FOOD AND DRUG ADMINISTRATION

ONCOLOGY CENTER OF EXCELLENCE

Product Development in Hemophilia

Public Workshop

Thursday, December 6, 2018

8:34 a.m. to 4:30 p.m.

FDA White Oak Campus

Great Room

10903 New Hampshire Avenue

Silver Spring, Maryland

C O N T E N T S

C O N T E N T S (continued)

C O N T E N T S (continued)

C O N T E N T S (continued)

C O N T E N T S (continued)

P R O C E E D I N G S

(8:34 a.m.)

Welcoming and Opening Remarks

DR. MARKS: Good morning, everyone. I'm Peter Marks, director of the Center for Biologics Evaluation at FDA, and on behalf of FDA, I just want to welcome everyone in the room and online to the FDA hemophilia workshop. I want to thank you all for attending.

Before I get started, I want to thank a number of colleagues at the Center of Drug Evaluation and Research, in the Oncology Center for Excellence, in FDA's Office of Patient Affairs, as well as in our own Office of Tissues and Advanced Therapies at the Center for Biologics for putting together what I think will be a very stimulating program, which I think should lead to some good discussion.

As a hematologist/oncologist by training and as someone who has cared for numerous people with hemophilia, as working in a hemophilia treatment center, it's really a very exciting time

to be getting together because there are quite a number of products now in development, not just novel protein therapeutics, but also gene therapy, which has been on the horizon for more than two decades for hemophilia, may finally be becoming a reality.

Really, from the evolution of hemophilia for us in the Center for Biologics is pretty impressive because this is something that went from blood transfusions in the 1920s; to the use of cryoprecipitate in the mid-1960s; to the use of crudely purified factor concentrates in the 1970s; to the use of recombinant concentrates in the 1990s after the hemophilia community was particularly badly hit by the HIV epidemic.

Now, we're on the horizon of novel protein therapeutics that are either bispecific monoclonal antibodies, conjugated proteins, and as I've mentioned, gene therapy.

So really, I think, today, it will be a great discussion around aspects of product development ranging from appropriate surrogate

endpoints, to patient-reported outcomes, to appropriate clinical trial designs, and that discussion is really quite timely.

With that, to try to keep us somewhat on time, I will shorten my opening remarks a little bit and just thank you once again for coming today either, again, here in the room or online, and I will introduce Dr. Al Deisseroth, who will talk about the FDA 101. Thanks very much.

Presentation - Al Deisseroth

DR. DEISSEROTH: Thank you, Peter.

So as Peter indicated, my name is Al Deisseroth, and I'm going to provide some background information for the standards used by FDA for the approval of marketing applications and the ways in which FDA can expedite review and approval of applications for new therapies. I have no conflicts to report and the views that I will discuss are my own.

In 2018, the FDA carried out 32 approvals in hematology; 12 new molecular entities, 5 biosimilars, and approval of 2 products for

hemophilia, recombinant pegylated hemophilic factor for hemophilia A and one of the bispecifics, emicizumab, for prophylaxis in patients originally with inhibitors and now without.

The top half of this diagram includes the 14 approvals of non-malignant indications; 2 therapeutic antibodies; 3 agonists of the thrombopoietin receptor; 2 ESAs; 3 filgrastim products; 1 anticoagulant; 1 TKI; and the 2 hemophilia-related products. And the bottom summarizes malignant hematology.

So as you can see, drug development and product development in the area of hematology has been quite active, recently. The basis for all of these approvals is a demonstration of efficacy with acceptable safety, and adequate well-controlled trials, and the ability to generate chronic labeling, which defines a patient population and enables safe and effective use of the drug product.

For a full or regular approval, evidence of the clinical benefit is required as measured by increased survival or improvement in the quality of

life through ameliorations of symptoms.

FDA, however, has the authority to exercise flexibility in the application of these standards. One way that the FDA exhibits flexibility in its regulatory activities are the programs that provide for expedited review and approval of products.

There are several methods that the FDA is entitled to use. Fast track applies to products that have preclinical or clinical data that suggest that there's a potential to fulfill an unmet medical need.

Breakthrough therapy relies on clinical data showing substantial improvement over available therapy as measured by clinically relevant endpoints.

Priority review is applied if the product would provide significant improvement in safety or effectiveness.

The fourth method of expedited review is accelerated approval, which uses a surrogate endpoint other than one that can equate immediately to benefit, which must be reasonably likely to

predict clinical benefit.

This slide shows the difference between regulated and accelerated approval. As I mentioned, endpoints for regular approval equate to clinical benefit, whereas for accelerated approval, there is a surrogate endpoint reasonably likely to predict clinical benefit.

This slide is an example of an approval in the hemophilia area, emicizumab. And as shown on this slide, emicizumab is a bispecific antibody, which stimulates the functional effects of factor VIII by bringing together factor IXA and 10. The endpoints that were used for this product analyzed bleeding rate and a patient-reported outcome instrument.

This slide summarizes the landscape for therapies; as Peter said, factor replacement by passing agents and applying specific antibodies in gene therapy. And the available endpoints can apply to each of these types of therapy.

One of the problems that the field encounters is when different factors or different

assays get discrepant assays, and I think this will be the topic of discussion today, to which we're looking forward.

The last method of expediting review that applies to cell-based therapies is the RMAT program, which is sort of a mixture of fast-track and breakthrough therapy, but doesn't have the requirements for demonstrating a substantial advantage over available therapy. I think CBER has received 31 requests and has granted 11 in this area, so it's quite a useful method for cell-based therapies.

Well, these brief remarks, I hope have convinced you that FDA is capable of exercising regulatory flexibility to expedite the approval process for promising therapies for which there is an unmet need.

I'm looking forward to the presentations today, which may lead to identification of ways of optimally managing expedited product development for hemophilia. I'll now give the floor to Dr. Jay Lozier.

Presentation - Jay Lozier

DR. LOZIER: Thank you, Al.

My task is to describe CBER's concerns for hemophilia product development, and I am a medical officer in CBER in the Office of Tissues and Advanced Therapeutics. I have no relevant disclosures, as you might imagine.

I will talk about CBER's mission and how we regulate hemophilia-related products in CBER, how we approach development of new products, and our regulatory experience, and then point to some special concerns, particularly for gene therapy, and then talk about our goals for this workshop and how they'll be addressed by our very capable speakers.

So our mission is to ensure the safety, potency, purity, and effectiveness of biologics and particular blood products, and gene therapies. The key words are "safety" and "efficacy," and that applies to all CBER products. Biologic products are defined as viruses, therapeutic serums, toxins, antitoxins, or analogous products applicable to the

prevention or treatment or cure of disease or injuries of man in the CFR, which is where we find all of our definitions. And the basis for our regulatory review is based on Title 21, Section 601 of the CFR.

Biologic products are reviewed mainly at CBER, but there are some that are reviewed in CDER. We regulate plasma-derived, recombinant, and gene therapy products for the treatment of hemophilia in CBER.

So product development; this is the standard product development at FDA, which applies to the hemophilia products. Often, there is an early interaction between people with particular notions about how to develop a product. There's an informal set of meetings, INTERACT meetings. And then when you have preclinical data, animal data, and in vitro data, and you think you're ready to go into humans, we have a pre-IND meeting typically.

Then when you think you're ready to go into humans, you submit an IND, which we have 30 days to review. And if we don't raise objections or we

iron out any differences we have about things, then typically, after 30 days, sponsors start on phase 1 clinical trials to establish the safety of the product. These may be first in human or those kinds of studies, and there may be more than 1 phase trial.

Assuming safety is established and doses are found, then you proceed to phase 2 studies of efficacy. Then, once you feel like you have a product that's ready to go and be tested, you do a phase 3 licensure trial, where you try to find some clinical endpoint and continue to demonstrate the safety to merit licensure.

The BLA stands for biologics license application, and that is when you come to us with your clinical data and say we want to market this. We then review this, and if you undergo an approval, then it's not over. There's postmarketing surveillance and postmarketing commitments to study safety typically in -- and this is particularly important for accelerated or expedited approvals. But there is postmarketing

surveillance of all products to some degree.

So when we regulate factor concentrates, the population we serve are the patients with severe hemophilia and bleeding risk. The natural history of these patients will differ amongst patients with severe hemophilia. For instance, those who have pre-existing joint damage and severe hemophilia may have a more severe bleeding phenotype than those with mild or moderate hemophilia.

We have used the average of the annualized bleeding rate, or the ABR, as the usual primary endpoint for efficacy for factor concentrates, currently. It's a subjective finding. It's a patient-reported outcome. And if we're going to use the ABR rate to describe a product as offering a benefit, you will have to enroll patients who have some bleeding episodes on replacement therapy to show a benefit for the new product or therapy.

Now, with widespread prophylaxes, essentially the de facto standard of care, often we have patients entering trials with ABRs on standard

therapy of zero or near zero. This sometimes is difficult for clinical trial design.

Factor levels are measured when we test factor concentrates in the clinic and we look at peaks and troughs. Seldom do we have a steady state that's achieved. Most of the measurements of factor levels for FDA clinical trials will be limited to pharmacokinetics and pharmacodynamic studies and determining the dose for routine prophylaxis management or perioperative management or control of bleeding.

There can be interpatient variability with regards to the pharmacokinetics and pharmacodynamics and there, as you will hear, are issues with the assays themselves because there can be discrepancies between chromogenic assays and the one-stage factor assays, which look at the activity via clotting methodology. The safety risk for factor concentrates these days is really centered on worries about inhibitor development.

There are some special concerns for gene therapy with regard to efficacy. We expect that

gene therapy will likely result in steady-state factor levels. If we look at factor levels as a surrogate endpoint for reduction in bleeding, we have a limited understanding of the relationship of factor levels and the reduction of bleeding risk.

More is better, but we can't necessarily say that a particular factor level, factor VIII level particularly, associated with a mild bleeding phenotype will necessarily translate to the same mild bleeding phenotype or no bleeding risk for gene therapy.

We do have issues with discrepancies between the chromogenic and one-stage factor assays, and they're really markedly different than our experience with recombinant products. In gene therapy, we are particularly aware that neutralizing antibodies in the vector may limit the initial treatment or re-treatment with a vector. So if you have an AAV vector, it's a very potent immunogenic set of capsid proteins that will elicit a very strong antivector response, so we typically think of AAV gene therapy as a one-time event.

We have an issue with whether we're going to see long-term durability of steady-state factor levels, and the jury is out on that because many of the clinical trials are still ongoing, and we still wait long-term data on the stability of the factor levels.

With regard to safety, we have concerns for liver-related toxicities. These now are, I think, pretty well understood, and anticipated, and managed in AAV gene therapy clinical trials, and those are usually pretty well managed with corticosteroids.

We have theoretical concerns about insertional mutagenesis, and given some preclinical studies in animals, we are certain that we will need long-term surveillance with any of the gene therapy vectors, whether it's AAV or lentiviral, retroviral, or whatever may be proposed.

We used to worry that we couldn't get enough factor VIII or factor IX to make a difference, and I remember writing any number of papers with everybody else in the room, saying, if

we could just get to 1 percent, we would make a difference, which we would. But now we have gene therapy trials where we're getting supratherapeutic levels, and we have to be at least concerned to some degree about the risk for thrombosis when you see factor levels getting up in the high 100s and 200 percent level, which 20 years ago I never would have predicted could have happened.

For pediatric patients, we need to know whether liver growth and development will affect the durability of the factor levels, and where we think currently gene therapy will be a one-time treatment, how do we design a treatment or can we design a treatment for children that can be a one-time treatment. That's an open question, and we have to worry about the risks for insertional mutagenesis and are these risks greater in children than for adults.

So our goals for the workshop are to address the efficacy issues. In session 2, we'll be talking about the physiology of hemostasis from an in vivo gene expression standpoint, the impact

of joint damage on the annual bleeding rate. In session 4, we'll be talking about factor assay method discrepancies, and in session 5, we'll be talking about the durability of factor level expression and adolescent liver growth.

In session 5, we'll be addressing safety issues for clinical trial design, particularly the risks for insertional mutagenesis and considerations for enrolling pediatric patients.

With that, I will end on time, and I will turn the microphone over to Laurel Menapace, who will talk about the CBER perspective on drug development.

Presentation - Laurel Menapace

DR. MENAPACE: Good morning. I'm Laurel Menapace, a hematologist and clinical reviewer in the Division of Hematology Products at the FDA. Before I begin my slide deck, which is relatively short, I just really wanted to thank all our patient advocates, physicians, scientists, and investigators who have joined us today. Really, without your participation, this workshop would not

be here. It took a lot of months of preparation in advance, and we greatly appreciate your input and your feedback.

As my colleague, Dr. Lozier, talked about, he briefly outlined the CBER mission and points of interest from a biologics perspective at the FDA, in terms of new product development and hemophilia. My presentation is really going to complement that and simply talk about drug development and were notably some of the recent approval we had in hemophilia. So without further ado, I'll get into that.

I'll just have a brief introduction of CDER's mission, again which complements the CBER mission and our role in drug development, and then bring up a few clinical and safety concerns we have regarding novel drug development in hemophilia patients.

Lastly and most importantly for me, I'd like to highlight a new field in hemophilia, patient-reported outcomes. Patient-reported outcomes have been heavily emphasized in oncology

and hematology trials, but we're beginning to see increasing emphasis on patient-reported outcomes in benign hematologic conditions, including hemophilia.

We're seeking feedback about patient-reported outcomes to guide us as we think about the future of patient-reported outcomes and incorporating them into clinical trial design specifically for patients with hemophilia A.

So when we think about the CDER strategic mission, there are two key points here. There are actually a total of three. I've only highlighted two here. Really, we promote public health by helping to ensure the availability of safe and effective drugs, and we protect public health by promoting the safe use of marketed drugs in the postmarketing setting.

What I've outlined here is really that we identify and develop new scientific methods, models, and tools to improve the quality, safety, predictability, and efficiency of new drug development.

The title of my slide demonstrates that the field of hemophilia A in drug development is in flux. It's changing and it's very dynamic. It's no longer static. We've relied on typical factor replacement products for many, many years, and now we're beginning to see novel drug development. And as such, the paradigm of treatment is shifting, and we need to best understand this and interact with our academic colleagues and investigators, as well as patients, again, to develop new ways of understanding these drugs, how these drugs should be implemented in clinical trials, and how we should approve these drugs.

Again, in the postmarketing setting, after we've once approved a drug, we are looking for early detection of new safety signals. We need to understand emerging safety signals with these drugs, and effectively manage these signals, and communicate with the practicing community in terms of mitigating these risks and how we should inform our patients moving forward.

You've probably already seen a similar

slide in Dr. Deisseroth's and Dr. Lozier's presentations, but again, I just want to highlight the fact that FDA and particularly my division, the Division of Hematology Products, our reviewers, which we have multi-disciplinary teams comprised of physicians, chemists, pharmacologists, toxicologists, and statisticians, as well as a number of other experts in the field, are heavily involved in the early process of drug development, even in the pre-IND phase, and then again heavily involved at each stage of clinical development, as you can see outlined here, heading from IND submission all the way to IND review, and then phase 1 through phase 3 development, and then ultimately submission of clinical trials for review of the agency for regulatory approval.

Again, our job doesn't end once we approve a product. We are constantly going through postmarketing surveillance, and looking for new safety signals with these drugs, and effectively communicating with safety providers and the public.

Some may ask, okay, Dr. Lozier gave a great

outline of some of the products the Center of Biologics is reviewing and responsible for, so what does CDER do in terms of hemophilia?

The two centers complement each other, and I would simply say what Dr. Lozier didn't present on his slides is what CDER is responsible for. But in terms of our hemophilia pipeline drugs, I just wanted to draw your attention to two.

The first is fitusiran, which is an investigational antisense therapeutic target which targets antithrombin. This has been in development for the treatment of hemophilia A and B with and without inhibitors and currently is in phase 3 of development after a clinical hold was lifted regarding some safety issues.

The other class of drugs that I want to draw your attention to are the anti-tissue factor pathway inhibitor antibodies, which we're beginning to see at the agency. And this is a class of drugs, and there are a number of drugs in various stages of clinical development, most in early stages of clinical development, including phase 1

and phase 2.

In regard to our recent approval in hemophilia A, most of you are familiar with emicizumab-kxwh or also known as Hemlibra. Emicizumab is a humanized monoclonal bispecific antibody that binds both activated factor IX and 10, thereby bridging the two and restoring effective hemostasis in patients afflicted with hemophilia A.

It is administered via a subcutaneous route, which is novel, and has a half-life of approximately 4 to 5 weeks. So the initial approval of emicizumab was in November of 2017, where we approved emicizumab for a routine prophylaxis to prevent or reduce the frequency of bleeding episodes in patients with severe hemophilia A with the presence of factor VIII inhibitors.

In a short period of time, the sponsor then submitted data from their pivotal HAVEN 3 and HAVEN 4 trials, and this led to an additional approval in October of 2018 where emicizumab was

approved for prophylaxis in hemophilia A patients without inhibitors, and additional dosing regimens were incorporated into the prescribing information.

In terms of safety concerns regarding emicizumab and questions for the agency as we move forward with this newly marketing drug product, some of these we're well familiar with and have been discussed extensively at other conferences and recently ASH. But most notably, with initial approval, there were concerns regarding thrombotic events, both arterial and venous, as well as the incidence of thrombotic microangiopathy, which occurred in patients who not only were receiving emicizumab prophylaxis, but were receiving high levels of bypassing products, high doses of bypassing agents for the treatment of breakthrough bleeding. This resulted in a black-box warning and, again, we're continuing surveillance in these patience.

More importantly, we didn't see any events in the recent HAVEN 3 and 4 clinical trials, but again, these trials enrolled patients without

inhibitors, so they were not receiving bypassing agents. They were receiving typical replacement products for breakthrough bleeding.

Furthermore, another more recent safety concern, which had been identified as a potential safety concern by many of us early on, was the development of antidrug antibodies. As we know, these are common with this therapeutic class of drugs, of antibodies, and they can result in clinical loss of efficacy.

There recently had been a report of a pediatric patient in the HAVEN 2 trial who developed anti-drug antibodies with clinical loss of efficacy. He was discontinued from the study and returned to his prior prophylactic regimen, and there were no other safety events. But moving forward, we have to think about this potential with emicizumab and monitoring in the clinic setting, and how we're going to handle these events in the future.

So just to highlight some overarching themes and topics that we'd like to see addressed

today and that many of our experts are going to go into great detail about, again, these are some questions we have for the future of emicizumab therapy as a novel product in hemophilia A.

These include therapeutic monitoring of patients receiving emicizumab prophylaxis, treatment of breakthrough or acute bleeding with factor VIII replacement products in patients without inhibitors, as well as bypassing agents in patients with inhibitors.

It's very important to note on the trials, particularly after the events of thrombotic events and TMA occurred, that the sponsor had redesigned their trials so that patients were receiving the minimally effective doses of replacement products or bypassing agents. And again, that's provided in guidance in the prescribing label.

This may not necessarily reflect a real-world setting, where you have an acute or serious bleed. This is something to think about.

We also have questions about emicizumab prophylaxis in the setting of surgery or acute

trauma, and as I previously alluded to, how we're going to monitor for develop of anti-drug antibodies, and the fact that, ultimately, even patients without inhibitors have the potential for delayed inhibitor development because they're still relying on traditional factor VIII replacement products in the setting of breakthrough bleeds.

So in the short term, we may be preventing this dreaded complication of hemophilia A treatment, but ultimately, they may still develop inhibitors.

Now, switching quickly to patient-reported outcomes, I just wanted to highlight, for those of you who are not familiar, this is considered a clinical outcome assessment. A patient-reported outcome is a measurement that basically comes directly from the patient about the status of a patient's health condition without further amendments or interpretation of the patient's response by a clinician or anyone else. For example, this may be a rating of pain on our traditional pain scale.

Why is the FDA interested in patient-reported outcomes and why are they so important in hemophilia? Patient-reported outcome instruments were utilized as secondary endpoints in all HAVEN clinical trials to support our regulatory approval for emicizumab prophylaxis in patients with hemophilia. And we're beginning to see an increasing interest from sponsors of drug development programs in hemophilia interested in patient-reported outcome measures and implementing them in clinical trial design.

For the purpose to keep my presentation brief here, I'm not going to go through this whole slide, but basically, I just want to highlight that patient-reported outcome assessments should be held to the same standard as other outcome measures in our trial, and that they should include a clear statement of objectives, well-defined and reliable assessments, and can distinguish the effect of the drug from other influences.

In terms of regulatory goals for including patient-reported outcome data, there are several

paths that sponsors and pharmaceutical companies can pursue. Sometimes, they're seeking just supportive data for overall benefit-risk assessment. Sometimes, they would just like to provide descriptive patient experience in the product label. Furthermore and lastly, some would like to make a claim of treatment benefit in the product label.

