hemostatic efficacy of endogenously produced FVIII following gene therapy may differ from what might be expected in patients with mild or moderate hemophilia A, which are associated with mutations. While ongoing clinical trials will provide additional data to answer this question more definitively, we recommend that ICER consider mapping FVIII and ABRs or use the actual ABRs from valoctocogene roxaparvovec clinical trials in any future reviews.	We acknowledge this limitation.
Genentech, Inc.		
1.	Use the real-world FVIII dosing in the base case, as it represents current clinical practice and clinical guidelines. Recommendations: The scenario analysis using real-world FVIII dosing should become the base case or be presented as a co-base case. The efficacy should be estimated based on target FVIII levels, which is consistent with the approach used for valoctocogene roxaparvovec. Specifically, assume FVIII levels at >3-5% to be consistent with clinical guidelines [9], and apply the corresponding annualized bleed rate from den Uijl et al [10].	We have changed the base case to reflect dosing more consistent with current use as seen in ATHN data. We have also changed the efficacy of factor VIII to be consistent with results from patients using the current dosing levels of factor VIII in the US.
2.	Rationale: The base case FVIII dosing referenced from the SPINART trial [11] is not consistent with real-world dosing, the Advate prescribing information, and clinical guidelines [9,12-14]. Therefore, the real-world dose provided by the American Thrombosis and Hemostasis Network of 118.2 IU/kg for Advate and 111.2 IU/kg for Eloctate should be used [1]. Doing so aligns with The World Federation of Hemophilia Guidelines that recommend prophylactic treatment targeting higher FVIII trough levels (>3-5%+) [9]	Same as above.
3.	The current base case substantially underestimates FVIII utilization and consequently, incorrectly lowers the total cost of therapy, thus overestimating the cost-effectiveness profile of FVIII prophylaxis [1,15]. Given the large discrepancy between the scenario and base case analyses, there appears to be high uncertainty in the results, bringing into question the validity of ICER's base case analysis. Using the real-world dose of FVIII is aligned with standard US clinical practice and allows healthcare decision makers to appropriately interpret the value of therapy.	See above.

4.	Describe the methods and results of each model separately to avoid an inappropriate comparison of results across models. Recommendations: (1) The methods and results for the valoctocogene roxaparvovec adult model (Model 1) and the Hemlibra child model (Model 2) should be presented as separate sections in the report. (2) Distinct tables and figures, with footnotes denoting each model as an independent analysis, should be added to the report. Rationale: The current report presents all figures, tables, and text for both models jointly, which invites comparisons that should not be made. Moreover, three different frameworks are deployed in conducting the assessments with unique underlying context. In attempting to fit these distinct models into one report, aspects of the review are misconstrued. For example, the model structure (Figure 5.1) incorrectly depicts the populations entering the models at the same Pettersson Score, when in fact they are different [1]. Implications: By presenting the model results side-by-side, readers may erroneously interpret the results as being from the same model, and miss the important differences such as levels of underlying evidence and the patient populations. Readers also may make incorrect cost-effectiveness comparisons between Hemlibra and valoctocogene roxaparvovec, when no such comparison exists. Broadly speaking, stakeholders with a lay knowledge of cost-effectiveness, such as the press, may make comparisons that are not valid and can lead to unintended consequences with regard to patients' access to therapy.	We agree and have edited to provide greater distinction between the two models.
5.	Include an in depth discussion of the therapeutic benefits that are important to PwHA within the clinical and economic evaluation. Recommendations: ICER should more thoroughly discuss the benefits of the interventions on lifestyle and life decisions, patient preferences, and FVIII inhibitor development. In addition, the Report-at-a-Glance and Executive Summary should also contain this information.	We believe having an entire early section (section 2 of the report) addresses this issue earlier and better than the suggestion here. We remain uncertain how emicizumab prophylaxis impacts inhibitor development but agree that this is an important issue.
6.	Rationale: Section 2.2 describes restrictions of hemophilia such as career flexibility, education choices, physical activity, family structure, and geographic mobility. ICER's previous 2018 hemophilia A review highlighted the positive impact of less frequently dosed therapies on these restrictions [16]. However, the potential impact of therapy on these concerns is not addressed in the current clinical or economic evaluations. In a rare disease with a high patient burden, an intervention's potential benefits and contextual considerations on patient-relevant outcomes such as patient preferences and quality of life should be emphasized [17,18].	We focus somewhat more on issues around adherence and the decision to take prophylaxis in this report because prophylaxis with factor VIII is substantially less burdensome, more effective, and less restricting than prophylaxis with bypassing agents. We have, however, added information on this to potential other benefits.
7.	An additional important patient-relevant outcome is Hemlibra's potential to delay inhibitor. FVIII inhibitor development has historically been an unpredictable and burdensome treatment complication. A sizable portion of PwHA (~30%) [9,19] will develop inhibitors and this population will face higher costs [20]. There is strong clinical rationale for Hemlibra delaying FVIII inhibitor development. This outcome is linked to FVIII exposure, which is lowered in PwHA treated with Hemlibra [7,21]. Published economic models have accounted for the risk of FVIII inhibitor development [7,8] and should be leveraged and weaved into the overall discussion in this report	As discussed above, we agree that this is potentially a very important issue in judging the value of emicizumab. Unfortunately, we currently have no evidence on whether prophylaxis with emicizumab increases or decreases the risk of developing inhibitors, heard conflicting judgments on the issue from experts, and note that there is sufficient equipoise that an IRB approved a randomized trial (cited in the report) to examine the question.