Just to highlight our CDER needs for the workshop in regard to PROs, which we'll be discussing in session 3, we'd like to introduce some commonly implemented PRO instruments utilized in the clinical trial setting, and we have invited several patient advocates, who will discuss the meaningfulness and utility of such instruments to adequately capture the burden of disease.

This is really important. Something that we're trying to highlight here at the agency is the voice of our patients, and the impact of such measures, and whether they actually have clinical relevance for these patients who are afflicted with hemophilia A.

Finally, we would like to gain feedback regarding the utilization of patient-reported outcomes and hemophilia clinical trials to support regulatory approvals from our colleagues. And at this point, I'll conclude my presentation. Thank you very much.

(Applause.)

Session 1

Moderator - Lori Ehrlich

DR. EHRLICH: Good morning. I'm Lori Ehrlich. I'm one of the medical reviewers in the Division of Hematology Products in CDER. It's my pleasure to introduce Dr. Ragni. She joins us from the University of Pittsburgh, where she's a professor of medicine and clinical translational science and the medical director of the Hemophilia Center of Western Pennsylvania in Pittsburgh.

Her career's been focused on clinical and translational research and novel therapy development and hemophilia. She's just going to provide an introduction for the rest of the day with an overview of the progress and challenges in

hemophilia.

Presentation - Margaret Ragni

DR. RAGNI: Good morning. Let's go through my disclosures. You might say we're in a golden age of treatment for hemophilia, considering how far we've come from whole blood transfusion, plasma prior precipitate, clotting factors, and recombinant factors, and now with gene therapy and some of these novel agents.

But with every advance, we've had complications, and the new novel therapies are certainly not alone here. Perhaps the biggest complication of hemophilia today is inhibitor formation, with about 30 percent incidence, both in those on prophylaxis, the solid line, or those who are on demand, the dotted line. There's a T-cell dependent B-cell response to exogenous factor VIII, and because it neutralizes your factor VIII, the treatment is bypass therapy, but it poorly controls bleeding with twice the hospitalization, 10 times the cost, and 3 and half times the mortality of standard therapy in a

non-inhibitor patient.

While we can look at risk factors and understand risks from race, genetics, family history, and early factor exposure, we really cannot predict who's going to develop inhibitors. And the goal clearly is better hemostatic therapy to prevent and eradicate inhibitors, which was the topic of a recent NHBLI workshop, State of the Science for Inhibitor Eradication.

But in addition to inhibitors, the burden of treatment is high with 2 to 3 times weekly treatment. Serious complications exist. Venous access is difficult. Compliance as they become adults is low, and breakthrough bleeds really limit activity, and protection from joint bleeds and joint damage is very limited. And finally, the global disease burden is great and factor is scarce, so we need novel therapies.

The three that I'm going to talk about are emicizumab, fitusiran, and gene therapy, as you heard recently, and these represent potential paradigm shift with fewer infusions, less invasive

route, longer protection from bleeds, improved hemostasis, improved quality of life, and potential for reduced immunogenicity, and even for potential phenotypic hemostatic cure.

But complications continue to persist with plasma-derived factor. We had hepatitis, HIV, and inhibitors. With recombinant factor, we've had inhibitors and a variable recovery. With bypass therapy, bleeding is poorly controlled and thrombosis may occur.

With extended half-life clotting factors, we had higher doses and frequencies, but it raised expectations. We ended up discussing the treatment quite a bit with our insurance colleagues to allow our patients to take what seemed to be working for them.

With gene therapy, clearly there's the capsid immune response, as well as other causes of hepatotoxicity, and with some of our novel therapies, hepatotoxicity and thrombotic microangiopathy.

In addition, we need to be thinking about

new measures of treatment response. Certainly, with both plasma recombinant factor, we were able to use factor VIII-IX assays as well as inhibitor assays, and with bypass, we couldn't specifically measure factors, but we use thrombin generation and thromboelastography, not available in many clinics.

With extended half-life clotting factor products, there have been variable peaks and troughs and the evolution of a population pharmacokinetic approach. With gene therapy, the question is what level are we trying to attain and discrepancies between chromogenic and standard 1-stage assays, and quality of life and its importance in assessing outcomes, as well as some of these patient and other core outcomes. With novel therapies, thrombin generation has been used as well as thrombogenic assays.

So let's talk a little bit about these novel approaches. I'm going to talk about the AAV gene therapy, emicizumab, as well as fitusiran. Let's start with emicizumab.

Emicizumab is a bispecific antibody that

binds factors IX and X. It's equally effective, whether the factor VIII is missing or an inhibitor is in place, and it basically mimics the factor VIII action to bind IX and X to effect hemostasis in a patient with hemophilia A or an inhibitor.

In phase 1 and 2 trials, there was clearly a dose-response curve, as you can see on the left, with increasing doses, increasing levels of emicizumab. This dose-dependent increase resulted in improvement in thrombin generation as you can see on the right. This was given once weekly intravenously, so had the potential for a simpler treatment.

As you can see here, this improvement in thrombin generation really was acquainted to an improvement in annualized bleed rate. And here you see in blue emicizumab prophylaxis, and you're comparing in pale blue no prophylaxis versus emi; in green, factor VIIa or FEIBA versus emi; and in yellow, factor VIII versus emi.

In every situation, there was improved

reduction in analyzed bleed rate, as well as in the large phase 3 trial comparing those with prophylaxis on the left or no emi prophylaxis on the right. There was a marked reduction in all bleeds in blue; in joint bleeds, partially treated bypass; as well as specific other bleeds. In each case, there was a significant reduction in the annualized bleed rate.

In these studies, other bypass was used for breakthrough bleeds. FEIBA was used in 27 percent. Recombinant factors VIIa at 33 percent or both in 12 percent. The most common adverse event was the injection site reaction in 15 percent. But as you can see, one of the most concerning findings was thrombosis, which occurred in 5 patients, all 5 of whom received FEIBA at a dose of 100 units per kilogram per day for over 1 day, and was associated with thrombotic microangiopathy.

So while emicizumab improves thrombin generation and reduces bleeds, there are some potentials for toxicity and also underscoring where our knowledge is lacking in risks of clotting and

risks of bleeding. And I'd mentioned, there were 5 deaths, all of which were thought not related to the drug, 3 of which you can see were in compassionate use; 1 in an expanded access program; and another patient who died of bleeding related to his hemophilia.

In terms of laboratory monitoring, emicizumab doesn't require activation by thrombin. It does artifactually shorten the APTT, so it would affect any assay based on the APTT, including single-factor VIII assays or the inhibitor assays. In fact, the APTT may be normal and the anti-VII may be zero in patients who are receiving this drug, while it may not reflect their true situation. But it is not affected by bovine chromogenic reagents, and for that reason, laboratory monitoring may use bovine chromogenic Bethesda assay or a chromogenic factor VIII.

Other assays are being evaluated, including thrombin generation, clot waveform analysis, and of course you heard a little earlier about the anti-drug antibodies in patients who seemed not to

be responding to drug. They may have developed anti-drug antibodies, 4 of the 18 in the HAVEN trials. And this was associated with reduced clinical efficacy and how best to manage that.

We don't all do ADA assays, but in a very interesting study by Nogami, he looked in vitro at anti-emicizumab monoclonal antibodies that compete with emicizumab and seemed to eliminate the effect of emicizumab in an APTT assay; so another approach that one might use in addition to measuring the assay.

So what are the issues about management in patients with emicizumab? Breakthrough bleeding should probably either minimize or avoid use of FEIBA altogether. Standard factor VIII dosing is quite reasonable, as is recombinant VIIa. And we have instituted in our clinic, and I'm sure in other clinics, that patients need to call the hemophilia treatment center if they are requiring continuing factor use for a bleed because we need to evaluate what the cause may be or symptoms of a blood clot. So we're making them aware, these are

patients with bleeding disorders, what a blood clot is.

Development of the anti-drug antibodies clearly in patients who have loss of clinical efficacy, increased breakthrough bleeds. We really need to think about that, and these patients need to be seen and discussed with us in clinic what needs to be done, clearly suggesting that patients who are non-compliant may not be candidates for this drug or we might need to figure out better ways to manage them.

Utilization of laboratory assays during emi treatment, APTT and anti-VIII are normal, as we mentioned, so we may want to use a chromogenic factor VIII or a bovine chromogenic anti-VIII to assess the status of our patients.

What do we do in surgery? It's clear that emicizumab alone may not be adequate for major surgeries. Certainly, we've used it alone in minor procedures. Patients with hemophilia are more likely to bleed than clot. We need to think about scheduling the surgery around the time of the

loading dose for hemostasis for emicizumab and then giving factor VIIa or factor VIII at the time of surgery, immediately before and after for several days, and monitor them very closely for bleeding.

Just a reminder; bleeding complications still outweigh thrombotic complications, so we need to manage these patients very carefully, but how you do that I think is not clear, and we're learning as we go along.

Immune tolerance; will emicizumab be efficacious if it's started before, or do we need to wait until after immune tolerance induction? There have been debates on both sides of this question, and certainly, long-term follow-up is necessary as are future trials of emi.

Cost-effectiveness, just to mention, the Institute for Cost and Economic Research has looked at the use of emicizumab in inhibitor patients and shown -- looking at the cost of bypass therapy, non-factor cost, long-term costs, including hospitalization, which is one of the most costly, and comparing it with bypass, with emicizumab, and

showed that it was clearly much more cost effective.

Is that true for patients who don't have inhibitors? That math has not been done, but time will hopefully tell, and we will be looking forward to hearing more about that.

So in summary, emicizumab may improve hemostasis, reduce treatment frequency. It may be less invasive by the subcutaneous route, and my patients love this drug, as I'm sure most of the physicians here will tell you. It may have comparable efficacy in inhibitor and non-inhibitor patients, but it may be thrombogenic if it's used concomitantly with FEIBA. We're very careful to tell every patient that issue and avoid prescribing it as much as possible.

It may be less immunogenic by avoiding factor exposure, but breakthrough bleeds, as was already pointed out, may still expose you to factor, so that question is out.

Future questions are, what about the treatment of acute bleeds? Surgery, how do we

manage it? Trauma, when do we use it in children, and do we use it in any other way in children? Certainly, it's been used in very young with excellent efficacy. And what about suppression of inhibitors? So there are a lot of unanswered questions.

We'll go on to the second drug, which is fitusiran or an antithrombin III knockdown. This works really by harnessing the RNA interference platform. It targets antithrombin production, MRNA in the liver. It interferes with its translation, binding to it in the hepatocyte, degrading the MRNA, and silencing gene expression, resulting in reduced or prevention of antithrombin synthesis, which clearly can be shown to be related in subcutaneous dosing weekly here at 0.75, 1.5, and 3 mgs per kg in a dose-dependent reduction in antithrombin level.

This is of course associated in the phase 1 study in hemophilia A with monthly dosing subcutaneously to show a dose-dependent lowering when it's given monthly, and that's associated with

once the drug is stopped, it's reversible, as you can see past day 80, 90, 100, and so on.

This reduction in antithrombin is associated with increasing peak thrombin, as you can see in this graph, and that's associated with reduced annualized bleed rate. As you can see on the far right, it is dose dependent.

This is also true in patients with hemophilia A with inhibitors, again increasing antithrombin lowering and is associated with greater peak thrombin generation and reduction in annualized bleed rate.

In terms of side effects and safety, I would point out that injection site pain is the most common, but in this particular study, there was cerebral sinus thrombosis, and this occurred in a single patient who used multiple doses, high doses of factor VIII, which were contraindicated in a study and for which the study was stopped.

I would also note that the fitusiran also is associated with hepatotoxicity primarily in patients who had hepatitis C and who are HCV RNA

positive. That is not treated with antiviral therapies. This drug is degraded by plasma and intracellular nucleases, targets the liver, but does not seem to be an inducer of P450.

So why did this happen? Fitusiran certainly may cause stress signals in HCV damage to hepatocytes. If this is the potential mechanism, we're not sure. It may lead to increase in LFTs, and the LFT elevation occurred only in those who were HCV viral-load positive who had not received treatment.

Going forward, patients must receive antiviral therapies, and that is part of this mitigation procedure; that they must first be treated with anti-HCV therapies before on studies. For breakthrough bleeds, we ask them to keep diaries, use low doses of factor VIII, IX, VIIa, APTT, and to call if they need continuing dosing.

For surgery, if there's a major surgery, we try to schedule it at the nadir; that is, 2 weeks after the dose; and use factor VIII or IX or VIIa as needed. If it's a minor procedure, we've been

able to just dose at the time of the nadir with no additional dosing with patients who refused to take any other agents, and we found that to be the case in small minor dental and port procedures.

It's important to educate our patients about the symptoms of thrombosis so that they are well aware of those things. I will finish with hemophilia gene therapy.

As you know, hemophilia is really a model disease for gene therapy because it's monogenic, and there's a wide range of factor levels affected. It is a one-time potential cure, and what it really offers is potential global treatment for many who were affected for which there are no treatments, and they are shunned in their society or die young.

You've seen this graph many times at ASH and here, but in general, we really don't know what the level that we would like to see here is. What we want to do is avoid bleeds entirely, and as time has gone forward, we know that, at least with the 12 or 15 percent level, we can do that.

Are higher levels better? Are we getting

into a range in which we're worried more about thrombogenesis? But what we really want to do is to convert a severe phenotype to a monophenotype and avoid bleeds altogether.

As you know, there are multiple approaches and strategies for gene therapy, but the AAV is the strategy used in hemophilia. The wild-type AAV is minimally pathogenic in humans. There are many different serotypes which offer tissue specificity. But there are some potential cons with a small packaging capacity, and pre-existing immunity is known in at least 30 or 40 percent.

In general, the strategy is that you load the cargo into this AAV vector with factor IX cDNA of up to 1.3 kilobases or factor VIIIb domain-deleted CDNA of 4.7 kilobases. And basically, the gene is inserted into a vector, infused intravenously into the patient, goes into the hepatocyte as expressed in the circulation. We draw those pictures for our patients. They seem to understand that quite well.

Once you've inserted this genetic material

into the wild-type genome, you use the capsid for tissue specificity. And here you can see AAV 8 is specific for the liver, as is AAV 5, and some for the musculoskeletal and heart.

I'm going to just talk very briefly about 4 gene clinical trials. There were two more and several more talked about at the ASH meeting, but in general, we have two here, University College of London, St. Jude. It looks at an AAV Factor IX and BioMarin and Spark with a factor VIII AAV vectors.

As you can see in this University of College of London study, one of the first in 3 dose ranges, you can see that the mean factor level was 5.1 percent, but it was fluctuating, but even at that level offered a 90 percent reduction in annualized bleed rate and over 90 percent reduction in factor use. So the major limitation was AAV capsid T-cell response, which seemed to be responsive in many cases to steroids.

Here, you can see that you can actually increase that efficacy, that is that factor level, even to a 33 percent steady state in this factor IX

gene therapy using the Padua gene, which is at an increased 8- to 12-fold higher factor IX, and this also results in greater than 90 percent reduction in bleeds and in factor use.

There were capsid-immune responses. They seemed to be steroid responsive, and the gene therapy was well tolerated, and these levels seemed to persist.

In the factor VIII BioMarin AAV 5 factor VIII trial, you can see that there was a wide range of factor VIII. These patients, many were started on steroids empirically to avoid immune response, and these levels ranged between 12 and 219 percent with marked reduction in both annualized bleed rate and factor use.

In the Spark study, which is still ongoing, the dose ranges were 11 to 14 percent, as you can see here. These patients had also marked reduction -- from the ASH meeting, a marked reduction in annualized bleed rate and in factor use, and still had some capsid-immune responses, suggesting that maybe empiric steroids may be an

improvement. They offer an improvement and longer-lasting higher levels.

What are the problems with gene therapy? Certainly, there is hepatotoxicity, transient liver function elevation noted in some patients to not just capsid-immune response, but also interactions with other hepatotoxic drugs. One hemophilia A patient receiving efavirenz, or Sustiva, as part of a highly active anti-retroviral heart therapy for their HIV, developed a grade 3 liver toxicity after AAV gene therapy.

Efavirenz has a black-box warning. It is one of the most highly hepatotoxic drugs, and it can induce oxidative stress and endoplasmic reticulum stress.

The mechanism of the liver function/dysfunction in AAV gene therapy is unknown. The temporal onset a few weeks after gene therapy and rapid reversal on stopping this drug certainly suggests that there may have been some synergistic hepatotoxicity, and we really need to learn more about this. But caution is urged to all

patients who want to do gene therapy to avoid potentially hepatotoxic drugs, and this is really a critical message to all our patients and treaters.

Finally, we talked a little bit about assay discrepancies. We know there are discrepancies between the 1-stage and the chromogenic assay. Which one should we use? Do we need to do both? There's also inverse discrepancy between factors such as the B domain-deleted Xyntha and gene therapy results.

The mitigation is to either use both assays or to standardize chromogenic assays. One of the questions, of course, is, are most hemophilia centers now going to adopt chromogenic assays?

What is the gene therapy's success? Is 10 to 15 percent a sufficient measure of success? Is greater than 15 percent better, and who will decide? Where does thrombosis fit into this picture? How high do we need to go, or do we need not to worry about it?

Is more better? Does getting to greater than 100 percent make you stronger, able to do more

work, et cetera, et cetera, activities? And are alternate measures of success a reasonable approach?

This is quality of life, some of these very important core outcomes, freedom from fear, happiness factor, as one of my patients told me, or even looking at outcomes from liver transplant patients as a yardstick to measure how patients do once their levels are corrected.

Certainly, for mitigation, more data are needed to assess factor levels after gene therapy, understand the discrepancy between factor and gene therapies, and determine what optimal therapies there are for gene therapy.

We should mention that a cost-effectiveness analysis has been done using a more cost-safe transition model looking at quality-of-life years gained. And clearly, as we compare gene therapy with factor VIII and model using literature and Medicare reimbursement measures, using a one-way and probabilistic sensitivity analysis over a 10-year time frame, and doing over hundreds of

thousands of simulations, gene therapy was clearly dominant in 92 percent of those simulations, and it's likely to be cost effective in severe hemophilia A as compared to factor VIII prophylaxis. This was one study. More need to be done.

In summary, what we're seeing with these novel therapies is improvement in hemostasis, both in hemophilia A and hemophilia with inhibitors. We're noting issues and questions that arise with 1-stage versus chromogenic and whether thrombin generation and TEG are the ways to monitor some of these.

These require less invasive administration subcutaneously. Patients love it, and it is an amazing change for these patients; reduced bleed frequency, looking at annualized bleed rate, improve clinical measures, whether you use quality of life or these core outcomes, as we mentioned.

There's an improvement in laboratory measures, but clearly discrepancies exist. It may reduce factor VIII or IX immunogenicity just by

avoiding factor VIII exposure, but breakthrough bleeds may still remain a problem. And it may induce potential toxicity such as the hepatotoxicity and thrombotic microangiopathy we talked about, and future considerations are really understanding these drugs long-term, real-world use, and use in ITI or with surgery and in children. Thank you.

(Applause.)

DR. EHRLICH: Thanks, Dr. Ragni.

MR. COSSENTINO: I just want to make one announcement real quick. After the break, we're going to be doing some interactive audience questions and polls using a website called slido.com, and we encourage everybody to log onto Slido during the break so you become familiar with it, and we have a test poll up right now.

Just go to slido.com on your phone or laptop, and enter event code 3355. It doesn't require any login or personal information, and you'll be able to ask questions and answer polls in real time. I'll display the directions on the

projector during the break in case you missed any of that, and there are handouts as well that have the directions. Thank you.

(Whereupon, at 9:36 a.m., a recess was taken.)

Session 2

Moderator - Najat Bouchkouj

DR. BOUCHKOUJ: In order to stay on time, we're going to go ahead and start. I'm Najat Bouchkouj. I am a pediatric hematologist/ oncologist and a clinical reviewer at the Office of Tissues and Advanced Therapies at CBER. I will be the moderator for session 2, which is titled "Clinical Endpoints in Hemophilia."

Before I introduce our speakers, I just want to give you an outline about this session. We're going to have two speakers, two presentations, 20 minutes each, followed by a panel discussion. We will leave the questions to the end, and we'll take questions from the audience who are present in person and online as well.

So if you can submit any questions you have

online, we'll try to go through them as time permits.

Just before I introduce our speakers, I just wanted to pose a couple of questions for the audience, and I hope you have joined Slido already so we can get your feedback about a couple of questions. There might not be a hard right or wrong answer, but we'll ask the questions right now, and then we'll ask them again after the presentations.

The first question is about a 30-year-old male with severe hemophilia B, who is currently on prophylaxis therapy with factor IX product. He has moderate activity, swimming and brisk walking 3 times per week. He is considering to be enrolled in a gene therapy trial.

What target factor level at steady state, which is a constant level, would be optimal to reduce his risk of bleeding; 1 percent, 5 percent, 15 percent, 35 percent, or 40 to 100 percent?

(Audience responds.)

DR. BOUCHKOUJ: We have about 32 answers,

and about 48 percent says 1 percent, and that goes down. And no one said -- 1 percent is the low. So no one said zero percent.

Okay. Let's go to the second question. The second question is about a 16-year-old boy with severe hemophilia A, who's currently on prophylaxis therapy with a factor VIII product. He has two target joints and he plays soccer. He wishes to consider gene therapy treatment.

What target factor level at steady state would be optimal for him to reduce the risk of bleeding? Again, 1 percent, 5 percent, 15 percent, 35 percent, or 40 to 100 percent.

(Audience responds.)

DR. BOUCHKOUJ: As I said, there is probably no correct answer, but we will go through the presentations, and then we will ask the questions again and see if you change your mind.

I have the pleasure of introducing our first speaker for this session, Dr. Bob Montgomery from the Medical College of Wisconsin. He is a senior investigator at the Blood Research Institute

at the Blood Center of Wisconsin and professor of pediatric hematology at the Medical College of Wisconsin, Children's Hospital of Wisconsin in Milwaukee. He's a physician scientist who has studied von Willebrand factor and its relationship with factor VIII.

Presentation - Robert Montgomery

DR. MONTGOMERY: Thank you very much, and thank you for the invitation to speak today. These are my disclosures.

When we're dealing with normal hemostasis and we have circulating levels of von Willebrand factor, factor VIII platelets, in this cartoon, when we have vascular injury, we expose the subendothelium, which becomes a nidus for von Willebrand factor binding. And that von Willebrand factor binding organizes itself and has the recruitment of platelets.

When those platelets are adhered, they activate, and it's that activated surface that factor VIII will in fact bind to. I show factor VIII coming from the fluid phase, but as we

know, factor VIII is carried in plasma on von Willebrand factor. Once that happens, we have the factor VIII that brings together the factor IXa and X with the ultimate formation of the clot. And after healing fibrinolysis, hemostasis is restored.

We'll be talking a bit about von Willebrand factor and its impact on factor VIII and also touching on some issues with factor IX.

There are two cells in the body that make von Willebrand factor, and one of those also makes factor VIII. There is no factor VIII in platelets. In the megakaryocyte, in the formation of alpha granules, von Willebrand factor is produced and is stored along with a host of other proteins.

If you don't have von Willebrand factor such in a type 3 patient, you actually still have alpha granules in platelets. So therefore, it's not that those platelets are dependent upon von Willebrand factor, as we'll see different in endothelial cells.

These megakaryocytes ultimately form platelets, and it's these platelets that have the

stored proteins, including von Willebrand factor, in the circulation. There is a secretory pool of von Willebrand here that's in platelets. There is no factor VIII unless it was put there genetically.

In the endothelial cell, we have the formation of Weibel-Palade bodies, which are the secretory granule of the endothelial cell. These Weibel-Palade bodies are actually formed because of von Willebrand factor. And if you don't have von Willebrand factor, you actually don't have Weibel-Palade bodies either. So it's a very different relationship.

This is also a secretory pool of von Willebrand factor, but when you secrete von Willebrand factor, as I'll show in a bit, you also secrete factor VIII, and that's different from platelets. In addition, we use DDAVP as a way of releasing these Weibel-Palade bodies to increase von Willebrand factor and factor VIII so that those storage pools are clearly different.

Both von Willebrand factor and factor VIII are acute-phase proteins and are increased with

surgery, with pregnancy, with physical stress, with mental stress, and with aging. All of these phenomena result in changes of the secretory pool, and it's something that we'll come back to, that at least current approaches to either replacement therapy or gene therapy don't necessarily replace this part of the process.

Von Willebrand factor can acutely be released by DDAVP, which also releases factor VIII, and this can be used if patients have mild or moderate deficiency.

A number of years ago, we actually asked the question, based upon secretion, where do these two proteins first meet. To make a long story short, here we see a patient with mild hemophilia who was treated with DDAVP. You can see that factor VIII goes up and the von Willebrand factor goes up, and both can be elevated into a therapeutic range.

However, if you take a severe hemophilia patient who's on prophylaxis and receiving factor VIII -- in this case, it actually was every

6 hours, for a variety of reasons -- but give that patient DDAVP, the von Willebrand factor goes up as expected, but there's no budge of factor VIII.

What's important there is that, therefore, you can't replace the stress pool or the secretory pool of factor VIII by infusion even though you can definitely stop bleeding.

If we look at the von Willebrand patient, again, DDAVP will release both proteins. These will be similar in a stress response. Interestingly, if you take a type 3 von Willebrand patient who makes no von Willebrand factor and has a baseline level of factor VIII usually around 5 percent of normal, and now you prophylax with von Willebrand factor concentrate that has no factor VIII in it and now give DDAVP, what's interesting is the von Willebrand patient's factor VIII has now been normalized because of changing the survival in the presence of von Willebrand factor. So factor VIII level is now normal and that's endogenous factor VIII.

Yet, if you give DDAVP, even though there's

endogenous VIII made in all the places, it doesn't create a secretory pool, and I think that's something that we'll touch on.

Two laboratories recently were able to show somewhat the same thing in a single issue of blood, and that is to study the amount, or the relative amount, of factor VIII that's in fact produced in endothelial cells. This was done by two different approaches -- I'll talk a little bit more about our own -- in which we floxed the factor VIII gene, which meant that if we took that animal and crossed it with an animal that was making, let's say, we'll say albumin Cre, the albumin Cre would cut out the factor VIII so that every cell that was making albumin would stop making factor VIII.

This actually can be shown. Here is the floxed factor VIII mice. Here are the ones in which we knocked out the factor VIII in albumin-synthesizing cells, and there was no effect.

In contrast, if we move to the cadherin and the TIE2, or the TEK Cre, you can see that factor VIII is essentially eliminated just like the

knock-out, suggesting at least it doesn't say what endothelial cell is making it, but it is saying that virtually all factor VIII is made in endothelial cells in mice.

More recent studies have suggested that the different beds of endothelial cells can have a dramatic difference, such that it may be that vascular endothelium may contain both VWF and factor VIII. Sinusoidal endothelial cells have factor VIII but may not have von Willebrand factor, and lymphatic endothelial cells are similar.

Recognize, though, that if we don't have von Willebrand factor, the only place in these models would be the peripheral vascular system that you had a secretory or stress pool of factor VIII.

Factor IX is less controversial, maybe, and factor IX is made in the liver by the hepatocyte. Here is a recent paper showing the various organ systems in the body, and the only one in which there was an identified factor IX mRNA was in the liver, not surprisingly.

If we went within the liver and now looked

at the cells within the liver itself, you'll see that LSECs, or sinusoidal endothelial cells, do not make any factor IX, and it's only made in the hepatocyte, not surprising.

Some other recent studies, however, by Darrel Stafford and his coworkers at Chapel Hill have demonstrated the importance of factor IX binding to subendothelial collagen-4. This bound factor IX provides an important extravascular pool of factor IX. Certainly, it's the intravascular that is physiologically important, but the extravascular may be able to support that in the long run.

Circulating levels of factor IX do not all predict the full hemostatic potential, and as shown using a K5A mutation in a mouse in which collagen-4 binding was eliminated, there was normal in vitro clotting, but reduced in vivo clotting, so that the fluid phase effect was easily measured even though the systematic effect of collagen-4 is not binding in a traditional clotting assay.

Now, great strides have been made that have

significantly affected gene therapy, and we'll hear a lot about that from other speakers today and some of the problems with the assay. But really, factor IX Padua has changed the field of producing a protein that has increased specific activity and is genetically modified, and along with some other modifications can produce many-fold higher levels of factor IX expression based on a mole-to-mole basis.

Here's a model of the assembly of the Xase complex. It's relatively straightforward that VIIIa binds to form the Xase complex. But this step may actually be more complex than that. I put in here von Willebrand factor because of the benefit of von Willebrand factor to increase the local concentration of factor VIII, something that doesn't necessarily happen unless von Willebrand factor is present.

There are a number of binding sites, the one that's been traditionally known for a long time as the GP1b/IX binding site on platelets that bind to the a1 domain of von Willebrand factor. It's

also been known for many years that 2b3a on platelets binds to the RGDS sequence that's present in von Willebrand factor. And therefore, that, along with the binding of factor VIII to VWF, could in fact facilitate the local delivery of factor VIII.

We now know, both in studies that have been published by Veronica Flood and another one presented at ASH this year on myosin, that these are also extra platelet binding proteins at the local vascular injury site that can augment, number one, the binding of von Willebrand factor; and number two, the delivery of factor VIII to formation of the Xase complex.

We also know that IXa here can bind to collagen-4 so that even von Willebrand factor is brought into close proximity with its factor VIII to factor IX that might be bound to collagen as well.

Great strides have been made through emicizumab. Emicizumab clearly can take over this function of bringing IXa to X, to the formation of

thrombin generation. If we think about it, that's a fluid phase protein and not necessarily something that's necessarily delivered with increased concentration.

I think there are still issues that need to be worked on where you need to think about comparing what is the local delivery of factor VIII to the systemic delivery of factor VIII and things that might augment clotting and regulate function.

What questions remain concerning factor VIII or factor IX? For factor VIII, is the site of synthesis important? Is a storage pool of factor VIII important? If it is, the site of synthesis becomes important since you won't have a secretory pool if you synthesize the factor VIII in cells other than the endothelial cell.

Does stress increase factor VIII or just release it from stores? Is there a problem with uncoupling factor VIII from von Willebrand factor as far as the physiology of local hemostasis?

Does von Willebrand factor actually serve as a protein that delivers factor VIII to the

evolving thrombus? This is something that one can speculate on but is only evolving better proof of that phenomenon.

For factor IX, does IX need to be made in a hepatocyte? There are certainly studies of it being well-made in muscle as well as other cells, and as we heard in ASH by Qizhen Shi, also, factor IX can be made in megakaryocytes in platelets. But if made in another cell, there needs to be both adequate furin and adequate gamma carboxylation.

The final issue is, is Padua safe? There are issues around its specific activity and its immunogenicity. Everything seems to be very favorable, but there are things that we just need to continue to be aware of.

How important is subendothelial collagen-4 binding as a store? What Darrel Stafford's group showed is that infusing high levels of factor IX actually can have a binding to the collagen-4 and actually caused sustained benefit over a longer period of time than necessarily measured in plasma.

With that, I'll thank those that worked with me, and thank you for listening. Thank you.

(Applause.)

DR. BOUCHKOUJ: Thank you, Bob.

Our next speaker is Dr. Marilyn Manco-Johnson from the University of Colorado. She's the director of the Hemophilia and the Hemostasis Center and the Children's Hospital of Colorado. She will be talking to us today about factor VIII and IX correlation with breakthrough bleeding and optimal joint endpoints of new therapies.

Presentation - Marilyn Manco-Johnson

DR. MANCO-JOHNSON: Thank you, and thank you very much for the opportunity to present today. Here are my disclosures.

I'm here talking about therapies for hemophilia A. I've tried to compare this to what we in hem-onc are more familiar with in the cancer world; that is, a complete response, a partial response, and no response. But a complete response would be normal biochemical and clinical outcomes,

while a partial response would include reduced bleeding, reduced factor consumption, and reduced morbidity with things like intracranial hemorrhage, hospitalizations, and other severe bleeding events. A partial response may be desirable, particularly in the short term, regarding the risk of excessive levels with a thrombotic potential.

When we look specifically at the important outcome of hemophilic arthropathy, what outcomes can we have? Well, certainly, I think we can never get away from the restoration of plasma factor activity; so factor VIII and IX both have ranges, ranging from 50 to 150 percent of a population mean. You can look at their activity or look at the protein content in the blood.

Certain surrogate markers for factor activity that are important to be applied to non-factor therapies would be looking at correction of the partial thromboplastin time, the normal thrombin generation, thromboelastography, and more recently, interesting markers of bone metabolism, which have shown to be altered in the absence of

factor VIII or factor IX and restored by the replacement.

Clinical effects of protein restoration, we have focused primarily up until now on no spontaneous bleeding. Bleeding has been used to consider clinical or determined to recognize bleeding; no bleeding beyond what a normal person would experience in trauma or surgery because, obviously, we all bleed given enough of a stress, and normal bone density, which is a more subtle and refined indication of thrombin generation. And we'll talk a little bit about no or reduced onset, or reduced progression of joint disease.

The benefits of direct and indirect outcome, if you look at factor VIII levels, we widely understand what that means. A normal level is normal. There's no reason to expect that if any therapy got someone within the normal range, that it wouldn't translate to normal clinical hemostasis.

Indirect evidence on bleeding and joint damage is more relevant to the patient. It's a

functional marker of efficacy. And the con, a very important con to this, is that all indicators of outcomes on joint damage and joint bleeding are going to work better in young patients with normal or minimal pre-existing arthropathy, because patients with damaged joints may experience variable outcomes relative to joint pain and bleeding depending on how they came into the trial.

I think this is very important because in all therapies for hemophilia, through the FDA, we do start with adult patients, who are better able to give consent and we feel are less vulnerable as research subjects; but on the other hand, they have developed and fixed cartilage and bone structures.

If you look at the effects of hemophilia, I want to argue as a pediatrician very strongly, that the effect of blood is much more severe on growing cartilage and growing bone. And we know that most of this damage is not reversible, so if we're going to come out with good adult outcomes, we need to start with the very young children and protect the cartilage and bone as it's growing.

So our functional outcomes are the prevention or stabilization of arthropathy, and we have physical joint scores such as the Hemophilia Joint Health Score, and we have imaging scores using both ultrasound and MRI. We have very important patient-reported outcomes, including quality of life, activity, participation, and pain, and these are going to be discussed later by Dr. Kempton.

This is a presentation that we made actually in 2013 by Tom out of Glorioso and colleagues, and it looks at joint outcomes with age. And very interesting, at a very young age -- so for all of these images, looking at joint bleeding, joint physical exams, and joint MRI scores, you can see that hemophilia is marked by a huge heterogeneity, with a huge variability in scores among patients. And of course that makes our registration trials with relatively small numbers of patients difficult. But bleeding gets to about a mean of 20 bleeds per year, and you've reached that very early in life, and it's about the

same throughout life.

Physical exam scores, while very variable, peak out in the young adulthood and don't really change much, whereas MRI changes are consistent throughout life as long as they've been measured. When we look between soft tissue and osteochondral changes, this is primarily the osteochondral change.

We looked at changes in the Hemophilia Joint Health Score, and this was presented at the World Federation this year, at individuals who started prophylaxis before 3, between 3 and 6, 6 to 10, et cetera. And we found that you could only blunt the curve of physical damage over time if you started below 3. And among all these other ages of starting, there was no difference.

On this scale, you see the Hemophilia Joint Health Score. We do these annually in Colorado, and looking at the positive score means you're worsening; negative score is improving. This is severe, moderate, and mild hemophilia. Right is on prophylaxis; blue is on demand.

You can see that there's tremendous overlap and the worsening of scores regardless of the severity of hemophilia, so mild hemophilia is way better than severe, but it's not great, and it's not the goal that we aspire to.

In the joint outcomes study that we reported at ASH in 2006, children given 25 units per kilo of recombinant factor VIII, starting before the age of 30 months, were found at the age of 6 to have significantly less osteochondral damage compared to children who use this on prophylaxis, such that the relative risk of joint damage was 6-fold if you did not use prophylaxis in the preschool years.

These children using Kogenate had a mean half-life of 12 hours and a mean 48-hour trough of 4 percent. So when you're looking at troughs, this is a baseline for what you get for 4 percent.

In the outcome, we found that there are many children who had relatively little bleeding, but evidence of bony change, and conversely, children who had lots of joint bleeds who had very

little joint damage, such that MRI showed a modest correlation with a number of hemarthroses so that we could only account for 13 percent of joint damage that could be explained by clinical or recognized bleeding. And this drove me to come up with a concept of subclinical, unrecognized, or micro bleeding.

Now, I want to emphasize this is in young, intensively treated children. It probably doesn't hold to 4 years ago, when individuals had relatively little treatment and big clinical bleeds.

Looking at that population, at the lifetime average of joint bleeds of individuals who started prophy at an average age of 1.3 years was 1.5 joint bleeds throughout childhood until age 18, whereas those who started at age 7 continued to experience more bleeding, with an average of 4.3. And if you considered only the time after they were on prophy, they still had 4 joint bleeds per year compared to 1.6 on the early prophy.

So if we look at clinical joint bleeding,

ABRs, in clinical trials, it depends if the patient was on prophylaxis or not and how early they started prophylaxis. So heterogeneous trials that enroll individuals from different backgrounds are going to be affected by bias.

We found that the odds ratio of joint damage between early prophy and delayed prophy was 14 at the age of 6, but held up as still an odds ratio of 6 at the age of 18. And I think 18's an important cutoff because most growth centers are fused and you have pretty full cartilage and bone development by that age.

Well, when we looked at our clinical, easily used surrogates for joint outcome, the clinical exam score, the joint ABR, the total ABR; unfortunately, none of them correlated with osteochondral changes on MRI. So the indicators we're using in our trials are not correlating with long-term bone and cartilage outcome.

The only predictor of the MRI osteochondral damage was the number of bleeds suffered before the age of 6, and this so strongly correlated with

whether you were on prophy before the age of 2 or on demand, that it really was a surrogate for early prophy.

Looking at the osteochondral changes over time, those who had early prophy unfortunately continued to accrue some osteochondral damage, but this was less than those whose prophy was delayed until age 7 and less than those who never had prophy. So at the age of 18 to 20, we had a total 6-joint MRI score of 7; if we started early prophy, 13; if the prophy was delayed, towards 7; and 20 if you never had prophy. So outcomes are dependent very much on the age it's starting.

The physical exam scores trend exactly the same way, that they do worsen over time, and at the time we did the joint outcomes study, 25 units per kilo every other day, this group had excellent adherence over 90 percent that you still accrue some damage, but it's less than then if you delay prophy until 7.

So going back to the lack of correlation, with recognized bleeding and with physical exam

scores, this again supports a subclinical unrecognized bleeding in our current population of young, intensively treated patients, and is very important and is probably as or more important than the clinical numbers of ABRs.

I'm not going to dwell on this because Dr. Montgomery just gave a very eloquent presentation of this. We know that factor VIII ranges fivefold in healthy people, and we know that both exercise and inflammation raise factor VIII, and we know that continuous factor VIII will not respond to physiologic stresses.

What is the optimal goal of factor VIII therapy? Should we be aiming to mimic physiologic levels or should we be attempting the lowest level that results in no clinical symptoms for the widest range of patients? And I've already given some arguments why the clinical symptoms are not necessarily the best.

But if you look at clinical bleeding, if you were to choose a trough, the work of Den Uijl with moderate hemophilia, looking at endogenous

level and number of bleeds, suggested that about 20 percent factor VIII, you would have very few bleeds without significant trauma.

These are negative binomial analyses. And I just want to point out, with hemophilia studies of joint bleeding, you have lots of people who have zero bleeds, and then you have tail-outs to the very high numbers. And this distribution makes it the most difficult to get accurate statistical modeling.

Well, I kind of edited the work of Mike Soucie, presented at ISTH in 2015. He came out with a conclusion, looking at factor IX in yellow and VIII in the dashed black, that 15 percent would be an optimal level. And 15 percent works pretty well for the adults, but if you want to prevent the joint damage while cartilage and bones are still growing, you have to focus on those growing-aged children, and 25 to 30 percent actually looks like a much better level to be targeting.

This just happens to be WAPPs PK curve. I'm not talking about inhibitor tolerance, but to

show that in factor VIII replacement, you have peaks, and you have troughs, and you have area under the curve. And even with the extended half-lives, the longer the interval between infusions, the longer time you're at a very low level.

If we are to consider unrecognized bleeding or oozing into the joint as being a significant pathogenesis of joint disease, then those curves of long tails are not necessarily optimal. If you were to consider that peaks are important for trauma, for sports, for surgery, then a consistent level at 15, 20 percent is also not going to work well.

With a standard replacement, we can manipulate this. This is a boy with a tolerized inhibitor on 30 per kilo every other day, and to play soccer, instead of taking 30 per kilo 3 times a week, he devised the 1 30-per-kilo dose, while he has 3 15-per-unit kilo doses, and has a daily dosing for 5 days a week and none for 2. And he's able to increase the area under the curve. He's

able to reduce or elevate the level of the trough, reduce the time in the shoulder and have no bleeding.