8.	Furthermore, one of the Comparative Effectiveness Public Advisory Council voting questions focuses specifically on "Potential Other Benefits and Contextual Considerations''. By expanding the discussion of patient-relevant intervention impacts, the voting members will be better prepared for a robust discussion. Implications: Incorporating these important considerations and outcomes will increase the relevance of this assessment, particularly to PwHA. Notably, including these considerations would improve the face validity of the analyses by addressing well-documented concerns and key complication of the disease.	We appreciate this comment and note that Council members will have such public comments available to them prior to the meeting.
9.	4) Increase transparency by ability to modify the shared Excel model. Recommendation: The Model Transparency Program should allow users to modify model inputs, and run scenario and sensitivity analyses. Rationale: The intended purpose of model sharing is to facilitate feedback on the draft report. The modeling good research practices from the International Society for Pharmaceutical Outcomes and Research recommend a model is transparent and can be validated to increase confidence [22]. This is achieved by allowing relevant parties to review a model's structure, equations, parameter values, and assumptions. However, many of the parameter values in the model (e.g., therapy switch probabilities) are locked, and screen protections in critical trace and calculations tabs prevent formula auditing and important pressure testing of alternate inputs (e.g., time-dependent efficacy). This presents challenges in running critical scenarios that assess the accuracy and validity of the model (i.e., sensitivity and scenario analyses cannot be replicated). Manufacturers, like Genentech, can and have provided accuracy checks that ensure end tools are error-free and clinically valid. Implications: A model with the ability to run scenario and sensitivity analyses will enable stakeholders to better understand and quality check the analysis, strengthening confidence in the results.	Draft models are shared with manufacturers to provide greater transparency into the model structure, inputs, and assumptions, with the goal of enhancing feedback on our draft reports. It is not generally possible to provide models with fully interactive user interfaces at this stage of model development, but the model we shared allowed for changes to a variety of inputs related to clinical, cost, utility, and mortality parameters, with changes reflected in the results.
Sanofi		
1.	Section 4. Comparative Clinical Effectiveness 1) Network Meta-Analysis (NMA): The NMA generalized all FVIII treatments based on a single study (SPINART- NCT00623480) representing a single FVIII product [rFVIII-FS: Kogenate® FS Antihemophilic Factor (Recombinant)].2 Based on the pharmacokinetic and dosing profile, this agent is categorized as a standard half-life (SHL) FVIII product.3 More recently available FVIII agents, categorized as extended half-life (EHL) products, contain modifications to their structure that allow for extended duration of action and less frequent dosing, which may affect the patient experience and treatment adherence.	In this revised Evidence Report, we decided the correct comparator for emicizumab is factor VIII prophylaxis at the doses that are typically used in the US today. For reasons presented in the report, this has led us to downgrade the rating for the comparison to C++. Given that we have made this change, we are choosing not to address every individual issue raised below, but we believe that the B+ rating is correct when comparing emicizumab with the doses of factor VIII used in SPINART.

2.	Using rFVIII-FS as a proxy for the entire FVIII drug class is a misleading assumption that fails to distinguish significant differences between both the EHL and SHL classes, as well as the individual drugs themselves. Iorio, et al. illustrated these differences when conducting an indirect treatment comparison (ITC) of rFVIII Fc fusion protein (rFVIIIFc) and conventional, or SHL, rFVIII products in hemophilia A patients. They found that prophylaxis with rFVIIIFc may be associated with improved bleeding rates and lower weekly factor consumption than more frequently injected rFVIII products. When comparing to rFVIII products with similar bleeding rates, the authors found that rFVIIIFc may be associated with lower weekly FVIII consumption rates while requiring fewer prescribed injections.7 Additionally, the presence and method of collecting joint health data further differentiates FVIII agents. For example, Kogenate assessed joint health using index joint magnetic resonance imaging (MRI) scores of osteochondral damage, while rFVIIIFc collected joint health data for both children and adults in the A-LONG and follow-on ASPIRE studies, leveraging the Hemophilia Joint Health Score (HJHS) for children under 12 years old, and the modified HJHS for use in population aged 12 or older.	See above.
3.	Therefore, the generalization between EHLs and SHLs, and the exclusion of data that differentiates each product's unique profile, may lead to the inaccurate interpretation of results. Given this significant limitation, ICER should clearly state in its report, that the results of the clinical comparison to emicizumab, and the corresponding evidence rating, are only applicable to rFVIII-FS alone. Additionally, as the results of the NMA are incorporated into the cost-effectiveness model, the output of that analysis should also only be applicable to rFVIII-FS and should be stated as such	See above.
4.	Comparative Net Health Benefit The conclusions in the comparative net health benefit section are misleading and inaccurate, and do not align with the findings of the NMA. The NMA results show that there is NO significant difference between emicizumab and rFVIII-FS when evaluating treated bleeds and treated joint bleeds. However, the summary results of the comparative net health benefit section of the Evidence Report do not align with this finding. The authors conclude that there is a difference simply because the bleeding rates were numerically lower for emicizumab. Additionally, as only treated bleeds (including treated joint bleeds) were the only bleeds evaluated, the statement "…appears to have lower bleeding rates (of all types) compared with factor VIII..." is incorrect.	See above.

5.	Section 5: Long-Term Cost Effectiveness 1) Limitations of Pettersson Joint Health Score The utilization of the Pettersson joint health score warrants a disclosure of the limitations of this measure. While the first ICER modelling approach accurately captures possible bleeding stages (overall vs target joint / treated vs untreated) and the subsequent consequence of repeated bleeds (arthropathy estimated via Pettersson joint score), the new proposed structure focuses heavily on tunnel stages of arthropathy using the same score.11 We agree that the use of the Pettersson score in economic models represents the most pragmatic approach to assess joint damage, considering the limitations of existing endpoints/measures. Stages built entirely using this score presume however that a) point categories in the score are related to a specific cost and utility of patients using literature proxies (PS = 0, PS 1-27 or PS = 28), which contrasts with the natural intuition to attribute costs and disutilities to each bleed type as occurring in real clinical setting b) no distinction between joint and non-joint bleeds and c) treated and untreated arthropathy patients have a similar risk to develop degenerative arthritis, which neglects the clinical benefits of early treatment of chronic hemarthrosis	The model tracks and issues disutilities for treated target joint bleeds that are different than treated non target joint bleeds. It also bases progression in the PS levels based on joint bleeds. While not perfect, the PS remains the best way to combine the short and long term outcomes associated with bleeding.
6.	Lack of Pediatric Data In the Heterogeneity and Subgroups section of the report, the authors identify that there is insufficient data to conduct a comparison between emicizumab and FVIII therapy in children. However, in the cost-effectiveness model, the age of the cohort begins at one year old. As a large percentage of prophylaxis treatment in Hemophilia A patients starts at a young age, it is understandable that there is a need to evaluate the cost-effectiveness of this population. However, given the limited data comparing FVIII and emicizumab in this population, we do not think that the model should extrapolate the results to include an age group not supported by data. Therefore, we recommend that the model start evaluation at 12 years old.	In general, in the absence of data on a subgroup, we assume similar relative effects in the subgroup to the overall group and that is what we have done here. The FDA label approves emicizumab for use in newborns. Trials of emicizumab in younger children with inhibitors show results at least as good as results in adolescents and adults.
7.	Use of Rurioctocog alfa (recombinant factor VIII: Advate®) and recombinant factor VIII Fc (rFVIIIFc) (Eloctate®) as "representative" of the standard half-life (SHL) and extended half-life (EHL) classes, respectively. Given the fact that the comparative effectiveness / NMA components of this evidence report are based on data from one clinical trial and a single SHL FVIII therapy [rFVIII-FS: Kogenate® FS Antihemophilic Factor (Recombinant)], the cost-effectiveness model should only reflect that specific FVIII therapy. To extrapolate clinical findings from one factor therapy to two entire classes of FVIII agents discards the wealth of data that differentiates these products. Furthermore, the inclusion of drug costs and dosing data from two separate agents, not included in the comparative effectiveness assessment, is misleading and erroneous.	The drugs were chosen as they are relatively highly prevalent drugs in their class. Further we based the average doses on representative ATHN data. We acknowledge that there is high variance in drug dosing and part of the sensitivity and scenario analyses illustrate the potential variance from changes in those.