So we know that the counterpoint down side of this is that it's very frequent IV injections, and that's very difficult to tolerate over time. But it's more recreating the physiologic state of being able to be high and low as you need it.

This is a really elegant work of Carolyn Broderick from Australia, where she looked at sports participations in people with hemophilia using the NHF categorization of level 1, 2, or 3 sports, and 3 is the most vigorous. She found that at a factor level of about 35 percent, your increased risk of bleeding was very modest. It was only 1 and a half to 2 times that of sitting in a chair reading a book with severe hemophilia; so that's a very acceptable rate.

Her work would suggest for an active boy being 35 percent at the time of activity. Another graph she showed was that almost all bleeding is within an hour of the active participation.

So is hemophilia in the 25 to 30 percent range optimal for therapy to consider both safety and efficacy? Our future projections are based on our experience with the disease and with our imperfect treatment, so we really don't have the data to predict that.

I want to suggest that clinical bleeding predicts the onset of joint disease. So whether you'll have joint disease or not is very well predicted by the number of bleeds, but not the severity of the damage.

Again, this is the subclinical bleeding, and talking a little bit between MRI and ultrasound MRIs, the gold standard, very good with bone and cartilage, excellent on soft tissue. It's a long study, expensive, and not always available, while ultrasound is a point-of-care test.

It's available in the clinic. It's inexpensive, but you can't image the central joint structures where the joint bleeds actually occur. It's operator dependent. It's tricky to distinguish synovial fluid from hemosiderin, and

it's a very, very nice discussion of these pros and cons by Dr. Soliman from Andre Durie's group at the University of Toronto in sick kids.

Just to point out a little bit, these are normal ankle, and this nice dome on the talar dome, you see it flattens when you get a lot of bleeds. And these little white dots are cysts in the bone. It's also very good. The bright white is fluid and this black is synovium.

So these chronic changes over many years are very well-picked-up by MRI, but they're not good for a 1-year or 2-year study to show you're not going to get that interval change quickly.

In ultrasound, this is a clinical study done on a little 5-year-old boy whose parent was using extended half-life factor VIII twice a week at the dose recommended on the package insert, feeling that she was giving her boys cadillac treatment, and yet this widening in the right knee joint, compared to the contralateral joint, was representative of fluid in the joint, and this soft tissue in here is some clotted blood in the knee.

These findings were present in both knees and both ankles, so this little boy who had no evidence of joint bleeding, obviously had imprints on ultrasound that he was oozing or having some bleeding into joints, and that was not an extended therapy.

Just to show that extended half-life products so far have not really been able to extend the time without a significant time at a low trough; whereas with factor IX, extended half-lives, we've done a lot better and can maintain a trough near the gold standard.

In conclusion, factor level is a key endpoint, but there are differences, fundamental differences in therapies that do or don't have peaks. Longer-term secondary endpoints will be better assessed in young patients with less pre-existing damage.

We need patient-reported outcomes. For factor IX at target level, as close as we can get to the normal range is desirable, but I think factor VIII, for all the reasons Dr. Montgomery

discussed, requires more data accumulation, and we don't know yet what the optimal therapy will be. And I'll close right there. Thank you.

(Applause.)

DR. BOUCHKOUJ: Thank you, Dr. Johnson.

We're just going to put out the questions again and just ask for your feedback to answer the couple of questions that we asked before, and see if you've changed your mind after the presentations.

So again, this is a 30-year-old male with severe hemophilia B, who has moderate activity, and what would be his optimal constant factor IX level to reduce his risk of bleeding.

(Audience responds.)

DR. BOUCHKOUJ: Okay. Following question? This is the 16-year-old with severe hemophilia A, who is active, and what would be his optimal factor VIII level.

(Audience responds.)

Panel Discussion

DR. BOUCHKOUJ: Thank you.

I guess what we can do; perhaps I can ask our speakers what would be your answer to the questions. Maybe Dr. Manco-Johnson, if you want to comment on that.

DR. MANCO-JOHNSON: I would say, with factor VIII in the second boy, probably I would say 40 to 100 percent, if we were confident that we weren't going to 200 percent because this is in the normal range, and he already has 2 vulnerable joints.

DR. BOUCHKOUJ: And for the first question, do you have a --

DR. MANCO-JOHNSON: Optimally, a cure is a cure, and I would like to see people in the normal range, although I think, from what we know, that 35 percent for most things, except surgery, would be acceptable.

DR. BOUCHKOUJ: Thank you.

How about your thoughts, Bob?

DR. MONTGOMERY: I think on the first patient, I would think 35 percent seems the ideal level, and the second one, I think the

normalization of a child to be able to do athletics is important and think that it does carry with it an added burden of need of clotting factor. I think that really probably is over 35 percent, but I'd probably shy away from 100 percent.

I suppose 100 percent, without having any acute phase response.

DR. BOUCHKOUJ: Maybe I can ask Dr. Montgomery, does the result circulating factor VIII or IX level after gene therapy result in the same physiological thrombotic risk as with endogenous factors?

DR. MONTGOMERY: Say that again.

DR. BOUCHKOUJ: The result in circulating factor VIII or IX level after gene therapy, do they have the same effect of thrombotic effect as the endogenous factors?

DR. MONTGOMERY: I think there's still a lot to be known, so I don't know that I have the answer for that. I think that, ideally, you'd like to produce the protein in its physiologic cell, and that hasn't been done for factor VIII for a variety

of reasons, and certainly has been done for IX.

But how important that is I think is an issue. We probably have for years planned surgery, trying to correct patients at the time of surgery to 100 percent, not recognizing that the normal patients that have surgery probably have 250 percent factor VIII at the time of surgery.

So I think the physiologic importance of that stress response is more intuitive than it necessarily is highly driven by science.

DR. MANCO-JOHNSON: I was going to say, one problem we're dealing with today is that the range of motion in functional outcomes of joint surgeries, which are heavily used by adults with hemophilia, is less than patients who don't have hemophilia, and the musculoskeletal community of the World Federation believes this is due to an intense inflammation related to lifelong bleeding.

So again, if we're going to improve adult surgeries, we have to start in childhood and remove that early inflammation and damage.

DR. EHRLICH: Can I ask you a question,

Dr. Manco-Johnson, about the idea of starting early in childhood? So in light of novel therapies, keeping in mind that you can show that early prophylaxis is better, but we haven't yet shown for the novel therapies if those actually reduce joint damage, when do you start to think about using something like Hemlibra in a child? Would you start with standard prophylaxis, or are you as a clinician considering moving Hemlibra earlier?

DR. MANCO-JOHNSON: So we have two issues, the highest rate of intracranial hemorrhage and epidural spinal hemorrhage. These life-altering hemorrhages are in infancy and early childhood. So I think Hemlibra does offer the opportunity to prophylax a child before they're weight bearing with a delivery route that's very possible.

So we don't have data on doing that yet, how effective it is, but theoretically, I think it's very attractive. And then, in terms of later childhood, I think the subclinical bleeding -- I like to call it micro bleeding -- probably starts when you're weight bearing.

So while Hemlibra could be a bridge in very early weight bearing, we don't have data yet if the current doses are high enough to really prevent the kinds of stresses on joints that need to be measured and need to be studied, and possibly factor VIII therapies could be more effective then.

DR. EHRLICH: Do you, in light of the development of antidrug antibodies, even though the experience so far is that those are rare, consider the possibility, once you develop an emicizumab antidrug antibody, then you've sort of lost the ability to use that later in life, that you should consider maybe saving that for later, when you've exhausted other therapies?

DR. MANCO-JOHNSON: I like to front-load therapies to get children to grow in a healthy structure and function. And it's not that I don't worry as much about adults, but I think the morbidities of adults can be better managed if you enter adulthood with a good body.

DR. SHARMA: I have a question for Dr. Manco-Johnson. Could you comment on how can we

best capture the subclinical or microbleeds in the context of a clinical trial?

DR. MANCO-JOHNSON: Yes. I showed that picture of ultrasound. I'd like a show of hands here. How many thought that looked like mumbo jumbo?

(No response.)

DR. MANCO-JOHNSON: No? Well, they're not as black and white, clearly beautiful, as the MRI image is. And I think they are operator dependent, and we're going to need a lot more training, a lot more standardization, a lot more validation before they're a good clinical tool.

On the other hand, with the ultrasound, you can see fluid in the joint, and actually, that can be pretty well characterized. My husband did a lot of work in developing ultrasound, and with the ultrasounds and MRIs, he used to look at the joints of young children with hemophilia and say 10 percent of the joints have too much fluid. It's very minor, but objectively, you don't see this in healthy children.

I know now that 10 percent of children had subclinical bleeding in their joints, and that's what he was seeing. And he kept feeling a little -- he read the outcomes of the joint outcomes study and he was apologetic about it. But he said it's just more. I don't know what it is. I don't know why, but this is more than you should see.

But I think, for a clinical trial, ultrasound can show are you having a little bleeding now, because I don't think in a registration trial, we have the time. You need 5-10 years to look at MRI outcomes. But if you're accumulating fluid while you're on this therapy, then this therapy is not effective.

DR. SHARMA: Thank you.

DR. BOUCHKOUJ: Just by a show of hands, how many clinicians do we have with us in the room?

(Hands raised.)

DR. BOUCHKOUJ: In your practice, do you use ultrasound as point of care to evaluate bleedings on a regular basis?

DR. MANCO-JOHNSON: Yes.

DR. BOUCHKOUJ: Thank you.

I guess one question for Dr. Manco-Johnson I have, as we get better at improving, minimizing joint bleeding in general, would you recommend that measuring joint outcomes may be needed to assess long-term impact on treatment, for long-term treatment?

DR. MANCO-JOHNSON: Absolutely. And I think something like MRI, if you had a standard time at 18 years or 30 years, I think that that would be a gold-standard outcome right now because you can look at the effect on the center of the joint.

In ultrasound, you can see cartilage and bone abnormalities, but only in the periphery of the joint, but an MRI has to be reserved to a few time points and you need a good interval from baseline to outcome.

DR. BOUCHKOUJ: Are there efforts among healthcare providers to standardize the way these are assessed, the joints are assessed, in terms of

for recruitment of trials and so on?

DR. MANCO-JOHNSON: Yes. I think the International Prophylaxis Study group that was started and headed by Victor Blanchette at SickKids has done a lot of work to develop and validate physical joint scales for both adults and children, and then took on MRI, and they're taking on ultrasound.

I know Dr. von Drosky [ph] is also working on that, but I think that Dr. Blanchette's groups are multicontinental, multinational, and have a very wide interdisciplinary input.

DR. BOUCHKOUJ: Thank you.

I think what we can do; maybe we open up for questions. If you guys have any questions, please come to the microphone, if you want to ask the speakers and panelists.

We have some questions from Slido. You can submit your questions on Slido as well if you are listening online.

DR. EHRLICH: I just want to point out, there are a couple questions already on Slido, but

I think they'll be better addressed in a later session. So we're not ignoring you. We'll just bring them up in the appropriate session.

DR. BOUCHKOUJ: Question?

DR. GOLDING: I'm Basil Golding with FDA); a question for Dr. Manco-Johnson. You alluded to bone markers and bone disease in the hemophiliacs. Could you expand on that and tell us what you found, and whether you think that that is something we should look at in clinical trials?

DR. MANCO-JOHNSON: Yes. Jason Taylor, when he was at University of Oregon Health and Sciences University, did a lot of work. And although there were different patterns between factor VIII deficiency and factor IX deficiency, he generally found an increase in osteoclastic activity and a decrease in osteoblastic activity when the factor level was severely low, and then after replacement, he found a reversal or normalization.

For many years, we had known that people with hemophilia have decreased bone density.

Naively, I thought that because of joint disease, individuals were not doing as much weight bearing and this was a function-structure relationship. But he then gave a biochemical explanation that thrombin generation may also be necessary for the deposition of calcium into cartilage, the cartilage matrix.

This I think is a more subtle, maybe shorter-term marker that we could follow in clinical trials because, obviously, we would want optimal mineralization of our bones.

DR. BOUCHKOUJ: Thank you very much for our speakers, and we will move on to the following sessions.

Laurel?

Session 3

Moderator - Laurel Menapace

DR. MENAPACE: Shifting gears, we're headed into session 3, which will be an overview of patient-reported outcomes as I previously discussed in my introduction. It is my distinct pleasure to introduce Dr. Elektra Papadopoulos, who serves as

the associate director of the clinical outcomes assessment staff in the Office of New Drugs in the Center for Drug Evaluation and Research.

Her staff and office provide consultation to CDER review divisions, as well as other FDA centers on clinical outcome assessments regarding their development, validation, interpretation, and overall suitability to support regulatory approval of labeling of new hemophilia drug products.

Dr. Papadopoulos, can you come forward? She'll be providing a brief overview of patient-reported outcomes, so sort of broad-sweeping strokes before our other speakers present their information. Thank you.

Presentation - Elektra Papadopoulos

DR. PAPADOPOULOS: Thank you very much, Laurel, for the kind introduction. It's my pleasure to be here this morning.

As Laurel mentioned, our group works across therapeutic areas. We focus on measurement issues with regard to clinical outcome assessments of which patient-reported outcomes are one type.

Without further delay, before I delve into the details of clinical outcome assessments, I always like to take a step back and remind ourselves of, really, what are we trying to accomplish. I think this really sets the stage nicely in terms of what is a patient-centered outcome. These are really outcomes that are important to how patients survive, how they function, and how they feel in the here and now in their daily lives. In the case of patients who can't express this, sometimes we have to rely on caregivers and others.

Now, this was referred to in earlier talks, but our mandate at FDA when we're making drug approval decisions is to really weigh the clinical benefit against the risks of a medical product. Clinical benefit as described here is a positive clinically meaningful effect of an intervention on how an individual feels, functions, or survives, and clinical outcome assessments are the tools that we use to measure the clinical benefit of medical products.

Importantly, how we describe this clinical benefit to patients, providers, and other stakeholders is determined by what we call the concept or the outcome that was measured. This slide was shown earlier, but it shows the array of types of outcome assessments that we use to assess clinical benefit. Again, we call them clinical outcome assessments.

Importantly, patient-reported outcomes are not the only types of patient-centered outcomes, and very often we have to rely on a variety of clinical outcome assessments in a complementary fashion to really demonstrate the evidence of clinical benefit.

For example, if we need clinician judgment to make an assessment, we would use a clinician-reported outcome, or in the case of young children or those who may have cognitive impairment and we'd like to get a measure of how they're functioning in their daily lives, we may need a caregiver assessment called an observer-reported outcome assessment. Oftentimes, we'd also like to observe

patients performing specific tasks in a standardized setting, and in this case, we would use a performance outcome.

Now, of course the focus of this session is on patient-reported outcomes, but we should not forget our youngest patients who may not be able to provide self-report.

How do we review clinical outcome assessments? Essentially, we ask the question, does the instrument measure the outcome of interest? Our regulatory standard is, is the instrument well defined and reliable? Is it appropriate for the target population, for the target indication, and does it have adequate measurement properties? I'll get into that in a little more detail.

The 2009 FDA PRO guidance defines good measurement principles to consider when we use these tools to provide evidence of clinical benefit, but importantly, all clinical outcome assessments can benefit from these good measurement principles, so they don't really differ

fundamentally.

Important to remember is that this guidance provides an optimal approach, but other approaches may also be considered and used depending on the situation, and we always need to exercise regulatory flexibility and judgment to meet the practical demands of medical product development.

Now I'll go through some of the key characteristics that we evaluate when we're looking for adequate and well-controlled assessments. First is content validity, and this is really critical from a regulatory perspective because it's important for labeling claims.

Our labeling claims must be accurate. They must not be false or misleading, so content validity is critical because it really tells us are we measuring what we set out to measure; are we measuring the concept that we think we're measuring.

This measurement property is supported by qualitative and quantitative evidence, so very often, we'll do qualitative research with patients

in the target population to document this measurement property.

Other measurement properties are largely quantitative in nature, and importantly, these can't be really interpreted unless you first have evidence of content validity. Measurement properties such as reliability or how reproducible the measure is, construct validity, which essentially is the measure associated with other variables as we would expect, an ability to detect change, these are all critically important, of course, but they tell us really how well we are measuring. They don't necessarily tell us exactly what we're measuring unless we have that content validity piece first.

Now, I'd like to just highlight some common issues that we encounter when we're reviewing clinical outcome assessments for their use in drug development. First, we ask ourselves, is there input from the relevant stakeholders, and if not, we may be omitting what is most important and relevant to those patients. We may include

irrelevant questions in our measures; the instructions, questions, and response options may not be well understood.

We next consider is the instrument appropriate for the study design, the population, or the research question. If we don't have this piece, the measure may be poorly matched to the severity of the patients, so that may hinder ability to detect change.

It may not be a reliable, valid, or responsive to change, and it may capture something that's important to patients, but not really what the drug is targeting or what's expected to change in a clinical trial with a therapeutic intervention. We also ask is the instrument's concept clear and well-defined, and this is of course important for labeling considerations.

I just wanted to highlight this meeting. It was a public meeting, part of the 21st Century Cures patient-focused drug development meetings, and it occurred not only with hemophilia A but also other heritable bleeding disorders in 2014. You

can find online a Voice of the Patient report, which faithfully summarizes the input that we received from that meeting.

I've shown on this slide some of the very important concerns that patients experience, of course including unpredictable bleeding; joint soft tissues, muscles, and brain; limited mobility due to joint pain and deterioration; and the participation in social and work life are extremely important. All the psychological issues of course are critical.

This slide I won't go into detail, but all three medical product centers here at FDA have a multitude of ways that we can engage with our stakeholders, not only in the context of a drug development program, but also we have meetings, and there's also a qualification pathway where we can provide advice on the development of tools for drug development.

I just have some closing thoughts, and that is a clinical outcome assessment development and implementation, it's not an easy endeavor, and it's

really a multi-stakeholder, multi-disciplinary endeavor. We have pathways for review and advice, and we're very importantly open to multiple approaches to instrument development or modification. Very often, we need to consider how do we leverage existing measures, or if we don't have appropriate existing measures, we may consider modification or development of new measures.

With that, I thank you for your attention.

(Applause.)

DR. MENAPACE: Thank you, Elektra.

It is my pleasure now to introduce Dr. Christine Kempton, who is an associate professor in the Department of Hematology and Medical Oncology at Emory University School of Medicine, where she is the director of the Hemophilia Center of Georgia Center for Bleeding and Clotting Disorders of Emory.

She also serves as the regional medical director for the southeastern region of the Hemophilia Treatment Center Network, and her clinic and research focus is on hemophilia and its

complications.

Dr. Kempton is going to speak about specific patient-reported outcome instruments and tools that have been utilized in hemophilia studies as well as recent clinical trials. Thank you.

Presentation - Christina Kempton

DR. KEMPTON: Thank you. I appreciate the invitation to be here today and to speak with you, and here are my disclosures as well. Before I get started into the specific instruments, I want to talk just briefly about why we might care about PROs with maybe a little bit of my editorialization that brings together some of the discussion here today and adding into the overview of selecting PROs for clinical trials.

Then I'm going to dive into the SF-36 and Haem-A-QoL tools that have been used recently, and I'll talk about them in more detail, with some comment of using these two tools as well in clinical practice.

Just looking at kind of why we might think about using PROs, this is where I think about

hemophilia and what we're doing in the context of Maslow's hierarchy of needs and that there are some parallels with how we're talking about hemophilia.

We long ago have gotten done with treating major bleeds and then more recently done a pretty good job at preventing bleeds. Now, we're talking more about how to impact disability, moving up in this hierarchy, how we're preventing disability by things that maybe we can't see exactly in our subclinical bleeding.

But I would also submit there is an even higher level to that, even when we get the function down, that contributes to that anxiety and depression, and that's where we meet these patient-reported outcomes and quality-of-life measures to really understand the full impact to the patient. And even when we've got good levels that are preventing disability, if we're not curing the disease in its entirety, we will still have impact of the disease.

So PROs in clinical trials can be used for a variety of endpoints. They can inform clinical

decision making. Clearly, they can be part of pharmaceutical labeling claims, which is what I'll focus on some here today, and could impact product reimbursement and influence healthcare policy.

To support these activities, we need to use the appropriate PRO instruments, and ideally these PRO instruments are supported by a conceptual framework. The conceptual framework helps to illustrate how concepts and instrument domains really hang together, and this really supports the face validity. This should make sense to a content expert how all these domains interact with each other.

As already mentioned, it's important for the instruments to be validated to be reliable, meaning they have retest reliability. There's internal consistency questions within a domain and looking at the same construct. Both content and construct validity is measuring what we want it to measure and it also fits in with other tools that we already know. If another tool is measuring the same quality of life, they should be going

together. If they're measuring something that's totally there that should be different, we should be seeing them divergent, and that's divergent validity.