8.	Inclusion of real-world dosing data into the cost-effectiveness modeling scenarios While real world evidence (RWE) is an important component of any assessment of drug therapy, the arbitrary inclusion of dosing data on Advate and Eloctate should be removed for several reasons, specifically: As stated above, these agents were not part of the comparative effectiveness analysis	The base case will be changed to reflect higher dosing seen in the ATHN data and bleed rates seen in practice in patients using doses of factor VIII at current levels. Data to allow specific dosing costs and efficacy across a broad spectrum of products are not available. We do conduct sensitivity analyses and scenario analyses to illustrate the potential variance in the model.
9.	There is no mention of planned incorporation of real-world dosing data in the modeling analysis plan (MAP) for this evidence report .12 The MAP also states that the HAVEN 3 trial will be used as the source for representative treatments, originally listing a breakdown of 87% SHL and 13% EHL. This ratio differs from the 71.18% SHL and 28.82% SHL used in this draft evidence report.	The MAP reflects initial thinking and strategy for the modeling that has changed based on feedback from several stakeholders and reviewers and internal discussions.
10.	Per ICER's own recommendations and publications on inclusion of RWE into coverage decisions, there needs to be a transparent and systematic process for evaluation and utilization of such data.13-15 There was no such process followed in regard to the inclusion of the unreferenced, American Thrombosis and Hemostasis Network (ATHN) dosing data. A comprehensive literature search would have identified many RWE studies on FVIII therapy demonstrating a range of doses. 10, 16-19 While of good quality, the ATHN data should not be considered as reflecting the typical treatments and dosing patterns for all Hemophilia A patients. The data from ATHN is representative of a hemophilia treatment center (HTC) focused population of patients. Dosing of FVIII therapy can be impacted by many things such as age (child/adult), severity of hemophilia, prophylaxis/on-demand, inhibitor status and other patient related factors.	We have provided additional details on the ATHN dosing data used in the evidence report. We have also compared those data with other RWE FVIII dosing studies in the literature.
11.	If RWE is to be incorporated into these analyses, all outcomes should be evaluated for inclusion, not just dosing. For example, real-world data on clinical effectiveness, safety and costs could considered for inclusion. Regardless of the type of data included, an open, systematic and comprehensive review is required.	We have included RWE on effectiveness of higher doses.
Takeda Pharmaceuticals		
1.	Deviation from a priori decision to use random effects model Takeda agrees with ICER's a priori decision in the protocol to prefer the random effects network meta-analysis (NMA) due to differences in the study populations. However, results from the fixed effects model were used to estimate uncertainty in the sensitivity analyses and seemingly used to interpret ICER's evidence grade for emicizumab. While the point estimates for the rate ratios for bleeds are equivalent under the fixed and random effects models, the estimated standard errors are lower (and the resulting confidence intervals are narrower) under the fixed effect model. In this particular case, the conclusion regarding statistical significance differs between the models. Since the random effects NMA was decided as the preferred model a priori, Takeda recommends that ICER utilize the credible intervals from the random effects model when conducting the one-way and probabilistic sensitivity analyses and also consider the lack of statistical significance when determining the evidence rating. Because of the wider intervals with the random effects NMA compared to the fixed effects model, uncertainty in the model-based results should be greater than currently reported.	In this revised Evidence Report, we decided the correct comparator for emicizumab is factor VIII prophylaxis at the doses that are typically used in the US today. For reasons presented in the report, this has led us to downgrade the rating for the comparison to C++. Given that we have made this change, we are choosing not to address every individual issue raised below, but we believe that the B+ rating is correct when comparing emicizumab with the doses of factor VIII used in SPINART.

2.	Underreporting of fundamental heterogeneity and lack of consideration of uncertainty in determining strength of comparative evidence While HAVEN 3 and SPINART , are well-designed studies, they have fundamental differences in design, populations, and outcome definitions. Since SPINART was specifically investigating secondary and tertiary (late) prophylaxis, it included a patient population that was not on regular prophylaxis therapy for ≥12 months over the previous 5 years.2,3 This requirement alone creates a population that is likely to be systematically different than a population that does not have that requirement. For example, patients included in SPINART may have poorer control over their bleeds due to foregoing any proactive continuous prophylaxis and/or they may have compromised joint integrity due to the nature of more frequent bleeding observed with episodic therapy which may lead to more recurrent bleeds during the investigation period. These differences introduce the risk of bias and uncertainty, which is not currently discussed in length in the draft report. Takeda recommends that several aspects of heterogeneity be highlighted in the report and how that impacts the strength of conclusions that may be drawn.	We disagree with the comment that the two studies are different in the patient population enrolled. Similar to SPINART trial, the randomized arm of the HAVEN 3 trial included in the NMA included patients who were not on prophylaxis prior to the start of the study. Furthermore, SPINART was similar to HAVEN 3 in terms of baseline characteristics, study design, and outcome definitions. Please see Table 4.2 in our revised report.
3.	Patient subjectivity involved in outcome reporting, differences in outcome definitions, and impact of differences in follow-up periods. A critically important issue we identified is the use of 'treated bleeds' as the outcome of interest instead of 'all bleeds'. This methodological decision is understandable considering the healthcare sector perspective. However, it limited the number of FVIII pivotal trials that met the inclusion criteria, thereby increasing uncertainty of bleed estimates for FVIII products. Treated bleeds were not reported in SPINART's 1-year follow-up and were only reported in the 3-year follow-up.2,3 While the population in the two follow-up analyses are the same, their bleed rates differed at the different follow-up times. In the 1-year follow-up analysis of SPINART, the mean total number of bleeds (standard deviation [SD]) were 2.0 (4.5) while the mean treated bleeds (SD) in the 3-year follow-up were 2.5 (4.7). Rate of total bleeds would be expected to be higher than treated bleeds, considering total bleeds include both treated and untreated bleeds, which was not observed. This may illustrate the progressive nature of severe hemophilia A or loosening of heavily controlled pivotal trial requirements during long-term follow-up. Regardless, it highlights the impact follow-up time may have on bleed outcomes and how it may be inappropriate to assume annualizing bleed rates adequately control for differences in follow-up times between trials (HAVEN 3 [24 weeks] vs SPINART [3 years]),	The SPINART trial reported number of treated bleeds and not total bleeds. According to the authors, bleeding event was defined as any episode of external bleeding (i.e., epistaxis), bruising, pain or limited function for which FVIII was infused. A joint BE (subset of total BEs) was defined as an event with pain, swelling, tingling, warmth or limited motion of an extremity for which FVIII was infused. Electronic diaries, (similar to what was done in HAVEN 3) were used to record infusion and bleeding data. Also, the mean treated bleed at 2 years vs. 3 years was 2.2 (SD:5.1) vs. 2.5 (SD:4.7). Please see Manco-Johnson Corrigendum 2013.
4.	Therefore, in order to include more evidence from FVIII trials and utilize more comparable follow-up periods, we recommend either the use of all bleeds to conduct indirect treatment comparisons, or more transparent reporting of the limitations and uncertainties inherent when comparing 1) bleed outcomes based on subjective patient decision to treat a bleed and 2) bleed rates over differing follow-up periods.	See above.

5.	Selection bias in HAVEN 3 intra-individual comparison confounding its use to support evidence rating1 HAVEN 3 intra-individual comparison of patients who previously participated in a non-interventional trial may not be accurate and reliable and is not an appropriate source of evidence to support differences in efficacy between emicizumab and FVIII replacement therapy. The non-interventional trial design is at very high risk of selection bias as investigators may have selected participants for inclusion who were poorly controlled, experiencing frequent bleeds while on FVIII therapy, and/or who would be thought to benefit from switching to emicizumab compared to the overall population . Kruse-Jarres, et al. raise this important consideration in their discussion4. Lack of adherence observed in the non-interventional trial support the presence of selection bias; this population may represent patients who were unsuccessful on FVIII treatment as opposed to the overall population of interest in this review.	See above.
6.	Takeda believes the B+ evidence rating for emicizumab ignores the inherent heterogeneity within the indirect comparison, lack of uniform outcome definition, and lack of statistical significance in the random effects NMA. Based on the ICER Evidence Rating Matrix, a B+ rating assumes the lower end of the confidence interval does not extend into the comparable range. While the "comparable range" was not defined, the confidence intervals for the annualized treated bleed rate ratio and the annualized treated joint bleeds rate ratio for emicizumab vs FVIII prophylaxis both include 1.0 (see Tables 4.4 and 4.5 from the draft evidence report, respectively). Thus, one cannot conclude that the two treatments differ with any degree of certainty. Instead, it seems more reasonable to declare a high certainty of a "comparable" efficacy rating based on the consideration of the cumulative topics in this section. Therefore, Takeda recommends reconsideration of the B+ rating for a lower rating, as evidence illustrates comparability and lacks strong evidence of any differentiation of efficacy.	See above.
7.	Scenario Analysis 'Representative' Label FVIII utilization for prophylaxis varies greatly based on individual patient characteristics. Stating that one dose is more 'representative' than another ignores the spectrum of utilization across patients. A recent study investigating real-world FVIII consumption illustrates the large variability in weekly FVIII consumption among patients with severe hemophilia A on prophylaxis. Additionally, evidence demonstrates that adult patients consume less FVIII than pediatric patients. Therefore, weekly FVIII consumption may decrease over time. Takeda understands this is a population level analysis and averages are typically used to help make population level policy recommendations; however, the wide variation of FVIII utilization among patients may make it inappropriate to make 'one-size-fits-all' interpretations. Takeda recommends relabeling this scenario as an 'alternative-utilization scenario' rather than one that is "more representative" to prevent confusion that a single dosing average represents an entire population. Payers should make policy decisions regarding the role of FVIII for patients based on individual utilization of FVIII. Stating the cost-effectiveness results based on a FVIII dose of 118.2 and 111.2 IU/kg/week as "more representative" may have significant economic consequences if the scenario analysis is interpreted as 'the truth' and is applied to all patients regardless of their actual FVIII utilization.	We believe the dosing seen in this scenario is more representative of current utilization of factor VIII in the US and have received feedback from several reviewers that this should be the base case. We plan to use the higher doses seen in the ATHN data along with efficacy estimates from a recent analysis of patient outcomes in US hemophilia treatment centers. We will add language to the discussion regarding the variance of doses. We also conduct sensitivity and scenario analyses to illustrate the potential variance in outcomes associated with that type of variance.

8.	Also, it is understood the American Thrombosis & Hemostasis Network was used to determine alternative dosing averages, however, it is largely unclear how the 118.2 IU/kg/week for ADVATE and 111.2 IU/kg/week for ELOCTATE were derived. For example, was FVIII utilization for breakthrough bleed treatment included in the average weekly dosing and what was the mean age? Takeda recommends improved reporting of scenario analysis dosing elicitation to better understand the methods and to promote replicability	The doses come from an average of initial prophylactic treatment regimens in the ATHN data, which we will add into the report. We also will edit to maximize transparency in the model inputs.
9.	Contradicting Conclusion Raises Confusion about 'Correct' Interpretation While the "Summary and Comment" section of the report begins each comparison with the base case result, the emphasis quickly shifts to highlight the results of the scenario analysis which have contradicting results. Interpreting both conclusions independently without explaining the individualized and dynamic dosing of FVIII may lead to confusion and misunderstanding of the results, especially for an audience that may not be technically trained in health economics. Thus, Takeda recommends that ICER consider presenting the conclusions in a more balanced fashion and further explain the dynamics of variability in dosing and how that may impact conclusions to prevent confusion on which results are 'correct'.	We are changing the base case as described above, but also will include as a scenario analysis what was the base case in the draft report. We will also include language in the discussion and conclusion reflecting the implications of that scenario relative to what will be the new base case.
Novo Nordisk		
1.	Relying on the SPINART study to represent efficacy for all FVIII products does not acknowledge innovations over the past decade which have led to lower ABRs with FVIII products. Novo Nordisk's understands the logic that ICER used to identify the SPINART study for the NMA, however, the issue is that the SPINART study was conducted from 2008 to 2011 with a SHL FVIII dosed at 25IU/kg, 3x per week.1 Over the past decade, hemophilia treaters have identified other dosing strategies to reduce ABRs and new EHL products have been improved with superior pharmacokinetics. Both have resulted in improvement in bleed protection which is not reflected in this ICER report. To highlight this point, Esperoct is the latest product to be available in the US which achieved an average ABR of 1.22 (slightly less than that seen with emicizumab in HAVEN 3) and well below the 4.56 ABR used for FVIII products in the modeled base case. Again, we understand why SPINART has been chosen, but we ask ICER acknowledges in the Uncertainty and Controversaries section of the Evidence Report that SPINART is older data and results may have been different if other sources of FVIII efficacy were used in this review.	In this revised Evidence Report, we decided the correct comparator for emicizumab is factor VIII prophylaxis at the doses that are typically used in the US today. For reasons presented in the report, this has led us to downgrade the rating for the comparison to C++.
2.	The Scenario Threshold Analysis where real-world dosing of FVIII products is compared to trial dosing of valoctocogene roxaparvovec and emicizumab is limited in its applicability in that it does not compare apples to apples. While real-world dosing is available for FVIII products, we do not yet know how emicizumab is dosed in the real world, and given the concerns about durability, it is unclear what the real world cost of care will be with valoctocogene roxaparvovec. Therefore, if ICER continues to keep the Scenario Threshold Analysis in the Evidence Report, it is suggested that ICER adds this point to the Limitations section of the Evidence Report.	We will add language to the limitations section regarding this issue particularly since we are now making the higher dose of FVIII the base case.