It's also nice that they're able to detect change. If we can't detect change as we make changes in medical treatment, they're not going to be all that useful in our clinical trials. Ideally, they'll have limited respondent and administrator burden, which is another important component, not just in clinical trials, where we accept a lot more respondent and administrator burden, but as we move into clinical practice as well, that's more key.

Ideally, they're able to impact clinical care. I think we're missing opportunity if we're using PROs in clinical trials that can never translate into clinical care.

So PROs can be generic versus disease specific, and there are some advantages and disadvantages to each of these. With generic PROs, the advantage is maybe that they capture more

common health-related, quality-of-life domains and really allow comparisons to a normative population; how close are we getting to a normal quality of life? Disadvantages are that they might not be sensitive to changes over time.

With disease-specific quality-of-life measures, they may be more sensitive to specific symptoms experienced by patients. However, they may miss domains affecting the patient, but unrelated to the disease under study.

Before I move into specific measures of health-related quality of life, I thought it worthwhile just to touch on the conceptual framework and illustrate it here, the conceptual framework for health-related quality of life.

In this framework, we see that there is a cascade of impact of biological function, impact systems, impact functional status, general health perception, and then overall quality of life. All of these domains are then impacted by characteristics of both the individual and the environment. This is what our health-related

quality-of-life tools are trying to understand, how these work together and impact the overall quality of life.

I'll transition to more specifics on the SF-36. It was a product of the medical outcomes study that was conducted in the 1980s and was a 4-year study examining specific influences on outcomes of care. There are originally 149 items. They ultimately then reduce down to a short survey, tried to include just 20 items, but there were significant floor effects.

In conjunction with the RAND Corporation, the SF-36 was then published in 1992 and has gone through several different versions. In use, you'll see version 1 and version 2 as well.

It's considered a general measure, and it has 8 health concepts: physical functioning, bodily pain, role limitations due to physical health problems, role limitations due to personal or emotional problems, emotional well-being social functioning, energy fatigue, and general health perceptions. It has asked patients to evaluate the

specific questions over the past 4 weeks.

It scored on a scale of 0 to 100 for each of the domains, 8 domains and 3 summary scores. The higher the score, the better the health. Scoring does require recoding and averaging in a specific domain, so it's not something that you can just add up as you're just looking at the responses. As I stated, there are 8 domain scores and 3 summary scores, a physical component score, a mental component score, and an overall health score.

It's been well-validated and translated in over 50 languages, and again, it's validated in numerous disease states. However, it's only recently been actually validated in hemophilia specifically in the PFIX [ph] study.

This demonstrated good internal consistency. You want to see a Cronbach's alpha of greater than 0.7 to demonstrate good consistency. It has good test/retest reliability. It demonstrated known group validity as well as content validity with correlations greater than

0.6.

It has been used as part of labeling in the Advate clinical trials. It's definitely used in lots of other clinical trials in studies as well, but I'm just going to focus on what's been used in the product labeling here.

We can see that with Advate prophylaxis, there was improvement in bodily pain domain and the physical component score in patients receiving prophylaxis compared with those on-demand therapy. And this is what we would expect to see as we're reducing joint bleeding and improving our short-term functioning.

I think it's relevant, and what is part of the benefit of this quality of life is that you might be able to then evaluate maybe some balance measures as we're getting more and more aggressive, particularly with prophylaxis, and the demands of the care may be more complicated or there are some other downstream effects.

Although this wasn't clinically relevant, I do think it's just notable that the mental

component score has a point estimate that was in the negative direction; again not clinically significant, but this is maybe a way that one can look at balancing or are therapies actually having some downstream negative effects that maybe we didn't quite understand.

Typically, quality of life is a secondary outcome in clinical trials. This is an interesting meta-analysis that looked at the concordance of the primary outcome with changes in health-related quality of life as measured by the SF-36.

We can see here there were 21 studies that had a primary outcome that was significant with the SF-36 that was significant. So that was a concordance of about 65 percent, whereas 25 percent had a non-significant SF-36 in the face of a significant primary outcome. So they don't always jive together, which I think is probably, then, one of the challenges of interpreting the results and what it means and challenges for you guys at the FDA.

So of the 33 studies that had

non-significant results as well, about 69 percent of them also had a non-significant result in the SF-36. So about two-thirds or so are concordant with the primary outcome.

Shifting gears to the Haem-A-QoL A, which is a disease-specific measure, it was developed in 2004 in adults with hemophilia, and there is a corresponding questionnaire in children as the Haem-A-QoL. The measure was developed using qualitative interviews of patients and physicians, and the initial draft contained 159 items. Pilot testing took place in 10 Italian hemophilia treatment centers.

The current measure has 46 questions in 10 domains, including physical health; feeling; a view of yourself; sports and leisure; work and school; dealing with hemophilia; and treatment, and it asks the participants to evaluate these areas over the past weeks.

Raw scores are transformed to a score also of 0 to 100, though lower scores indicate better health. This is in contrast to the SF-36, where

higher scores indicate better health and quality of life.

So the Haem-A-QoL, again, it's been used in more than two clinical trials, but they reported it in some of the labeling with Eloctate, and then more recently with the emicizumab. These two clinical trials were what were used to document the internal consistency, this is from the A-LONG and B-LONG study. The Cronbach's alpha was greater than 0.7 in 8 of the 10 domains. The two where there was less internal consistency was dealing with hemophilia and treatment.

In terms of validity, known-group validity was good except for family planning and dealing with hemophilia domains, and then convergent validities showed strong correlations with the EQ-5D-5 level, and the total scores physical health and feelings domains of the Haemo-QoL-A. There were moderate correlations with the HJHS with 5 domains and the total score.

The Haemo-QoL-A has also been used in the emicizumab clinical trial that supported its label.

We've been talking a lot about emicizumab today. This is the baseline information for groups A, B, and C, and I'll highlight here the physical health domain and the sports and leisure, which were the highest domains in these groups. And, again, a higher score is worse report for the patient.

The next two that were poor were the view of yourself and future, the next lowest scores. View of yourself have questions like, "I envied healthy people my age," with a report of a frequency, or "I felt comfortable with my body." Those were the types of questions that might be in view of yourself.

This was recently published as a HAVEN 1 study, that the total score showed clinically meaningful differences, which means there was a 7-point reduction in the total score. We saw that started at about 5 weeks and continued out to 25 weeks with the top dashed line being those on on-demand therapy. So we see clinically meaningful reductions or improvements in health-related quality of life as evidenced by reductions in the

Haem-A-QoL score.

This was also matched with a physical health score, where we see a 10-point reduction that's considered clinically meaningful reduction; and again, the same time frame. They were reached by about 5 weeks and persisted through the study.

The physical health score was what made it into the product label with the adjusted mean reduction of 32.6 points, or mean of 32.6 compared to no prophylaxis, which was 54.2. It's important, again, as I said, that the health-related quality-of-life measures can help us ensure that gains in physical domains are not offset by losses in other domains.

So although physical health was in the product label, I think it's always worth taking a look at the other domains. As we saw in the Advate label, the mental health component didn't really improve all that much, whereas we can see, at least with this Haem-A-QoL, with emicizumab, the view of yourself did improve, though clinically meaningful differences for these domains haven't been

established, and also the future also did improve as a domain.

There are also significant improvements in feelings in work and school, though again, clinically meaningful differences are not known.

So just touching on PROs in clinical practice, which Chris Guelcher will also talk on further, when considering them in clinical practice, to me, I like the idea of a value compass. With this compass, we take a balanced approach and consider not only clinical outcomes that are the hard ones and easy to measure, but also the functional health status, as well as satisfaction and total cost. PROs are really best suited to measure the satisfaction and the functional health status.

In clinical practice, we can use them for screening, monitoring, promoting patient-centered care, supporting discussions about patient priorities, promoting self-efficacy and adherence, and also supporting multi-disciplinary team communication and evaluating our quality of care.

Now, turning specifically to the ones that I've discussed today, the SF-36 and the Haem-A-QoL, some have been used. The SF-36 has been used pretty extensively in the orthopedic populations, a lot by payers and accountable care organizations, so these are really still looking at a population level rather than an individual level.

Instituting these into clinical practice, given the complexities with scoring, needing to transform, reorganize, et cetera, really requires informatics to support that on a real-time basis, as well as clinically meaningful differences need to be established in the hemophilia population to really know what these mean.

With the Haem-A-QoL, I couldn't find any reports of use in routine practice, and Chris may have some differences for us. It's a pretty burdensome questionnaire. You have to read very carefully each of the questions. It's kind of hard to scan through. The scoring requires transformation and, again, preventing use in kind of a paper format and requiring some level of

informatics. The lack of meaningful change cutoff outside of the total score and physical score limit its use.

In conclusion, as our treatments get better, we can expect more, not just functional improvements, but also improvements in our health-related quality of life and our psychosocial status. It's important to have these measures to assess these therapies and ongoing clinical trials.

As their use and importance in clinical trials and labeling increase, it's important that we move beyond really the ABR. Their use in clinical care will require improvements in informatics, identification of meaningful changes, and instruments with minimal response burden. Thank you.

(Applause.)

DR. MENAPACE: We'll now be transitioning to the patient speaker part of our session. And just to provide a little bit of background about what we asked our speakers to discuss today, we provided them the specific instruments that

Dr. Kempton just introduced, the Haem-A-QoL, which is a hemophilia measure as well as the more general SF-36 form.

So all patient speakers and advocates were given these surveys to review and look at the content and also provide their interpretation of the meaningfulness of these surveys as patients themselves.

We have four patient advocate speakers joining us today. Their biographies are listed in their packets. I'd like to introduce all four right now, including Mr. George Stone, Ms. Miriam Goldstein, Mr. Christopher Templin, and Mr. Shelby Smoak.

I believe Mr. George Stone has volunteered to provide his first talk. Please come to the podium. Thank you.

Presentation - George Stone

MR. STONE: Well, good morning. It's great to be with you guys today. This is an exciting time for those of you who are in the hematology world as doctors and nurses and in the lab work and

development. And it's a very exciting time for those of us who are patients, particularly in the developed world; maybe in the undeveloped world not so much.

Briefly, I'm a 65-year-old severe hemophilia A patient. I've gone through all the things that you would normally expect someone in my situation and age to go through. I'm the proud owner of 5 artificial joints. It was fun going through security to get here this morning. I was beginning to wonder, is this going to happen today or not.

The reason I'm going to kick this off is we've talked quite a bit about emicizumab or Hemlibra this morning, and I was a HAVEN 3 study patient, between March of 2017 and October of this year, so about 18 months. First of all, I'm very pleased to tell you, zero bleeds, so that's most important.

Now, with respect to these surveys, I regret to inform you that my view of these surveys is a little bit different. I had to complete these

surveys as part of the Hemlibra HAVEN 3 study. At first, I think it was monthly; then it was quarterly. I had seen these surveys before in hemophilia clinic. They aren't new.

I have to tell you, I realize these are translated from Italian, but when you see awkward language like, "Shelby, how are your swellings today?" Come on. Right away, as a patient, I go, "They can't be serious. Who designed this? Do they know anything about what they're asking? No hemophilia patient talks like that."

So that begins some skeptical view of the whole thing, frankly. So I think my number one observation is these need to be tweaked for the United States. They need to be put in proper English in America. Think about that.

Then the relevancy of these questions; well, to get on to the HAVEN 3 study, I had to be on prophy 3 times a week with Advate for a year. If you're trying to measure the difference that emicizumab is making today, you need to know my baseline. Am I coming in as a patient that's been

on prophy, or am I coming into the study as a patient that's been treating on demand?

You need to know a little bit more of my overall physical condition. I think it would be helpful to know, have some background for these questions that, in my case, I may not have a lot of pain in my joints because, well, many of them have been replaced.

A lot of this information that I would think you would want to know isn't captured. And I don't know whether it's captured by my hematology team and provided to the surveyors or not. There are many times when I'm trying to fill out the questions, I go, "You know, I'd like to add an explanation here," but I can't. You're limited to answering the questions that are put before you.

I just don't know that any of these questions are all that relevant when it comes to a patient outcome with respect to Hemlibra, in my case in particular.

What are the outcomes that I would think would be important? Well, for one, ease of

administration, is probably number one on the list. I did get some additional questions from Genentech during this study, and they were asking questions about are you satisfied with this treatment, are you okay with subQ?

One of the questions, which they didn't even really have to ask, was would you rather stay on Hemlibra or go back to factor? Really? I think maybe 5 percent actually said they wanted to go back to factor. I never quite understand that.

So I'm very good at one thing; maybe two things. One of them is internet research. And I found this little thing on the website, "Genentech's Hemlibra, clinical outcome assessment, data only partially swayed U.S. FDA.

"Hemophilia A drugs' labeling reflects data on physical function improvement because FDA deemed that portion of the Haem-A-QoL instrument fit for purpose, while other questions were viewed as insensitive to change or irrelevant. Review documents suggest agency was unimpressed with results from the health status instrument

frequently used in economic analysis."

What I found on the Web was, especially for the Haem-A-QoL, many countries are using it, more to probably convince their governments that it's worthwhile to help pick up the tab for factor for their hemophilia population rather than much else.

So in sum, I think it's probably a good idea that we revisit this issue, and I would say that I believe that the national hemophilia organizations, with a little prodding, probably would be willing to sit down with the FDA and industry, and maybe get a few hemophilia individual patients as well, and see if we can come up with something that's a little more direct, a little more pinpointed, and probably a little more accurate for what you all really need to know. With that, thank you.

(Applause.)

Presentation - Christopher Templin

MR. TEMPLIN: Good morning, everybody. Bear with me as I read off of my paper so I don't go off the reservation.

First of all, I'd like to thank FDA for giving me the opportunity to speak today about my thoughts and opinions on these patient-reported outcome surveys. It's important that the patient has a voice, so I'm pretty honored to be here. come with sort of a different view, I think, being old school, living with the way treatment was back in the day. It's pretty amazing what it is today.

We sort of went from the stone age, where treatment was I spent days, weeks, months in the hospital. I remember spending a whole year there once as a kid, and that was quite the year, to now having product at home available at a moment's notice and being able to pretty much infuse and get rid of all the waste stuff in 15 minutes, and your day really isn't impacted if I have the ability to pay for it, which that's not a topic for today's discussion.

But any day I wake up in a bed, not in a hospital or prison, must mean that my clotting factor is keeping me from bleeding, and my health insurance company has done their job to keep me

with medication because the Department of Corrections takes good care of their prisoners.

I often get curious to the actual true value and usefulness of survey-based data due to the ability of a person to embellish or dramatize how their hemophilia or bleeding disorder and/or their medication impacts their daily life on a minute-to-minute or day-to-day basis.

I believe that our needs to always group people into a box is sometimes a disadvantage because I don't think it tells the true story, sort of as I know a lot of folks that have mild hemophilia, and they're always told about how hard it is to get prophy, or how hard it is to get a new script, or they can't really tell their doctors the truth because they won't get factor. But somebody with severe hemophilia, they seem to have the truck delivering the product to their house every week if they need it.

I actually know severes who bleed like milds and milds that bleed like severes, so just going by the factor level is sometimes a detriment.

The goal should really be not to bleed. Whatever factor level it takes for an individual is that individual's factor level need. We're all different. I have a brother; him and I, completely opposites. We don't even look alike. Imagine that. He must be the milkman's kid.

But my biggest fear is that I'm not going to have access to my clotting factor because these new treatments might cause a company that currently makes a product to go off the market or reduce capacity. There's actually been some shortages in the factor IX space. Some folks I know have had some issues getting some product, and they had to switch to a different product, and it's sort of scary to know.

At least they're in this country. It seems like there's a lot of factor, but the price of it determines everything. So I think between products and even the level of care in the future, if the centers go away, doctors don't know what they're doing, try to give me factor VIII instead of IX, it's not going to help too much.

I believe that my feelings of pain, physical ability, anger, discomfort, all the nice stuff, changes on a daily basis, but is even made worse when you have to worry about is my doctor going to be there next week. The doctor I go to, she's pretty old, and I know she's getting ready to hand the center off to somebody else. Hopefully, those folks are committed to that facility because I don't want to have to travel further to get the level of care that I get.

One of the big problems that I see is people seem to inject factor, and they think that it's some superpower agent, and it turns them into the $6 billion man or $6 billion woman, because women do bleed, too. I have a daughter with hemophilia B. It's pretty crazy.

But I'm concerned that the level of benefit from these agents isn't able to be determined by checking a box because, like George said, maybe I want to explain, but there's no place to explain, or I think a little bit into the question. One was can you walk like a mile. And I was like, "I can

walk a mile if I maybe take a break like halfway through or take a little time."

I walk my daughter to the bus stop every morning, and it's funny, everybody else drives their car, and it's like a half a mile walk. But it's nice to get out in the morning, and get your gloves on, get your hat on, put your scarf on. While everybody else is driving, I get my exercise.

I get concerned that maybe we're moving a little bit too fast. We're trying to put everybody into the box. We're trying to really just make it bigger, better, stronger, faster, but we really need to think about the future a little bit more and just put the brakes on a little bit.

We have product. We don't want to substitute one expensive drug for another expensive drug, and here again, I'm talking about cost. I don't infuse. My daughter doesn't infuse this product because we just want to stick needles in our arms and cost the insurance companies money, bother the doctors with writing scripts, and all that stuff. We take it because it's a truly

life-saving drug.

I think we all need to remember that factor needs to be looked at as a life-saving, life-sustaining, keep Chris out of the hospital, out of the morgue, keep him at work so he can cause trouble there. But it's not a lifestyle drug, and I don't know of anybody with hemophilia that's taking this stuff because they want to take it. I mean, there may be, but that's few and far between.

So we just have to make sure that whatever surveys are used is something that is really being beneficial because I get a lot of surveys, and I get a lot of questions. And sometimes you're in a hurry, and you just check, yep, yep, everything's great. You go to a meeting. You get the survey. Everything's great. Here's your survey. See you later. Got to go. You want the people to take the time to put in the effort to do it, so you get the best bang for your buck.

I actually think a conversation-based method is better. When I go to the treatment center and talk to the social worker, or the

psychologist, or psychiatrist, they can actually tell if you're sort of BS-ing a little bit. The doctor might come in, "Yep, everything's great," bing-bang-boom, because it's 2 hours, 3 hours, and I'd rather go somewhere else.

You can learn more by having a conversation instead of just checking a box, especially as some of the questions are sort of hokey, like how are your swellings and stuff like that. So thank you for your time.

(Applause.)

Presentation - Miriam Goldstein

MS. GOLDSTEIN: Thank you. My name is Miriam Goldstein. My own disclosure is that I work at the Hemophilia Federation of America, but I'm here today in my personal capacity, and my views do not necessarily reflect the views of HFA.

I should also note that my personal experience with instruments like the ones that Dr. Menapace circulated for us to review is as a caregiver for now adult sons who are filling these surveys out on their own. So I speak from a

vantage point of a caregiver and a member of the larger hemophilia community.

It was very interesting to get the history of these tools because looking at them again in preparation for the session, they are clearly very, very dazed [ph]. They really seem to reflect a period before prophylaxis was commonly used in adults, and that seemed like a fundamental shortcoming.

They also are not inclusive, so one obvious area of omission is they omit questions that would be relevant to women with bleeding disorders. They take a one-size-fits-all approach to a community that is highly diverse, so baseline differences about age of patients, the stage of life, their childhood experiences all seem to be omitted from the survey.

Whether the clinician brings that in, in their own review of the instrument, is obviously a completely different issue. So personal goals and life experiences, also a high degree of diversity. And finally, the Haem-A-QoL was heteronormative, so

kind of shocking to come across questions about personal relationships and sexuality written in that way.

So even when these tools are applied to a very specific or limited demographic, it seemed to me, as a potential respondent, that it was very hard to tell what they were getting at. Are they trying to get at the overall quality of life of the respondent or to how someone is faring on a particular therapy, and that confusion sort of colored my reading of the entire survey.

In view of the complexity of hemophilia and the diversity of the population, I would agree with Chris and George that multiple-choice, check-the-box questions really don't capture the patient experience very well and that there's need for more elaboration. I realize that's intention with Dr. Kempton's remarks on how these have to be easy for providers to administer, but some kind of accommodation between those goals seems important.

I think George mentioned that patient groups might be able to come up with more nuanced

survey questions, and I will say from my work experience, I know that HFA has experience in working with patients to come up with patient-centered as well as patient-reported questions through CHOICE and CHOICE 2.0.

Finally, I'll just close by saying that while I recognize that the survey instruments are trying to capture a particular point in time, as a patient or a caregiver, the longer view is also really, really important to me; so some kind of longitudinal or follow-up is really important in terms of likely success, life outcomes on any therapy.

Again, I am familiar because of my employment with HFA's own patient portal, which does provide a tool for tracking patients longitudinally and even if they change providers. So I think I will end there, and thank you very much.

(Applause.)

Presentation - Shelby Smoak

DR. SMOAKE: Hey. I'm Dr. Shelby Smoake.

Although I live in the world of Dante and Milton, I think I understand most of this, so hopefully my remarks will be adequate to your needs.

I am a severe hemophilia B. I'll just start and say I've been in numerous clinical trials my whole life. I've experienced all kinds of therapies. Most recently, I was actually in a hep C trial. Happily, I was able to clear the virus, and that was a great, great day. I can't even explain that.

One of the things that I think we should think about -- and it was briefly mentioned, but I've wondered about the venue of these reports. No one has brought this up. But it seems to me you might want to consider a variety of venues. And I'll use myself. When I was in the hep C trial, if you know anything about D.C. traffic, it's horrific, and my PI in that study was only able to meet at like 3:30 or 4:00. And I can remember times where it was like I just needed to get on the road so I could get home at a decent time, so I did rush through them.

On one particular occasion, even being a PhD, they had "strongly agree," "strongly disagree," that kind of thing, I quickly did it. And they were switching them, and I didn't know that. So I got a call the next day of deep concern because I had answered the wrong way and I had to correct that.

So there can be mistakes. So I've wondered if trying to mix a virtual testing with what I do think is important -- I do think you have to have that face-to-face. I think sometimes the answers are skewed when you don't have that. So that would be a suggestion in that regard.

In terms of the therapy we have, I remember growing up, having two products. So to be here and to be experiencing the different available therapies the way our biologies respond differently, it's very valuable. It's very important.

I'll just briefly mention the metrics that are being used, we are engrained with factor level studies, and I think gene therapy studies are good

to use that, but I do like the movement of moving to ABR. But I think this is going to fail the real advantage of clinical trials because the real advantage to me is you're moving away from a rise-and-fall therapy, and it's really the troughs that destroy us, and the vantage of a clinical or a gene therapy drug is that that trough is removed.

So how do you capture the trough or how do you capture the sustained factor level? Thinking long term, how do you prove to the insurance companies that you can have a normal replacement factor product that is equal to a factor level at certain points in the spectrum, but the other one's going up and down, and gene therapy is not?

We have to figure out a way to make that kind of data capture because that's going to be the essence of selling this when it goes to market, and it's the real advantage.

I also can tell -- we know our bodies very well, and there's a certain point when you're in that trough, I feel like a rusty machine. I know that something's going on. I do like the idea of

Dr. Manco-Johnson maybe doing the MRI, something that's tangible, but there does need to be that capture.

In terms of the QoL, I can't emphasize enough the relationship between hemophilia and stress. So I think there needs to be questions that bring in stress. You need to look at how stress is maybe impacting the product, but you want a drug that's going to offer coverage during stress.

So if you want to remove it as a factor and say it was stress induced, that's one thing. But I know when I had an undue year of stress, I went from having an average bleed rate of 2 to 3 bleeds to something like 15 bleeds in that one year, one of which was a prolonged bleed of almost 8 weeks that sent me to total knee replacement surgery.

That stress incidentally enough was related to insurance. I ended up with $18,000 out of pocket that year. How does that happen? Well, you start the year with the $6,000 out of pocket. You change jobs, so that's another $6,000 out of

pocket. And then you find out your employer is not renewing on a January to January but an October to October, and so you hit another $16,000. So I have the equivalent of a car payment without the advantage of a car loan, and bleeds resulted.

I think those are some points to make. I think as far as PROs, I'll second my colleagues and say these really do need to be more specific. The rhetoric, the language is off, and we just need to utilize more appropriate language that is perhaps more specific.

So those are my thoughts, and I want to thank everyone for being here today and especially FDA for including us in this process. It's a very valuable thing. Thank you.

(Applause.)

DR. MENAPACE: I'd like to thank our patient speakers and patient advocates. Your feedback and input regarding these patient-reported outcomes is truly essential to the mission of the agency and the FDA, as well as the academic community. So again, we greatly thank you for your

participation.

Moving forward, I'd like to introduce Chris Guelcher. Chris is a pediatric nurse practitioner, who has been a hemophilia nurse coordinator at Children's National in Washington since 1997. Ms. Guelcher was promoted to lead advanced practice provider within the Center for Cancer and Blood Disorders at Children's National in 2017.

Christine will be providing some clinician perspectives today regarding PROs and PRO instruments and how we attempt to successfully incorporate them into clinical practice. Thank you, Chris.

Presentation - Christine Guelcher

MS. GUELCHER: So I want to echo previous speakers by thanking FDA for inviting me, and I will disclose that when Lori asked me, I said, "You don't really want me." I'm not an expert, but I think I've come to peace with the invitation in that I am sort of representative of my peers who probably aren't experts with patient-reported outcomes, and that's an area, a gap, that needs to

be addressed, so thank you for the opportunity.

I also want to apologize that I don't have a disclosure slide, but I have been on advisory boards for Genentech and Active Pharma and Novo Nordisk. None of that is relevant to today's talk.

Probably everybody has seen this model, which is the centerpiece of our model of care, with the patient being at the center and caregivers providing a multidisciplinary approach to address multifactorial issues in patients with bleeding disorders. And we know that that has reduced morbidity and mortality, and in the pediatric realm, less missed days of school, and for my parents, less missed days of work.

So with that as the background, how can we continue to include the patient's voice in the care that we provide? I think starting with the boots on the ground and where I think I can add to the discussion today is what is going on in a comprehensive clinic with a multi-disciplinary team.

If you think about adding in the patient's

voice to what's already at my center a 90-minute to 120-minute visit, and that's only with 5 core team members there -- so we have a hematologist, a nurse practitioner, a nurse coordinator, a physical therapist, and a social worker.

I originally had 5 slides because if I list everything that we all do, it takes up a whole slide. But I've in the interests of time pared it down just to highlight some of the more time intensive but important aspects of the clinic visit.

As Dr. Manco-Johnson mentioned, we do use clinical ultrasound to look at joints, and that has been a great tool to add to our visits and I think has really solidified what we're talking about, about joint changes and following bleeds over time for our patients. So that's been an excellent tool.

Our social workers, obviously, as alluded to by our patients, have an insurmountable task sometimes dealing with insurance issues and add that to a basic mental health assessment, the

impact psychosocially of this bleeding disorder diagnosis on the family unit in the community.

Then looking at the nursing component, traditionally, we've been looking at bleed assessment. And yes, that's gotten better on prophylaxis, but it's not absent. So it's important to be looking -- not necessarily we don't think of it in the clinical setting, at least at my center, as an annual bleed rate. I think that's more been a clinical trial definition, but it's important to try to characterize bleeding and how that's changed over time, and certainly with the advent of new therapies.

We also spent a lot of time talking about infusion teaching and home infusion, and that's changing a bit with the advent of some of the new therapies, and we're moving to what is an easier administration. But I think, as I said, walking in this morning with Miriam, we're going to have a generation of patients who may not be able to home-infuse factor when they have bleeds.

So how as nurses are we incorporating that

into our care? Bleeds may happen less often on these non-factor therapies, but it's that disadvantage, that familiarity with what is a bleed and how to treat it at home.

Then at the end of the visit, sort of tying it all up in a bow with discussing research, which has expanded exponentially, talking about new therapies, which is growing exponentially as well, and then recommending treatments. I think, as you've heard from the patients, there's not a one-size-fits-all approach. And while we may think as clinicians something is the latest and greatest, we have to respect the perspective and opinions of our patients that may evolve over time.

So all of that is a pretty meaty clinic visit. And not to belabor the point, there's a lot of actually hands-on implementation that's going on. We may be spending time going over any number of clinical trials. Somebody might be looking at consent for the CDC surveillance registry. They may be looking at an authorization for the ATHN data set. They may be eligible for industry

studies. There may be some investigator-initiated studies. All of that takes time to explain and make sure that our patients are fully aware of risks and benefits.

From a sort of practical standpoint, we offer patient choice, so we need to know from our patients if their insurance allows them what product they want to use and what home care they want to use. Our federal partners have some mandates of us, so we have the Patient Engagement Survey for our patients that are over 13.

At our center, we use transition guidelines, sort of a quiz approach that we've developed in our region to gauge where they are, what their understanding is of their disease state, and how that changes over time. Then for women with bleeding disorders, we also might be doing the Bleed Assessment Tool.

Either during or after clinic, the providers have some pretty big tasks. Maybe we're entering data into our clinical manager, which is our tool to track our patient visits. That could

translate into the 20 core elements that are part of the ATHN data set, which then translates into the hemostasis and thrombosis data set, which is a responsibility to a federal partner.

Patients that are participating in the CDC study, there is a CDC surveillance form that needs to be completed, and any number of ATHN, 1 to 10, that patients are participating in. Then of course, industry studies may be ongoing throughout the year with more frequent visits.

So all of that takes a lot of time and effort by the clinicians, so it extends beyond obviously that annual or biannual clinic visit.

Outside of just seeing our patients in clinic, it's important for us as clinicians to be aware of what's going on in the literature. And in the hemophilia literature, this is just a smattering of papers that are out there, many of which were authored by some of the clinicians that are here today.

We have lots of discussions of the landscape tools that are measuring different

aspects, and uniformly, everybody has said there are great tools out there. They are reliable and valid, but picking the right tool to meet your needs can pose a challenge. And then, of course, having so many tools then makes it difficult to measure from one study to the next if we're using different tools.

To echo what one of the patient speakers said, I think in the literature, the use at HTCs of these tools for investigator-initiated have been more to sort of demonstrate a need. The advantage of a tool like Haem-QoL-A is it's translated into a number of languages. In these two cases, these centers were able to take their data and compare it, so that is an advantage of using a tool like Haem-A-Qol, but it may be challenging to incorporate that into the clinical setting, which I think Dr. Kempton alluded to.

From my perspective, having the patient-reported outcomes in labels is an opportunity but it's also a challenge. One of the things that I spend a lot of time doing in clinic

is interpreting. Historically, it's been what is recovery study in a half-life? What is the area under the curve and how do you explain that? Now, I'm trying to explain how a level is not a factor level, but it might be on par to hemostasis and does that change the area under the curve?

So adding interpretation of patient-reported outcomes is just another way to try to meet a patient where it might be meaningful. So a patient that goes cross-eyed when I start talking about peaks and troughs, this may speak to someone. So it's important that we have that as an opportunity, but I think it may also be missing the mark. So I don't know that we want to put too much emphasis where it's not relevant. I guess we'll know more as these discussions happen in clinic.

I can say from just my current clinical use, this hasn't been the focus for most of our patients. They're really intrigued about the more classic reduction in bleeds currently.

I'll end echoing what Dr. Kempton said, that I fully respect that the patient is the center

of care, and I went into nursing because that's all I ever wanted to do. So I feel very responsible to hearing the voice of the patient, but I want to be realistic that in order to administer these tools, there has to be a way to present it where we're going to get meaningful information.

If my clinic's on Monday afternoon and I'm in D.C., so that same traffic. I have parents that need to get out of clinic, and pick up kids from school, and make dinner, so I need to be respectful that in order to get meaningful results, they need to have time to complete it.

Because there is so much going on in clinic, are patients just going to check boxes, and are we going to see results that are really based on survey fatigue? Then the impact on the resources at the treatment centers; we have a lot of -- I guess it's not fully fair to say unfunded mandates, but we have a lot of responsibilities to our partners.

So entering that data and incorporating the data, more importantly, into our plan of care, how

do we do that? If we're seeing patients once a year and they're filling out a survey, are they going to see that as valuable if I can't turn around and tell them how that's making a difference in their care or in the care of the community?

Ultimately, I think Chris alluded to this, patients that participate in clinical trials -- and George I think gave the other perspective -- may be coming at this use of clinical-reported outcomes differently.

If you're a patient that wants to be in a clinical trial and you've taken that approach, are your answers the same as somebody who's not in a clinical trial and coming to clinic? So I think we need to be cautious about the differences in why patients might be responding.

So with that, thank you very much for your attention.

(Applause.)

DR. MENAPACE: Thank you, Chris, for providing some real-world pearls of wisdom in terms of how we think of patient-reported outcomes in the

clinical setting, particularly for patients with hemophilia.

Moving forward, we're going to have more of a panel discussion with four of our internal reviewers at the FDA. We all have different job aspects in terms of how we review patient-reported outcome data, but basically, we're all interacting with stakeholders, whether it be pharmaceutical companies or patient advocacy groups, academic investigators who have questions about patient-reported outcomes and how best to utilize them in their own clinical studies or clinical trials.

I'd like to introduce two reviewers. Virginia Kwitkowski is the associate director for labeling in the Division of Hematology Products. In this role, she advises review team members and division leadership on methods for developing clear, meaningful, and scientifically accurate prescription drug labeling that conforms to regulations, guidance, and policies issued.

She is also a patient-reported outcomes lead for the Division of Hematology Products, and

we heavily rely on her expertise in this area, and she certainly has helped guide me in a number of challenging situations.

Ms. Kwitkowski completed her master of science degree at the University of Maryland graduate program, with a certification as an acute care nurse practitioner in oncology.

The second reviewer I'd like to introduce is Dr. Belinda King-Kallimanis. She is a psychometrician working in the Office of Hematology and Oncology Products, and she provides support to the three oncology divisions with respect to clinical outcome assessments as well as patient-reported outcomes.

She works on advancing science with respect to understanding how current clinical outcome assessment strategies in cancer clinical trials can be improved. Belinda has been working the field of COAs in patient-reported outcomes for the past 10 years across both academia and industry.

So I would invite Gini, as well as Dr. King-Kallimanis, to come up to the podium if

they would like to provide some further comments or thoughts about their reviewer's perspective. Thank you.

Presentation - Virginia Kwitkowski

DR. KWITKOWSKI: Thank you, Laurel.

I really appreciate being here, and I just want to thank, again, the patient representatives here. The information they provided regarding the clinical outcome assessment instruments that we shared with them are really meaningful and helpful.

I just want to start by saying that we expect that these instruments are developed with patient participation, and if they're not, if they're initially developed with clinicians, expert clinicians, they would be reviewed with patients. So it's disappointing to hear that we've managed to collect patients here that don't agree with the items, and that's very interesting for us.

So I when I'm looking at an instrument -- and again, I've been a clinical reviewer in the past and now I focus mostly on labeling and patient-reported outcomes -- we're

always taking into consideration our previous experience as clinicians and whether or not the instruments and the items in the instruments appear to be relevant to the patient's feelings and the experience that they have with their disease.

So we're looking at content validity from a very high level, but we're expecting that the development of the instrument actually looked at that in a very focused way, with patients, with clinicians who are experts in the disease area.

So those are some things that we look at as clinicians, is to sort out whether or not content validity has been established because that's the most important part of the instrument evaluation.

Other things that are really important, and sometimes where our regulatory goals my counteract what the patients want to see in an instrument, would be, there are some disease symptoms that are not really mobile, so you may have a permanent injury that is really important to you as a patient and that you would want that captured in any instrument that was drafted for a patient with

hemophilia.

However, if it isn't mobile, if it won't move with treatment, it isn't important from a regulatory standpoint because if you're rating it on a scale of 0 to 5, and you're rating it as a 3, and there's no chance of moving that, whether it be the mechanism of action of the drug or whether it's just a fixed deficit, we would not be able to see movement in that particular item, and that would be problematic, especially if it were incorporated into a total score. So we have issues with those as well.

I think that what's really important, sometimes we get submissions where we have instruments used to collect data, and there's actually no real good evidence of what the clinically meaningful change is; so when they say, "Look, our patients had a 3-point change on this scale of 0 to 5," and we have no data to support that a 3-point change is important to patients.

That information can be established in multiple ways, but if it's not established at all,

or it's not established in an adequate way, we have difficulty deciding whether we should put it in labeling at all because we really don't want to put non-useful information into labeling.

Those are my thoughts, and I'll just turn it over to Dr. King-Kallimanis.

Presentation - Bellinda King-Kallimanis

DR. KING-KALLIMANIS: Thanks, Gini.

I think what we heard from patients a lot in this session has been that the items have to be relevant, and I think this goes back -- if you look at the Haem-A-QoL questionnaire, you can see that there is evidence that it has reasonable measurement properties. But what we're hearing is that the questions are not relevant and that they may not map to a relevant research question.

So one of the things we've been pushing for a lot in IND applications that are coming in today, that the PRO questions being asked are actually being thought out a little bit more carefully. In the past, it's just been we want to investigate health-related quality of life, but how and what

elements of that are important and when is it important to measure that.

So we start to develop more clear and concise research questions, and we can then go and look for the right instrument versus put an instrument in that maybe captures a lot of the concepts that are interested, but not particularly well, and then try and fit a question to it after the fact. It's difficult, and we often then find ourselves asking questions that are not relevant.

So it's this balance between capturing concepts that are relevant and overburdening patients and having something at the end that we want to have an answer to. So I think that's where we're needing to move, and we've heard a lot of that today.

I think some of it's just that we're in a time period where patient-reported outcomes have become very popular, and we want to be able to include that information more in the label, but the trials were designed 5 years ago or something like this, and it wasn't such an important outcome at

that time. So we're sort of in this growing pains period, and I hope to see that change as we start to move forward.

(Applause.)

DR. MENAPACE: Thank you.

I think we've reached the end of session 3, and we're going to be opening up the discussion for a panel discussion. We do have a couple of Slido questions that we'd like to pose to the audience to kind of get the conversation rolling. But anyone within the panel or from the audience who has questions, feel free to come up front to the microphones once we're done with the question aspect of this segment.

DR. EHRLICH: I think the first question should be on your Slido on your phone, but I don't think we're going to display it here, but we'll display the responses when they become available.

The question is, prior to today's presentations, describe your baseline knowledge of PRO instruments and their use in hemophilia clinical trials.

DR. MENAPACE: Can you repeat that question again? Here we go.

DR. EHRLICH: Are the results there? They're displayed here.

(Audience responds.)

DR. EHRLICH: I think it looks like most people have answered now, so I'm going to close this poll. It looks like a significant number of people in the audience have had at least some experience and some extensive experience with PROs.

The next question should be coming up now. This next question that should be now on your Slido I think is perhaps a little bit of a loaded question. But the question is, is it useful to have patient-reported outcome information included in the prescribing information for specific hemophilia products?

(Audience responds.)

DR. EHRLICH: It looks like we have most of the responses. It definitely tilted towards the yes, but some nos. I think it would be interesting if we could break this down by people's roles in

product development, whether it's sort of patients versus industry versus FDA. There might be a different answer to this question.

Our third question is here now, so which of the following patient-reported information would you consider most important to include in the prescribing information? There's functioning, emotional health, ability to go to work or school, side effects, or other.

(Audience responds.)

DR. EHRLICH: All right. I think we have most responders now, physical functioning being the clear winner on this one. I think our next question is sort of the flip side of this. What do you feel is the least important to be included in the prescribing information? And "write other" is a little bit of a tricky one here.

(Laughter.)

(Audience responds.)

DR. EHRLICH: I think we have the bulk of responders now. Perhaps a surprising response here, and maybe this will come up some in our panel

discussion, but side effects seem to be the winner here. Just one more question before we go to the panel discussion.

How much time are you willing to devote to the PRO surveys that include relevant items during each study visit?

(Audience responds.)

Panel Discussion

DR. EHRLICH: I think we have most responders now, so a pretty decent spread here. It seems like 5 to 10 minutes is the winner, but a decent kind of bell curve on the amount of time being devoted here.

I think we can move to the panel discussion. There's one question on Slido that we can maybe start off the discussion with, and then we can maybe move on to other questions. But the question on Slido is does the FDA consider ABR as a PRO; and if so, how does one assess the reliability and validity? If not, how does it not meet the criteria of a PRO?

I can actually start answering this

question also, unless anyone else has comments. I think at least in CDER, which is where my experience is, this ABR as a PRO is shifting from what used to be kind of a clinician-reported outcome and is now shifting more towards a patient-reported outcome.

Our most recent experience has been with emicizumab, as you probably know, and in this case, they were developing a new electronic tool to sort of better capture bleed-related data as a patient-reported outcome.

This is an example that as the technology moves ahead, then the data that we're getting and how we review that data is changing. But certainly, in this trial, it was a patient-reported outcome.

In the development of this drug, there were a lot of discussions between the commercial sponsor and the FDA clinical review team as well as the COA team to develop this tool and make sure that it was answering the question that we needed it to answer to ensure that the tool was functioning as we

needed it to.

An interesting outcome, which was also presented at ASH this past weekend, was that I think it was a little bit surprising that what we had previously seen as a clinician-reported outcome was generally treated bleeds. And now with this tool, there was a much bigger report of untreated or all bleeds.