3.	Additionally, given higher real-world doses compared to clinical trials are being used in the Scenario Threshold Analysis, it is likely that the ABRs would be lower. While ICER, addresses this on page 77 by stating "We are uncertain of the added efficacy of these higher doses, but even if these doses completely eliminated all bleeding events (and thus had greater efficacy than emicizumab), emicizumab would remain cost effective." While that may be true, ICER could still use real-world bleed rates in the base case for this Scenario Threshold Analysis as it has implications for the conclusions made. Malec and colleagues (2020) analyze the ATHN dataset and find an average 1.3 ABR for all patients receiving FVIII (1.0 and 1.9 for EHL and SHL, respectively).3 Using these ABRs may likely change the conclusion on page 66 such that while valoctocogene roxaparvovec and emicizumab would still be projected to save costs, QALYs may not be higher in versions 1 and 2 of the model. Novo Nordisk asks that ICER include this in the model and update this conclusion	We appreciate this suggestion, and we are now using the Malec paper results in our base case model.
4.	3. The Scenario Threshold Analysis utilizes dosing from ATHN (118.2 IU/kg every week for Advate and 111.2 IU/kg every week for Eloctate) which may not be representative of other FVIII products. ICER has used Advate and Eloctate as 'representative treatments of each type, and typical doses for those products' (page 53). While they may be the most prescribed SHL and EHL FVIII products, their labeled dosing regimens are anything but typical for each type of FVIII they represent in this report. Advate and Eloctate both have significant dosing flexibility as per the FDA label which may have contributed to the higher real-world doses derived from ATHN and may not be representative of other FVIII products.	We recognize that there is high variance in the potential dosing of these and other products. We are moving forward using these products for the base case given their prevalence within these major classes of treatments. We also are conducting a full set of sensitivity and scenario analyses to help characterize the potential variance in the results.
5.	Advate is recommended to be dosed at 20 to 40 IU/kg every other day (which is up to 140 IU/kg per week) or every third day with dosing targeted to maintain FVIII trough levels ≥ 1% (allowing for even higher weekly doses). Eloctate is recommended to be dosed in the range of 25-65 IU/kg at 3-5 day intervals (which is up to 152 IU/kg per week). Other EHL FVIII products have less flexibility and lower maximum dosing amounts per week [Adynovate: 40 to 50 IU/kg twice per week (up to 100 IU/kg per week); JIVI: 30 to 40 IU/kg twice per week (up to 80 IU/kg per week) or 50 to 60 IU/kg every 5 days (up to 84 IU/kg per week)] while others have simple, fixed recommended dosing (Esperoct: 50 IU/kg every 4 days = 87.5 IU/kg/week). Therefore, these FVIII products likely have lower weekly doses in the real world than what ICER has modeled in this scenario. In fact, a study was published earlier in 2020 found statistically significant differences in weekly dosing for EHL products.	Again, we recognize that there is high variance in the potential dosing of these products. We believe using the doses seen in the ATHN data set represent a data driven representative basis for projecting drug costs for these classes of medications. We also conduct sensitivity and scenario analyses to help assess potential variance in the results. We will also add language to the discussion section regarding the potential that alternative products in the class could have different comparative results.
6.	To address this issue, Novo Nordisk suggests that ICER publish the weekly doses for each of the FVIII products in the Evidence Report as captured within the ATHN database for readers to assess if the real-world dosing for Advate and Eloctate are indeed representative of other SHL and EHL FVIII products.	We only have the average doses for Advate and Eloctate from that data set. It is beyond the scope of our analysis to analyze every dose of every product. There would also be sample size issues with several of the products.

7.	Novo Nordisk also suggests that ICER discusses the issues that are present when using Advate and Eloctate to represent all SHL and EHL products in either the Uncertainty and Controversaries or the Limitations section of the Evidence Report. Here, ICER could highlight differences in labeled dosing regimens which have an impact on the real-world dosing for each product. Furthermore, in the case of EHLs, ICER could state that the different methods used to extend half-life have had different levels of success when compared to SHL products (ranging from Adynovate's 40-50% half-life extension in adolescents and adults5 to Esperoct's 85% half-life extension in children6) and pharmacokinetic differences between EHLs7 which further explain labeled dosing differences.	We will add language to the limitations section regarding this issue.
8.	Finally, given the implications of this scenario analysis, ICER would best serve readers of the Final Evidence Report by adding more information on how these dosing amounts were derived from the ATHN Dataset. Were FVIII dose and frequency provided by HTCs or patient dairies? Does it represent their prescription or actual product utilization? If it based on prescription, we suggest ICER states if 100% adherence to the prescription has been assumed. As ICER has noted throughout the report, adherence is a factor to consider with hemophilia treatment and Novo Nordisk advices that ICER adjust this real-world scenario for an appropriate adherence rate given this is being represented as real world dosing and cost. Furthermore, ICER could state if these doses are median or mean values. All of this information is important to provide context to the reader so that they can put these findings into context.	We have added further description of the ATHN data particularly as it is now the base case. We do admittedly abstract from adherence in the model as patients may switch or temporarily discontinue medications. Adding potential pathways related to adherence would complicate the model substantially but would not change any of the conclusions related to the base case as can be seen by considering the full set of sensitivity and scenarios analyses that are included in the report.
9.	In summary, we suggest that ICER more clearly informs the reader of the Evidence Report that there are differences between FVIII products, specifically when it comes to EHLs, and using one product to represent all of them, has limitations. Given the approach to use representative products for SHL and EHL, we suggest that ICER rethink the use of blanket statements defining the value of all SHL and EHL products when only Advate and Eloctate were included in this model. At a minimum, ICER could note this in the Uncertainty and Controversaries or Limitation section of the Evidence Report.	It would be helpful if manufacturers of FVIII products generated high quality evidence from head-to-head randomized trials showing whether there are or are not important differences in safety and efficacy.
Other		
Analysis Group, Inc.		
1.	In identifying model inputs for FVIII dose, we felt that matching real-world dose was critically important. Our dose of 40 IU/kg three times per week reflects the median from a recent analysis of the ATHNdataset (Croteau et al 2019), which reflects dosing among a large number of adult patients managed with standard half-life prophylaxis. Our model base case compares valoctocogene roxaparvovec to standard half-life FVIII products and not to Eloctate as was stated in the draft report summary. This dose does differ from dosing used in the ICER base case (80 IU/kg per week) and is at the upper end of the dose range for prophylaxis on the Advate label (20 to 40 IU/kg, 3 to 4 times per week). However, when patients elect treatment with gene therapy, they are able to forgo regular use of FVIII at doses that reflect real-world usage. The choice is not between valoctocogene roxaparvovec and FVIII prophylaxis dosed per clinical trials; rather, it is between valoctocogene roxaparvovec and real-world use of FVIII prophylaxis. Thus, we felt that real-world dosing for FVIII was the most appropriate model input.	We are now using the higher doses from the ATHN data as our base. As such we are also now using upper bound efficacy results for Factor VIII from the Malec study.

2.	We fully acknowledge that our choice to use the median wholesale acquisition cost (WAC) of standard half-life products means that costs in our model do not reflect the cost of a specific standard half-life product to any single payer. However, our model projects that valoctocogene roxaparvovec would be cost-saving on a lifetime basis as compared to standard half-life products even at the substantially lower per IU cost of $1.08 per IU used in the ICER base case analysis. Likewise, we chose to use a $2 million per patient average cost for valoctocogene roxaparvovec based on prices of currently marketed gene therapies in other therapeutic areas. However, our model also projects that valoctocogene roxaparvovec would be cost-saving even at the $2.5 million price point used in the ICER draft analysis.	ICER's analyses used an estimate of net prices rather than WAC for FVIII as more reflective of actual payments in the health care system. We chose a placeholder price of $2.5 million for valoctocogene roxaparvovec based on the range of prices mentioned in statements from the manufacturer. Note that our base case has changed from the draft report, as detailed above.
3.	In summary, the main conclusion of this draft report is extremely sensitive to the assumed dosing of FVIII: in fact, it switches from valoctocogene roxaparvovec being highly not cost-effective to being dominant (and hence preferred). We do not see a good reason for not using real-world (vs. labelled) dosing.	We are now using the ATHN based dosing as the base case. The dependence of the result on the dosing is also included in the discussion.
4.	Our inclusion of a 0.0004 disutility per infusion was an attempt to capture what some patients perceive as a burden associated with FVIII prophylaxis treatment. While the draft report background section describes FVIII prophylaxis as “burdensome” and notes the required intravenous access is “difficult to master and painful”, there was no attempt to quantify the decrease in quality of life associated with frequent intravenous infusions or to include it in the cost-effectiveness model. Patients treated with valoctocogene roxaparvovec (and to a lesser extent with emicizumab) require a much smaller number of infusions over a lifetime, and the effect on quality of life is almost certainly non-zero. We did struggle to identify an appropriate value for the size of the disutility associated with infusions, and given the lack of hemophilia A-specific evidence, we used an estimate of injection disutility from another therapeutic area (Matza 2015). A more robust value for the input would be desirable, but we feel that including an imperfect estimate of the disutility associated with infusions was more appropriate than ignoring the impact of infusions on quality of life.	As we do not have adequate literature based estimates of a disutility of infusions in this patient population, we are still not including that in the model. However, we do now report the number of infusions projected in the model and we have added language to the discussion regarding this issue.
5.	Now we would like to highlight two additional differences between our modeling approach and ICER’s, which have significant impact on model outcomes. First, our model assumes that the cost of valoctocogene roxaparvovec will vary by patient weight. In particular, the heaviest patients will have a higher cost of valoctocogene roxaparvovec and also will have a higher cost of FVIII prophylaxis while the lighter patients will have a lower cost of both valoctocogene roxaparvovec and FVIII prophylaxis. In the ICER model, the cost of FVIII prophylaxis is linked to weight, while the costs of valoctocogene roxaparvovec is assumed to be $2.5 million per patient regardless of weight.	We are assuming in the model that valoctocogene roxaparvovec is given once to patients at the age of 18 and as such our drug cost can be viewed as being relative to a patient with a weight equal to the average 18 year old in the US.