There was an improvement in this all-bleed category, but I think the rate of ABR with all bleeds was a little bit surprising, and we did a better job of capturing that with a patient-reported outcome.

DR. MENAPACE: Thank you, Lori, for responding to that question.

Just to follow up on the information you've already provided, in some ways, it was almost a little bit of a hybrid with electronic diaries that they used most recently in the HAVEN 3 and HAVEN 4 studies, where patients were essentially able to log and bleed-related and treatment-related data for a period of, I think, approximately 7 days they

logged, or every 8th day.

Then at each subsequent study visit, the investigator or clinical nurse investigator who was working with the patient had the opportunity to review that data with the patient. And if there was an error or an omission of a significant bleed-related event, go back and amend those diaries.

So it is interesting in the sense that we're heavily relying on patients to report their own bleed-related outcomes, which I think is novel and an important advancement in this field. But at the same time, they were still relying on physicians and other providers to help them translate bleed-related data and also help them if they had forgotten or omitted any bleeds in their electronic diaries.

MS. GOLDSTEIN: I just wanted to add something that I didn't mention when I was up there that is kind of on par with that. I think the opportunity to discuss patient-reported outcomes can't be understated and to get the context that

Chris was talking about, not just the checking the box.

I think something else to think about is if, for instance, the advocacy groups like HFA and NHF are opening up patient portals, how is that information going to be communicated, if at all, with clinicians?

I was a former board member of ATHN, and I am no longer on the board, but I've always been a proponent of having tools that communicate with our clinic EMRs, so that if a patient is documenting bleeds, that that's able to be communicated with the clinicians who can then put it in the context of the clinical picture and communicate with the patient about how that's impacting on things like missing school, and work, and their prophylaxis regimen.

So to not have double data entry and to have patient portals communicating with clinical manager, to have study forms that we can incorporate, I think all of that in the advent of EMRs is something -- there are opportunities there

that would make things much more smooth.

DR. MENAPACE: Great. I think we have a question from the audience.

DR. PIPE: Steve Pipe, University of Michigan. One of the themes I heard this morning so far was, within the clinical trials and the need to demonstrate some patient report outcome measures, the sponsors are limited to the validated tools that are currently in existence.

At least I would assume to see that those PROs end up perhaps in the label, but if we have some agreement that these tools aren't necessarily capturing the kind of information we need, particularly on the patient experience side, what's the agency's position on the ability to elicit that kind of patient experience in the context of a clinical trial, even if a validated tool isn't actually used to collect that?

So if we feel like we all need to get better patient experience as part of these clinical trials, sometimes the questions that need to be asked may be fairly specific. And I think

Mr. Stone gave a good example from his experience, where he felt that the validated instrument tools weren't really getting at what he was feeling for his participation.

So if experiential questions are collected in the context of a clinical trial, how are we going to see this information brought forward at the regulatory level?

DR. EHRLICH: I think that highlights an important question that got brought up throughout this panel discussion, and it's a difficult question. We certainly do have pathways available where sponsors can propose a new tool, a novel tool, and there are pathways to validate those tools. However, that can be challenging, that takes time, and you can't really validate the tool just within your own trial. They have to be validated in a larger perspective. So it is challenging.

I think we've presented these two surveys, and we actually don't have any allegiance to these two surveys other than that's what's been presented

to us, that we've only been able to review the data within the context of what's been presented to us.

I think we've highlighted here that these tools perhaps have problems that are insurmountable that maybe we weren't even internally fully aware of throughout the review. But I think we've also highlighted that what we were trying to do within the context of these tools that were presented was differentiating between what metrics are important on a more global lifestyle or lifelong perspective for patients, and what we can capture within a clinical trial, and what can be modified by treatment, as Gini also pointed out.

So for example, we included the physical functioning metric because that seemed to have a reasonable expectation that both represented patients' outcomes that could be sort of modified within the context of a 24-week trial and could be modified by a drug, where things like partnership and sexuality either couldn't be captured in a short period of time or couldn't be modified by the drug.

So we were able to use those tools and parse out some of what could be contextually validated.

DR. PIPE: I would also suggest that the tools that we have at our disposal right now are covering a very broad range of levels of care or at least how they're applied. So for instance, many of these instruments can be used in countries that don't even have patients on prophylaxis.

So to be able to use these tools and move the needle, so to speak, when you introduce a prophylactic therapy, et cetera, is not nearly as difficult as in a context where you might have access to more complete therapies. And going forward, if you look at where the field's heading, where you're going to get into gene therapy later, what the comparison is going to be against is really against optimized prophylactic therapy, and the ability to move the needle on that background with the tools that we have available would seem to be particularly challenging.

So I think, now even more than ever, the

patient experience and maybe drilling down into elements that are not even captured properly by these tools is going to be, practically, really important going forward.

DR. EHRLICH: Yes, I agree. I think some of the issues that were brought up such as ease of use, obviously, is going to be important to capture with the subQ administration. And then with gene therapy, obviously, it's a one-time administration, so maybe ease of use is not the right terminology for that but also can be important.

I think at the FDA, we look at things a little bit more globally, that we can take into context both the factor level bleed rate that's been captured as well as some patient-reported information, whether or not it's captured with these tools or other tools, to make our benefit-risk analysis.

DR. PAPADOPOULOS: I just have something to add, and that is I think patient advocacy groups are well-positioned to undertake either development of instruments or optimization of existing

instruments that could be used across medical product development, so that we would have standardized measurements that have been adequately tested with patients and have had that patient input piece.

The patient advocacy groups can really help foster that in a pre-competitive setting so that each medical product developer doesn't have to do that by themselves. And we do have a pathway for that to occur, where we can provide advice on tools that are being developed for unmet medical needs within a qualification program. And ultimately, these tools we expect to be made publicly available so that they can be used in medical product development broadly. So I think that's a really key opportunity.

MS. GUELCHER: I would just caution that advocacy groups are great, but they don't necessarily represent all of the patients. Hemophilia treatment centers see patients that may not be part of those advocacy groups, and we don't want to miss those voices.

DR. MENAPACE: Thank you, Elektra and Chris, for your comments. I believe we have one more question from the audience, and we are running into our break for lunch. So with this last question, we'll wrap up and conclude the panel. Thank you.

MR. SKINNER: Mark Skinner, patient with hemophilia, but also someone who does extensive research in the health outcomes field. I wanted to pick up on Steve's comment and then the last remark.

There was a core outcome set developed in hemophilia that identified a series -- at least 3 of the 6 elements were specifically patient-reported outcomes. We've covered ABR, but the two others were pain and mental health, the transformative aspect. Dr. Ragni mentioned the transformative piece earlier this morning.

Within the pain domain, I think that the group identified -- and it was the number one concern of patients coming out of the patient-focused drug development last year. Two-thirds of

the patients reported pain as the dominant outcome. It really hasn't been discussed today within the context of outcomes, nor within the pluses or minuses of SF-36 or Haem-A-QoL, both of which are deemed to be, at least by a lot of individuals, deficient and being able to differentiate between chronic and acute pain.

So now that we have a core outcome set, we're live, we're in the real world -- but that outcome set was developed in the pre-competitive space that was mentioned; but we're now in the real world and we're needing to collect that data with pain being the dominant outcome that the FDA was informed about -- what are the opportunities to bring in other instruments that would pick up the other elements of that core outcome set, to have them concluded?

Specifically pain, something that's more sensitive in terms of its occurrence frequency, differentiating how the drugs would change, and then bringing in the transformative piece since we now have those at least identified as core

important outcomes. Thank you.

DR. EHRLICH: I think within the FDA, there are always opportunities to have these discussions. We have mechanisms where commercial sponsors as well as patient advocates can just come and meet with us, and we can sit down and try to figure out a pathway to move these things forward. I know the COA staff does a lot of the earlier work in validating these tools and helping to incorporate these into clinical trials, but there certainly are mechanisms where we can meet and figure out a path forward.

DR. PAPADOPOULOS: The core outcome set that you referred was one that was developed in the context of use of gene therapies. My understanding of that is that the first stage of development was really having an agreement consensus around what are those concepts, what are those outcomes that are important to be measured in all gene therapy trials at a minimum, basically. It doesn't preclude other things from also being included. But at a minimum, those were the outcomes that were

decided upon. And my understanding is, then, now the next stage is to identify the actual instruments that will be measuring those outcomes.

So that's just a reflection on your comment. It's not complete yet. It hasn't been complete yet.

DR. MENAPACE: Thank you, everyone, for your comments. Just to echo everyone's sentiments, I think the FDA and the Division of Hematology Products, in general, is willing to engage with patients and patient advocates, and physicians, and physician investigators, as well as industry, to, as we previously referenced, move the needle forward in terms of patient-reported outcomes and clinical trials.

We'd be happy to answer any questions from any additional individuals over lunch or later on this afternoon, but thank you, everyone, for your attention, and we'll now break for lunch.

(Whereupon, at 12:26 p.m., a lunch recess was taken.)

A F T E R N O O N S E S S I O N

(1:20 p.m.)

Session 4

Moderator - Mikhail Ovanesov

DR. OVANESOV: Good afternoon, everybody. Welcome back and please be seated. Let's get started.

My name is Mikhail Ovanesov. I work for the Center for Biologics, Evaluation, and Research, also known as CBER. My office is the Office of Tissues and Advanced Therapies, OTAT, and my particular job at the Food and Drug Administration is the review of coagulation factor activity assays. I will facilitate this session today, a session on the use of coagulation factor measurements as surrogate endpoints in clinical trials.

Our agenda for today, just to go over it really quickly, there will be two presentations. The first one is on the analytical assays and reference standards, and the second presentation is on the clinical perspective on the assays used in

clinical trials. Then there will be a panel discussion. That's the second part of our session.

Our two presenters will be joined by three panelists. And together, the five panelists will represent the experts from the clinical labs in the United States and the European regulatory agencies.

There will be no questions and answers after each of the presentations. If you have a question to a presenter, please write it down and join us at the end of the panel discussion because we want to hear from you. We want our audience to participate in these questions.

Now that I went over the housekeeping items, I can proceed to introduce our first presenter today, Dr. Elaine Gray from the United Kingdom. Dr. Elaine Gray is working for the National Institute for Biologic Standardization and Control, NIBSC, with the Ministry of Product Health and Controls within the United Kingdom.

Elaine is an international expert in biological standards. She was personally involved in the development of the WHO international

standards for factor activity, many of which are used now for hemophilia diagnosis and treatment today. Elaine came to us from across the pond, and without further ado, welcome, Elaine. Thank you.

Presentation - Elaine Gray

DR. GRAY: Thank you, Mikhail, for this very kind introduction and also for the invitation to come here to speak. As my title indicated, I'll be talking about analytical perspective on methods and reference standards. This is my disclaimer.

Factor concentrates are biological medicines, and as we all know, it's dosing international units. There are a lot of advantages of the international unit. As we know, one international unit is typically found in 1 mL normal plasma, and that's how we define the international unit in the first place. This is equivalent to 100 percent normal in plasma.

Although we lay this international unit to normal plasma, the activity of normal plasma pool can change, and that normal pool from different labs are not the same. And even if you collect a

pool of plasma from the same lab, using the same donor over time, you'll find that actually would not be the same.

How do we know that it's not the same? That's because we have the international standard and the international unit. By comparing the different local pool to that, we find that there can be some differences.

For the international unit, once it's defined for the first standard, it is then fixed for subsequent replacement preparations. It is recommended that the local pools should be calibrated against the international standard or other reference preparation traceable to the international standard. This allows the laboratory to compare the level of activity.

It also allows us to potency label products in international unit and this international unit for the products that link to the plasma international unit. Therefore, this allows us to normal and deficient levels and helps the calculation of target levels for therapy.

Just to give you a quick example on how useful this is, this is data from the value assignment of the 2nd international standard for blood coagulation factor XI in plasma. First of all, this shows that this particular candidate was assayed against different local pool normal plasma, and you can see that overall geometric mean here shown, that 0.72 units per ampoule.

However, if this sample's assay by these 3 labs, as shown by the red circle there, you get about 0.65 unit per ampoule. However, the same sample assays in these other 2 labs, the value they have obtained were about 0.85. So you can see there's quite a wide spread of activity.

When we assay that same sample against the first international standard for factor XI, you can see that we get much better agreement, and the overall geometric mean, although not too different to that against the local pooled normal plasma, the actual GCV, the variability of the assay, came down to about 2 percent as opposed to about 7 percent. So this is really showing how good it is to improve

the laboratory agreement when we assay against a common standard.

So the role of the international unit is that it anchors down the potency labeling. This is very important in terms of ensuring the consistency of production. It is labeled in international unit and it's linked to dosing international unit. We know that for the products that are on the market right now, any of the products, in general, you can give more or less the same unit per kilogram body weight to raise activity by a very similar manner. So for factor IX, it's usually one unit of the product per kilogram body weight to raise activity by 1 IU per deciliter.

Ideally, the same type of assay method should be used for potency labeling and clinical monitoring. However, this isn't always the case. An example of that would be the factor VIII product in Europe has been potency labeled using chromogenic assay method. However, in the clinical lab, they're being monitored using 1-stage clotting assay.

The way that we prepare this standard, of course they have to be replaced from time to time. And you can see the history of the factor VIII concentrate standards here, which the first one was established in 1970, and now we are on the 8th international standard that was established in 2009. The characteristic of these standards tend to go with the availability of a product available at the time, so we went from intermediate purity to high purity material.

At the moment, the potency labeling of factor VIII and factor IX products, the plasma-derived and recombinant modified products are all traceable to the WHO international standard in international unit.

We talked a lot about functional activity assay today. We talked about the one-stage clotting assays, which is based on APTT. I don't want to go into detail about that, but we know that there's a lot of different APTT reagent with different phospholipid composition activators.

For the chromogenic assay, this is based on

using purified reagent, but we also have a lot of variations. For factor IX, there are two commercial assay kits, which is C-marked in Europe, but I understand that it's not registered in the U.S. yet. There are at least 6 commercial assay kits for factor VIII, plus there are a number of in-house assay methods.

I think we need to consider these two types of assays, as really within each assay type, there are a number of different variations, and they can be considered different assays.

These types of factor activity assay determinations require bioassays, which are actually relative potency assays. So it's not like a mass balance, where you just wait out something that we know what it is or it's not determined in terms of microgram or milligram. We require a reference standard.

The potency estimated for the test sample is relative to reference standard and based on the principle of assaying like against like.

In these assays, the reference standard and

test sample should have a similar characteristics, and the test dilution should behave as though it is the dilution of the standard. For us to do that, we have to minimize the matrix effect. We used a concentrate standard for assay of concentrated product, and for plasma standards for assays of patients sampled, especially for the congenital-deficient patient plasma sample.

The choice of the reference standard should be based on how well a candidate compares with all the product that it needs to cover. This is a huge challenge for the primary standard, as it needs to cover a product type with wide diverse characteristics.

Even for the plasma-derived material, although they're supposed to be native factor VIII/factor IX molecules, the excipient also will make a difference to the way that it's being assayed. This is something that we have to take into consideration.

Just to give you an example of how it can work, this is a von Willebrand factor concentrate

looking at collagen-binding activity. In this particular set of results, this particular concentrate has been assayed against the fourth international standard for VWF plasma, so this is a concentrate assay against a plasma standard.

Consider that we have two types of collagen reagent, type 1 and type 3, but even within type 3, collagen reagent, you can see we get a wide spread of results. It can be somewhere between 8.5 to about 16 or 17 units per ampoule, and the GCV came out to be 40 percent.

When the same sample is assayed against the first international standard for VWF concentrate, you can see immediately that we harmonized the results we get from all the collagen reagents, and the GCVs came down to about 7 percent.

Assaying like against like, the concentrate against concentrate, improved the interlaboratory agreement. It's also true that when we look at the actual factor VIII activity -- and here's some data where we assayed a concentrate against the plasma standard, you can see that the blue boxes are

1-stage clotting assays and the pink and the yellow boxes are neither the 2-stage clotting assays or chromogenic assays.

It's quite clear that we have assay discrepancy there. When we assayed this concentrate against another concentrate standard, you will find that here, as shown, the histogram outcome shows that they're all coming together; we have good agreement of values.

Even when we're looking at plasma-derived material -- this candidate is a plasma-derived material -- it's still important for us, in accounting [indiscernible], whether you assay against the plasma standard or a concentrate standard. We do have different WHO international standards for the measurement factor VIII and indeed factor IX for plasma and concentrates.

Assay discrepancy is nothing new. The most famous example is the B domain-deleted factor VIII, and we know that their clotting and chromogenic ratio is approximately 1.4 and that clotting activity is higher than the chromogenic activity.

Now, we're moving into the extended half-life factor VIII product, and I don't need to tell this audience how many we have. We have at least 3 extended half-life products for factor IX currently licensed. For some of these materials, they offer better yield, and they're longer acting, so it's better for the patients, but it creates a substantial standardization challenge.

We're now moving also into the gene therapy, and we have seen presentations on factor VIII and factor IX gene therapy. So again, do we expect that that's issued in terms of assay discrepancy? I think we know the answers to that.

The regulators are very concerned over the issues of assay discrepancy, and in 2013, the EMA ran a workshop to discuss the categorization of new clotting factor concentrates. I think that also showed there are issues related to the potency labeling as well as post-infusion sample monitoring.

The professional organizations like ISTH and SCC also came up with recommendations on how to

deal with these new products. This is a very well-cited decision tree, where it's based on statistical assessment of the assay of this new product against the WHO international standard for concentrate.

The idea is if you assay your product against the WHO international standard, you have to decide whether it's valid or not. If it's valid, you then go down one route and, if it's not, you can go down another route. It is based on statistical assessment. So I'd like in the next couple of slides talk about how we do this.

The estimation activity potency; you can use a single-point estimation for tests. To do that, you carry out a multiple dilution for your standard and create a standard curve. You test your test sample at 1 dilution. You can just read off the standard curve and you find out what's the concentration of that test sample.

This is a very common practice in clinical labs, although it is changing, especially in the U.K. The reason why it's a problem is that

single-point estimation for potency can be misleading when the dose-response relationship of the standard test samples are not parallel. It can see that when it's not parallel, in this particular case, the slope of the standard curve is less than the slope of the test curve, so this gives a slope ratio of less than 1.

However, when the test sample perfectly parallels each other, the slope ratio will be equal to 1. We need to do multiple dilution of both tests and standards in order to assess their parallelism.

In an ideal situation, the ratio of slope for standard and test should be 1, and I'm going to illustrate this in the next couple of slides. This is the results from the recent study that NIBSC carried out on the extended half-life factor VIII product.

Here I'm showing the results of the slope ratio, the standard to the test ratio for 15 APTT reagent and 6 different chromogenic assays. The boxes illustrate the 75 percent interquartile

range, and the mean is shown as the black line within the box.

We set out the acceptance criteria for slope ratio as 0.8 to 1.25, and this is represented by the two red dashed lines. This is based on historical data, what we understand from these types of assays that will give us good parallelism.

So we can see that this is a plasma-derived factor VIII against plasma-derived factor VIII concentrate, so this is the comparison best scenario. We only found that only 3 assays gave ratio outside 0.8 to 1.25 acceptance criteria. I think that what is also important to note is the boxes are very small, if you'd like, so that shows there's hardly any variability in terms of slope ratio for all these reagents.

When we look at the same picture for extended half-life product, you can see that, actually, for the majority of the reagent, the means are actually still quite close to 1 for the slope ratio, but the boxes are somewhat wider. And with the 2 reagents here, APTT-S local and the

APTT-automated local, the actual boxes themselves are actually outside the acceptance criteria.

However, out of the 350 assays for APTT assays -- I think there are about 170 chromogenic assays there -- we only have 8 assays that gave a ratio outside the acceptance criteria. This indicates and justifies that this product should be potency label against a factor VIII concentrate, international standard, and labeled in international units because, by statistical analysis assessment, the comparison against the international standard is valid.

However, just because the assays are valid, it doesn't mean that we're going to get the same potency. Here is another pegylated full-length factor VIII product. This is the results from an NIBSC in-house study, and it's quite clear that with APTT-SP and PTT, we're getting real low results. I think there were about 0.4 units per mL. But if you're using Actin-FS, you getting 14 units per mL. So this is a huge assay discrepancy despite the fact that we have

statistically valid assays.

The same kinds of pictures, you can see from a lot of field studies, and I think that all the extended half-life products have a few studies out there now. Just using Afstyla an example, you can see quite clearly that, for panel A, I think this is a sample at 4 percent and panel D is 100 percent.

If you're using a silicon dioxide based activator APTT region, you get a lot lower results. However, overall, I think that the studies have shown and have come to the conclusion that, overall, the results are quite consistent from the chromogenic to 1-stage clotting discrepancy, where overall, for all the range particularly tested, they gave very similar discrepancies, about twofold difference there.