6.	Second, in the draft ICER model, patients treated with valoctocogene roxaparvovec who experience a loss of response are assumed to initiate treatment with emicizumab. Our model instead assumes such patients would resume FVIII prophylaxis. Under our approach, the cost of treating patients in years after valoctocogene roxaparvovec loses effectiveness is identical to the cost of treating patients who never received valoctocogene roxaparvovec during those same years. Thus, our model is not very sensitive to variation in the model time horizon once the model horizon exceeds the typical duration of assumed valoctocogene roxaparvovec effectiveness. In the model presented in the draft report, however, the assumption of future treatment with emicizumab combined with the substantially higher treatment cost of emicizumab (compared to FVIII prophylaxis) means that model is very sensitive to time horizon. This finding is illustrated in the scenario in which patients receive valoctocogene roxaparvovec at age 40 instead of at age 18.	Given the enormous differences in costs at more currently representative doses of factor VIII, and particularly now that the ATHN doses are the base case, we believe it makes the most sense from an efficiency standpoint for patients to switch to emicizumab.
European Haemophilia Consortium		
1.	Report Readability In the report, having two models in single tables in section 5 may create some misunderstanding of the comparisons, especially as the FVIII prophylaxis models are different. Model 1 and 2 deal with slightly different cohorts, so direct comparisons are not possible. At first glance, a reader may misinterpret which model is the base case. For clarity, it would be better to split the tables dependent on the model used for each drug and base case. Alternatively, if the tables are to remain together, it would be preferable to have a single FVIII base case, rather than two separate FVIII base cases, which is confusing for the reader.	We agree and have made efforts to distinguish the two models and the corresponding results. We now refer to them as model 1 and model 2 and have separated the results.
2.	Additionally, in the PICOTS, there was an indirect comparison mentioned in terms of evidence for Emicizumab and Valoctocogene Roxaparvovec, but this has not been described in the model. It would be interesting to report this in the results section, or alternatively, removed from the early section. The former option would be beneficial in understanding the differences in benefits for each type of treatment.	Given the conclusions in the clinical section we are not comparing emicizumab and valoctocogene roxaparvovec in the economic modeling.
3.	Patient Relevant Data In the clinical trials of both treatments being assessed, Haem-A-QoL, a disease specific quality of life measure for haemophilia, was recorded and has been reported. While utility scores are beneficial in our community's understanding of the impact of such treatments, it would give greater context to the discussion if ICER were to consider reporting disease specific outcomes.	Thank you, we have expanded our discussions of these data.
4.	In the 2019 ISTH HAVEN study abstract1, data demonstrates that the proportion of participants with no missed workdays increased to ≥90% with Emicizumab prophylaxis in both HAVEN 3 and HAVEN 41. This may be a result of patients having extended periods without bleeding, and it could be included in the model under societal costs for Emicizumab, and may potentially be considered in those patients in the Valoctocogene Roxaparvovec model. If this is not possible, further expansion of the descriptive analysis should be provided.	We have literature based consideration of workdays associated with bleeds already in the model in the societal perspective scenario for the emicizumab model and in one of the dual base cases for the valoctocogene roxaparvovec model.
5.	In terms of patient relative data, there are two areas that we would like to see reported. The first is frequency of infusion. There	We will report model based estimates of the number of infusions associated with factor

	is a cohort of patients with needle phobia and poor venous access, which can have implications on adherence. The number of infusions required for a life-time of prophylaxis significantly impacts patients' quality of life2. It should be possible to report these from the model and give a better indication of the number of infusions that are avoided; whether they be subcutaneous or intravenous infusions, this is still several hundreds of infusions avoided in either case.	VIII products. We do not feel we have high-quality data looking at a causal relationship between number of infusions and quality of life impact.
6.	The second request is to consider using a pharmacokinetic model for the FVIII dosing, and report the time over a life-time that a patient spends above 10-12%. The rationale for this request is based on Den Uijl et. al. work, demonstrating the significantly reduced likelihood of bleeding above this level3. Signs of this are also seen in the ISTH HAVEN study abstract1. The combination of both of these would allow fairer comparisons when considering how changing to these treatments, or even remaining on a patient's current treatment, might impact their individual quality of life for the future. If possible, these might be considered by ICER when reviewing this draft.	A pharmacokinetic model is beyond the scope of our analyses. The base case doses and bleed rates reflect those seen in US hemophilia treatment centers. The base case model is driven by average levels that by construction would be similar if we had done a pharmacokinetic model and there are a variety of sensitivity and scenario analyses to address potential variance in the results. Though the exact distribution of results from a pharmacokinetic model may be different the average results and overall levels of variance would likely be quite similar.
7.	FVIII Utilization In 2018, ICER correctly identified that the data available on the inhibitor population was severely lacking, and took a pragmatic approach in assessing the current literature on treatment. As a result, ICER produced a model that mimicked the care available to patients in the clinic extremely effectively. In this case, there was just enough evidence defined in the literature review to qualify for a comparison of clinical trials versus randomized control trials such as SPINART. In this report, as a result, the discrepancy between modelled FVIII utilization and the real-world evidence provided by ATHN is significantly different. On this occasion, the lack of assessment of what is currently happening in clinics all over North America and Europe is totally different than the SPINART comparison. Historically, troughs of 1-3% were being targeted in these trials. In the EU and US, with patients having access to extended half-life or standard half-life products at higher doses, this paradigm is shifting significantly. The European Directorate for the Quality Medicines and Healthcare (EDQM) has stated that minimum trough levels of 3-5% be targeted in order to provide protection from joint damage4. EDQM recommendations are recognized guidelines endorsed by the Council of Europe.	We have changed the base case to reflect dosing more consistent with current use as seen in ATHN data. We have also changed the efficacy of factor VIII to be consistent with results from patients using the current dosing levels of factor VIII in the US. We have added the EDQM recommendations to the Guidelines section of the report.
8.	A pharmacokinetic (PK) based model would better demonstrate the reality of current FVIII use in clinics and would provide FVIII utilization closer to that seen in the ATHN dataset. Additionally, this would make the 'scenario' analysis more applicable. Across health sectors, there is an increasing recognition of the importance of personalized medicine. We hope ICER would use the same pragmatic thinking demonstrated in the past, and use more recent publication of FVIII utilization and trough levels, to inform the model.	See the two answers above. Our base case is now using the ATHN doses.
9.	Additionally, reporting the real-world analysis, identified through the ATHN dataset on pages 70-71, is one of the most interesting aspects of the report. As a reader, until this point is reached in the	See answers above. We have changed the base case but also include and highlight a scenario with trial based doses and efficacy.

	report, the message transmitted is that these products are highly un-cost-effective; then in this small section, it is suddenly highlighted that if real world evidence from ATHN is used, they do become cost effective. ICER's model, while not a pharmacokinetic one, is well structured. However, the huge difference between real world and model outputs makes it difficult to believe that the model is a true representation of what might happen if these products are made available. The model could be used in many countries as a reason for not introducing a cost-effective therapy, while in reality ICER identifies that they could be significantly cost-saving. Addressing the FVIII usage in both Model 1 and Model 2 would add validity to outputs on FVIII utilization. We recommend that, at the very least, this section is expanded and an additional scenario analysis be reported, identifying the new probability of these products being cost effective.	Overall, the dosing and highly cost ineffectiveness of higher doses of factor VIII remain a major issue which we discuss.
10.	Inhibitor Development The draft report indicates that de-novo inhibitor development was not considered. While we appreciate that ICER made this decision in order to simplify the model, this omission does not reflect real-world conditions. Additional information on possible implications of this decision should be added to the descriptive analysis for the reader; such an analysis should comment specifically on avoiding the development of FVIII inhibitors in the short-term in previously untreated patients (PUPs), balanced with the lack of data available on what might happen from a patient's perspective regarding delayed inhibitors due to lack of exposure to factor replacement.	As discussed above, we feel we addressed this in the section on uncertainties and controversies.
11.	Biological Activity of Post Gene Therapy FVIII The draft report notes that "The manufacturer of Valoctocogene Roxaparvovec has suggested that the low bleeding rate seen even as factor VIII levels decline imply that the factor VIII produced by gene therapy may be more biologically active than the factor VIII in patients with mild or moderate haemophilia." Such speculative statements are often reported, but do not actually have any basis or add benefit to the report. We would recommend removing this section from the report.	We try to address areas of uncertainty in our reports, and this issue was raised by prominent experts in hemophilia. We feel that we provided sufficient caveats around this.
12.	Durability This is a topic of significant debate currently, especially for FVIII gene therapies. In these models, it also has a significant impact on the interpretation of whether the treatment is cost effective or not. We suggest firstly that data is reported using the chromogenic assay. Secondly, if possible, we recommend that an analysis of duration be carried out and used to inform considerations towards a 'fairer price' for gene therapy. The main rationales for this is to prevent another redrafting of this report in the future, to help guide potential cost savings, and to further inform the types of payment models that would be recommended.	We chose to use the one-step assay results because this assay was used in prior reports of factor VIII levels and risk of bleeding. Given the changing costs of factor VIII prophylaxis over time, a threshold analysis looking at duration of benefit of gene therapy required for cost-effectiveness is, unfortunately, likely to require revision by the time a gene therapy becomes available.

International Hemophilia Access Strategy Council		
1.	Given variability in dosing and associated outcomes across factor VIII prophylaxis products in clinical trials and real-world studies, assumptions from both data sources need to be carefully considered due to its impact on cost-effectiveness outcomes. The estimation of factor utilization used for the cost-effectiveness analysis is critical. Because the model is highly sensitive to assumptions on utilization, we suggest ICER provides further detail about utilization based on varying ranges to better understand thresholds of cost offsets.	We agree that dosing and associated efficacy are a major issue. We have now made the ATHN doses the base case along with what we see as an evidence based bound on bleed rates associated with the representative doses from the Malec paper. We also highlight a scenario analysis around the trial based doses and the NMA in the clinical section.
2.	Differences between dosing of factor VIII prophylaxis and associated patient-relevant outcomes in clinical trials and that in the real world may be pivotal to cost-effectiveness findings regarding any treatment vs factor VIII prophylaxis. Regarding clinical trial utilization, there is a detachment that is not accounted for in the method of integrating utilization into the base case cost effectiveness model. Although there is a rationale for utilizing SPINART as the clinical trial to represent factor VIII products from an efficacy perspective, there are multiple classes of factor VIII products and using just one trial cannot accurately represent all approved factor VIII products, nor their treatment regimens, nor the outcome levels achieved. Additionally, we support the use of real-world evidence in the ICER draft evidence report; however, these real-world utilization rates should be supported by sound evidence and should also be accounted for by their effects on outcomes. In the report, the evidence is limited to only one real-world study without detail of the method in which it was captured, and thus enables the selected sample to influence the conclusion of the report completely.	See the answer above. We base the dose of the two representative products on average prophylactic doses of these products in US based hemophilia treatment centers. We also include a rigorous set of sensitivity and scenario analyses to help characterize the potential variance of the results related to dosing and other potential changes in model inputs.
3.	As such, we recommend that ICER identify and use data from multiple credible sources of data including clinical trials and studies using real-world evidence to represent factor product utilization and efficacy/effectiveness. This will ensure that treatments across multiple classes of factor VIII products are represented and that one study or treatment is not unfairly weighted in the analysis.	We acknowledge the variance in dosing and treatment protocols. We use highly prevalent treatments from the two primary classes of factor VIII treatments and data driven representative doses to characterize our base case costs. We also include a variety of sensitivity and scenario analyses and we discuss the importance of dosing to the conclusions.
4.	Given the lack of patient-relevant benefits currently incorporated into utility values, ICER should integrate these patient-relevant factors into the baseline and post-interventional utilities of their cost-effectiveness models. The goal of increased utilization is to target higher trough levels which results in substantial patient benefits, including reduced bleeds and improved outcomes over time (Lambert 2018; Peyvandi 2019; Jimenez-Yuste 2014). However, the cost-effectiveness model in the assessment does not comprehensively capture the additional patient benefits. The model limits any potential improvements that may arise strictly to productivity and reduced bleeds but does not account for the importance of improving the clinical phenotype and does not acknowledge the decreased disease burden associated with an improved phenotype or sustained protection.	Our model uses literature based utility scores and cost implications associated with bleeds along with literature based utility related implications of joint deterioration and surgery. These are consistent with other past and current models in the literature surrounding hemophilia A. We describe model validation methods that were also incorporated and highlight where our model differs from others. To the extent that clinical phenotype is captured in bleed rates, it would be included in our model. We discuss additional benefits in the section on Potential Other Benefits.

5.	A patient-centered outcomes framework for assessing value in hemophilia has been created and includes 3 tiers: 1) health status achieved or retained, 2) process of recovery, and 3) sustainability of health. Within each tier, there are several metrics that should be considered when assessing the value of hemophilia treatments (O'Mahony 2018). Currently, there are factors important to value assessment from all three tiers that the cost-effectiveness model does not incorporate, such as function/activity, pain, and health-related quality of life (physical, mental, emotional, and social functioning). The current model structure with its utility values strictly based on bleeds and Pettersson scores misses these key components that would enable a multidimensional integration of patient benefit.	We follow ICER protocols and adhere generally to expert recommendations for high quality cost effectiveness modeling. Also see the answer above. While certainly important to consider in certain contexts, it is not known how the specific items in the tiers of the value framework relate to the treatments in question and how they would map to costs and quality adjusted life years other than what our model projects.
6.	We respectfully suggest that the results of this report and its implications on the value of a gene therapy treatment option in Hemophilia A not be taken as definitive due to gaps in evidence. We understand ICER's limitations regarding available evidence when conducting this assessment which was further substantiated by the FDA's decision to ask for longer term efficacy data for valoctogene roxaparvovec.	We agree.
7.	For future reports on SSTs, it may also be beneficial (in situations where the price and durability of a product is unknown) to develop scenarios of testing varying duration of effect or alternative prices that may inform discussions around alternative payment models between manufacturers and payers.	We did look at alternative payment models. Given the changing costs of factor VIII prophylaxis over time, a threshold analysis looking at duration of benefit of gene therapy required for cost effectiveness is, unfortunately, likely to require revision by the time a gene therapy becomes available.
Paul Langley		
1.	Do you have a proof that the EQ-5D-3L/5L have ratio measurement properties. Do you have a proof that the TTO has interval measurement properties?	For a discussion of the scale properties of the QALY model (including TTO), please see: Roudijk et al., *Medical Decision Making* 2018; 38(6):627–634.