So in the packet insert, this is recommended that for this particular factor VIII, it should be monitored using a chromogenic assay, which reflects the accurate determination of the activity of this particular product, or if you use

a one-stage clotting assay, you should use a conversion factor of 2, so this is quite clear.

However, in this same paper, which is great because it will also show the chromogenic assay to one-stage clotting ratio for 3 other products, for these particular products, you can see that we have some kinds of dilutional linearity issues with the chromogenic to 1-stage ratio, where, for example, with NovoEight and Eloctate there, the increase in the chromogenic to a 1-stage ratio with increasing activity, whereas for Adynovate, it's the other way around.

So we do need to rethink a little bit about these dilutional linearity issues, especially when you're measuring peaks and troughs.

The same kind of story can be seen with the factor IX. Here are field study results, and this time, I think it's with factor IX, Fc fusion protein, which shows quite clearly we have overestimation or over-recovery at low level. Interestingly, the same kinds of results were obtained for BeneFIX, which is the recombinant

product.

For the recombinant longer half-life product, we know that, statistically speaking, they give you valid results, and according to the decision tree, if it's valid by both methods, clotting and chromogenic, you need to look at discrepancy and then agree on a single method.

However, what we haven't talked about is that this discrepancy so far, taking 1 stage to chromogenic discrepancy, but will happen when there's discrepancy within the method. So we know that this is an issue with APTT or 1-stage clotting method.

The next couple of slides are actually on gene therapy, which I'm not going to go through because I think Steve is going to talk about those in much more detail, but enough to say that we see the assay discrepancy for the gene therapy products.

So where are we now? Recombinant and modified recombinant product potency label against international standard, or in-house standard

calibrated against the international standard, using the manufacturer's own in-house assay and reagent. This international unit for these product anchors the relationship between the label potency, dosing, and recovered activity in the patient using these products.

For us, it is really important to keep the continuity of the international unit specific to each product, which after all, has been verified or supported by clinical trial data.

I'm going to run out of time soon, so I'm going to skip this one, but I would like to point out that, again, in the collaborative study that established a 5th international standard, factor IX concentrate, we put in a recombinant factor IX product, and we looked at the results against two other recombinant reference preparations.

Here at the top line, with this particular product assay against a 4th international standard, there's clear clotting and chromogenic assay discrepancy, where the clotting typically was 8.9 IU per mL but 7.1 IU per mL.

But when we assay this particular product against the recombinant preparation A or recombinant preparation D, we minimize the clotting chromogenic discrepancy. It also showed that we have improved interlab agreement. It's also important to note, with this particular set of data, that we obtained the same estimates for this particular recombinant factor IX product relative to all 3 reference preparations used.

So if we have done de-calibration of the standard correctly, it doesn't necessarily mean that we will actually shift into international units by using a recombinant standard. A recombinant factor IX international standard would have minimized assay discrepancy and provide interlaboratory agreement for pooling recombinant factor IX products.

Product specific standard can help solve assay discrepancy. This is actually old data shown by Mikaelsson in 2001. This is a post-infusion sample measured by chromogenic assay and clotting assay. You can see the arrow shows that there's

clear clotting and chromogenic assay discrepancy, when assay gets a plasma standard. However, when the same samples were assayed against the product-specific standard, we get perfectly good agreement between the two different type of method.

I think there's an advantage of having a product-specific standard. It does ensure and fix the traceability of international units as defined by the international standard and allowed interchangeability of the products because we know that currently the similar dose of these different products raises a similar level of activity in the patient.

This standard will also help the long-term stability of the product-specific unit. It will allow method independent testing assay, minimizing assay discrepancy because we will be assaying like against like. It reduces risks related to assay reagent and kit withdrawal, which is a real risk because the kit manufacturer can just drop the reagent when they think that it's no longer appropriate or they feel there is not enough people

using that particular reagent. It can definitely improve interlaboratory agreement.

So I think that a publicly available, stable, product-specific standard calibrated against the IS by manufacturer's method and reagent would support the safety and efficacy of these products.

I'd like to acknowledge our team at IBSC and also Mikhail for a very stimulating helpful discussion always. Thank you for your attention.

(Applause.)

DR. OVANESOV: Thank you very much, Elaine.

I would like to invite to the microphone our second presenter, Dr. Steven Pipe, from the University of Michigan. Dr. Pipe's biography -- and biography of our presenters can be found on the FDA website. But I just want to note that he has served on the board of directors for the Hemostasis and Thrombosis Research Society, as the chair of the board of directors for the American Thrombosis and Hemostasis network, and currently, he is the chair of the Medical and

Scientific Advisory Committee, MSAC, to the National Hemophilia Foundation. Thank you very much, Dr. Pipe.

Presentation - Steven Pipe

DR. PIPE: Thank you, Mikhail, and thank you for the invitation to participate in this great workshop. I'm going to be discussing the clinical laboratory perspective on assays with a particular focus on replacement therapy as well as gene therapy.

Why do we measure factor levels to begin with? They are certainly critical for clinical diagnosis, both diagnosing hemophilia, we assign severity based on the assay readouts and we depend on these assays for highlighting patients who have inhibitors and tracking their progress and treatment for their inhibitor.

We also use these in the clinical management of hemophilia for dose-adjustment, factor replacement, and monitoring factor levels during treatment in prophylaxis and even optimizing factor dosing for PK-guided prophylaxis. But there

are some important principles here, even talking at the diagnostic level, of why we need two types of assays to fully characterize our patients.

The assays are available in almost all healthcare settings. They're the activated partial thromboplastin time rate, TPT. You also have a mixing study that can be used to exclude the presence of the inhibitor. And we have the factor VIII and factor IX activity assays, which are based on this one-stage APTT-based assay. This has allowed accurate diagnosis of hemophilia and accurate disease severity assignment, at least for severe versus non-severe in almost every clinical practice setting.

But we do need additional assays to have full diagnostic precision. We need the chromogenic 2-stage factor VIII activity assay for accurate phenotyping of patients with hemophilia A in particular and to clarify discrepancies that exist between 1-stage and chromogenic assay results.

In some cases, factor VIII and factor IX genotyping is critical to fully understanding

patients' underlying disease mechanism. In some cases, we need factor VIII von Willebrand factor binding assays to sort out distinguishing against other presenting bleeding disorders. And we have even used molecular analysis of the VWF gene to help tease out so we're not misdiagnosing patients who may have type 2 and von Willebrand disease.

This is an often-presented schema of correlation of average annual number of joint bleeds based on a patient's underlying residual factor activity. This is looking at patients comparing severe hemophilia, those with factor VIII activity that is below 1 percent, the precipitous reduction in expected joint bleeding within the moderate range, and then even within the mild range, some continued improvement in risk for joint bleeding, until we get out to around 12 to 15 percent.

But we need to be careful of how much we're extrapolating from this graph. These are all hemophilia patients, all of whom have a mutation, and particularly if we're going to make judgments

about what's happening in this range of the curve, we need to understand that everything we know about this part of the curve comes from patients who have mutant factor VIII molecules, not replacement therapy.

So what do we know about some insights on mild and moderate hemophilia? Well, if we look at a number of mutations that have been described for mild, so basically non-severe hemophilia, we can see that these often are not just affecting the expression and secretion of the protein, and more often these patients have circulating abnormal functioning factor VIII. These are defects in factor VIIIa stability, thrombin activation, their inability to bind to and interact with von Willebrand factor, phospholipid binding, and even defects in factor IX interaction.

Particularly within mild hemophilia, about 20 to 40 percent of our patients exhibit 1-stage 2-stage assay discrepancy, and it can be in both directions, either one stage higher or two stage higher. If we look at those where the factor VIII

activity is higher by the 1-stage assay than the chromogenic, these genetic defects tend to cluster in the factor VIII domain interfaces between the a1, a2, and the a3. And these have been shown to cause reduced stability of the VIIIa heterotrimer and lead to increased a2 dissociation.

Alternatively, when the factor VIII activity is higher by the chromogenic assay than the 1-stage assay, these genetic defects tend to be clustered around thrombin cleavage sites and the factor IXa binding sites. So these are thought to cause impaired factor VIII activation by thrombin or an impaired binding of factor VIII to factor IXa.

If we think about these altered functions of these mutant molecules, it would be hard to suggest that these are only relevant in in vitro assays and couldn't also be contributing to the clinical phenotype expression of these patients' diseases. So extrapolating from mild and moderate hemophilia on the clinical characteristics, their bleeding rates, et cetera, purely based on a

factor VIII assay, without taking into account this aspect of the function of the molecule, I think is a potential mistake.

So now, let's shift to the other main arena, which is in clinical management of hemophilia A. This was demonstrated years ago as the principle for modern prophylactic therapy. It was a suggestion that we've talked about the peaks and troughs today of traditional factor VIII replacement therapy, but it seems that the time spent with factor VIII trough levels below 1 percent is directly correlated with bleeding risk. And the more hours per week you spend at those low levels, the reduced likelihood that you will remain bleed free.

But this is not an absolute threshold. This continuum exists whether you said time spent below 1 percent, time spent below 3 percent, or perhaps even time spent below even 30 percent. There is still some degree of correlation here with increased risk of bleeding.

If you look at the typical prophylactic

pattern of replacement therapy, this is demonstrating the peak and what we call the trough, and then with the next dose, you achieve the next peak.

Where you assign that critical level for what you consider optimal prophylaxis in a patient has a lot of interindividual variability. And we've learned years ago that programmatic prophylaxis may be able to deal with the majority of patients, but there's going to be outliers who need higher trough levels to maintain a good bleed control.

The advent of the extended half-life models does change characteristics of the curve overall, but we still have the principle of peaks and troughs. And although we can extend the area onto the curve if we really push the limits of the interval between dosing, patients can spend inordinate amounts of time with quite low factor levels towards the trough.

To counteract that, what has been used in the era of extended half-life is to even maintain

the same interval with the standard half-life, in which case re-dosing is occurring before patients get anywhere near these critical thresholds, and for individual patients, this has been important to gain real good control of their bleeds.

This was from a secondary analysis from a study in which all patients had their prophylaxis optimize. So they were all dosing at a fixed interval of every 3 days, and all of their individual pharmacokinetics was known so that the optimal dose could be given at a 3-day interval, such that their factor levels would never drop below 1 percent before their next dose.

So because we knew the factor level at any given time of the day, we could correlate that with the timing of their bleeds and make some assessment of what were some critical thresholds for breakthrough bleeding.

What this is showing is the continuum as far as predicted maximum factor VIII activity level at the time of bleed and the proportion of those who were without any spontaneous joint bleeding.

Some of the targets that could be identified is a target of 5 percent factor VIII trough level would have led to about 71 percent of patients achieving zero spontaneous joint bleeds. But approximately 15 percent of the patients would have required a factor VIII level well above 15 percent to have no spontaneous joint bleeds.

So again, even within this cohort, optimizing for their individual pharmacokinetics, we still see interindividual variability on the risk of them having breakthrough bleeding.

Assay discrepancies in clinical monitoring can also depend on the factor replacement product. Elaine has presented to us nicely here about issues reduction standardizing the products that we actually infuse into the patients, but even after infusion, there remain issues.

Discrepancies between 1 stage and chromogenic assays have been reported. Discrepancies may be exacerbated by B-domain deletion and sometimes maybe even the length of the B-domain linker. And some high discrepancies have

been reported with some pegylated B-domain-deleted recombinant factor VIII, and some of this may also be influenced by the reagents that are chosen for assaying that particular product. Nicely, these discrepancies can be overcome by using product-specific standards as Elaine has shown us.

One emphasis I would like to make is we have been dealing with assay discrepancy for a very long time. If we think about some of the challenges in the recombinant era, just the biochemical characterization of these products have shown that they may have altered post-translational modifications by glycosylation, phosphorylation, sulfation.

There may be presence of dysfunctional proteins that either have reduced through absent activity, reduced through absent binding to protein partners, the assay discrepancies that have been mentioned, and even discrepancies in the vial content versus the labeled potency.

On the clinical side, we've had to deal with altered pharmacokinetic parameters. Some

products have shown reduced recovery, shorter half-life, changes in the volume of distribution, dealing with clinical reports of reduced efficacy. When a patient has been on a particular product for their whole life, and they start on a new product, then they come back to the clinician and say, "I just don't feel that this is working the same as my previous product," even though the factor assays would give no insight as to why that would be the case.

Reports of increased inhibitor risk; this has been demonstrated from retrospective studies all the way through randomized controlled trials and some sense that there may be reduced efficacy in some immune tolerance induction applications.

But it only gets worse. That was with the so-called facsimile recombinant products, where we are trying to mimic the endogenous proteins, but this is the bioengineering strategies for enhanced biologics that are now being applied for modern-day replacement therapy. So as we make more and more bioengineering changes in these molecules, we're

apartment to see even more differential between these products.

This is emphasizing the same principles that Elaine's already shown you, but if you're using a particular EHL recombinant factor VIII, you increase the accuracy by 1-stage clotting assay when a product-specific standard is used. Here you can see that drift that Elaine showed us, as well as the wide spread across a range of concentrations, but this all collapses down with a product-specific standard. This can also be seen with the same product against a chromogenic assay, again, with this drift and the widespread in the assays, but then collapses down with a product-specific standard.

We shouldn't take from this that one particular assay is more accurate or reliable than the other because if you look at the variability in these assays, even when labs are using both the 1-stage and the chromogenic assay, we see really that we're seeing the similar types of variability within these assays, even if you were using a

product-specific standard. So we still have the same issues with both of these assays.

Now, if we look across the eras of treatment for hemophilia, we've been having to increasingly deal with these bio-engineered molecules, both in standard half-life and extended half-life products. But it's not going to end there because now when we move on to gene therapy, we are also having to deal with bio-engineered molecules.

We've talked about the point mutation of the factor IX Padua. B-domain-deleted factor VIII is the primary construct in gene therapy, but it's not the same, which I will show you in a minute. We've added codon optimization to these transgenes, and there's probably more targeted mutagenesis to come in subsequent upcoming gene therapy protocols.

So what's at issue with codon optimization? So in codon optimization, we're replacing rare codons. Because of the redundancy of the human genetic code, you can replace rare codons with frequently used ones to attempt to increase the protein expression. Because of the redundancy,

you're not changing the amino acid sequence of the molecule.

This has already been used for at least one commercial extended half-life recombinant factor IX in their production cell line, but it's a main stay now of factor VIII and factor IX transgenes for gene therapy. Adding codon optimization to factor VIIIb domain deletion substantially increases protein expression and allows you to either reduce the vector dosage or achieve higher plasma levels.

But there may be some unanticipated effects of codon optimization; altered protein confirmation and stability, altered post-translational modifications, and perhaps even altered protein function in a number of different areas.

The proposed mechanism is that codon usage determines the translation rhythm, so causing ribosomes to slow down or pause at specific sites. This can modulate the sequential folding events that occur co-translationally.

The thought actually is that codon usage

acts as a secondary code, so not just the codon determining the protein structure itself, but this secondary code because of these ribosomal regulations, these pauses, can actually guide in vivo protein folding.

How do we know that this really happens? Well, we can gain some insight from some work that was done by scientists here at the FDA, looking at a single synonymous mutation in factor IX that disrupts protein properties.

So here, this patient has a single nucleotide change, which does not change the coated amino acid for factor IX. Yet, because of this alteration, and this leads to altered messenger RNA, secondary structure, and codon usage. It alters the kinetics of translation, alters the protein confirmation and post-processing, can lead to enhanced protein degradation, and results in reduced protein expression and expression. This is the root cause of mild hemophilia in this particular case patient.

So we're talking about codon optimization

of factor VIII. The preclinical studies had actually predicted this. Codon optimizing B-domain-deleted factor VIII exhibited 7-fold higher expression from CHO cells, but there were some observed differences in post-translational modifications and in O-linked glycosylation, the degree of tyrosine 1680 sulfation.

Curiously, the specific activity was 1 and a half-fold higher by 1-stage clotting assay compared to chromogenic. This was not predicted from what we knew about B-domain-deleted factor VIII and other settings. As Elaine had introduced, this came to show up in the clinical gene therapy with these codon-optimized B-domain-deleted factor VIII as well, where we see about a 1.6 ratio comparing the 1-stage to the chromogenic.

Is this going to be an issue for factor IX? Well, actually, we're learning that it is. This was just presented at the ASH meeting from one of the factor IX trials. This is showing across the bottom here is chromogenic factor IX, then these 4 reagents that are chosen here represent about

90 percent of the testing that would be done in clinical laboratories across the U.S.

These are individual patients in the colored lines, but you can see that they're all showing this same degree of variability depending on what reagent is used. So depending on what the central lab is using, you're going to see very different results from the local lab and as compared to the chromogenic.

If you look at spiking the Padua variance specifically into a factor IX-deficient plasma, again, we see variability across these assays and clearly different from the chromogenic. This also exists for BeneFIX, but it doesn't show the exact same pattern as we're seeing with the Padua variant.

So assay selection is going to influence the readout of factor IX activity in these gene therapy clinical trials. We are lacking clinical correlates with the factor IX chromogenic activity. Elaine mentioned that we don't have an approved chromogenic factor IX in the U.S., which means none

of our clinicians have established any sort of correlation clinically with a chromogenic readout from a factor IX assay.

Endogenous expression of the transgene product also introduces potential for interindividual variation. If you're talking about standardizing a product, where you're controlling the cell line and the transgene that goes into that, you can get a fairly uniform product. But we are taking these transgenes, and we're putting them in individual livers, if you like, individual manufacturing facilities.

What kind of standardization can we do when every transgene that's expressed is coming from a separate patient?

In summary, measuring factor VIII levels is absolutely necessary for accurate diagnosis and phenotyping of hemophilia A as well as monitoring during treatment, but both 1-stage and chromogenic assays should be used for diagnosis and phenotyping.

Product-specific standards can overcome the

discrepancies we see in clinical monitoring, but correlation of factor levels with clinical outcomes is really context specific. Caution should really be exercised when extrapolating from one clinical context to another.

Mild hemophilia is not equal to replacement therapy and replacement therapy may not be equal to gene therapy. We have already highlighted here the issue of comparing patients who have mutant factor VIII molecules to those that are getting native molecules, and also the peaks and troughs of replacement therapy may be difficult to compare to the steady-state levels that are being achieved with gene therapy.

So hopefully, this will stimulate some conversation for our panel coming up. Thank you.

(Applause.)

Panel Discussion

DR. OVANESOV: Thank you very much, Dr. Pipe, for your presentation.

Now, I would like to direct your attention to our panelists, who will help us discuss the role

of factor activity as discrepancies in clinical trials.

We have several new people on the panel who were not introduced yet, and I'm going to go ahead and let you introduce yourself if you don't mind.

DR. FRIEDMAN: Good afternoon. My name is Ken Friedman. I'm the director of the Hemostasis Reference Laboratory at Blood Center of Wisconsin, which is now part of a group of blood centers called Versiti. I direct that lab, and I also am involved in hemophilia care of mostly adult patients, but also some pediatric patients. I've been involved in some of the monitoring of the clinical trials.

DR. DODT: Good afternoon. My name is Johannes Dodt. I'm from the Paul-Ehrlich-Institut in Germany, and we are a national authority for licensing blood products. I am involved in the quality aspects of these products, and we are also doing the licensing of the recombinant analogs. Thank you.

DR. MARLAR: I'm Richard Marlar, professor

at the University of New Mexico. I'm also the director of the Coagulation Laboratories at TriCore, which is a reference lab for about 16 hospitals in the state of New Mexico, as well as doing the special coag for the hemophilia program.

DR. OVANESOV: Thank you very much.

Let me introduce an overview of a very packed agenda for the discussion today. The discussion will be facilitated by three groups of questions; the first group about the clinical lab practice, the second one is factor assay discrepancies, and the third one, surrogate endpoints.

Now, without further ado, I will let our panelists respond to the first question. Is it practical for clinical laboratories to carry different factor activity assays for hemophilia patients on different therapies?

DR. MARLAR: From my perspective, I think that we need to look at laboratories in different ways. There are different types of laboratories. There's the large reference laboratories that see

many samples. They don't know what's in the sample. They get in and get a request to do a factor VIII or a factor IX. There are hospital-based laboratories that may or may not work with an HTC. And then finally, there are smaller the HTC-specific laboratories.

So I think, from that perspective, we have different ideas of what's needed. I don't think, in the majority of laboratories in the U.S., that we can handle more than 2 factor VIII or 2 factor IX assays at a time on that. And I think it really depends on how technological and innovative the director and the technical staff is to be able to set those assays up in there.

DR. FRIEDMAN: I'm going to also chime in on this question about using different reagents. Most laboratories actually have automation that they have validated, and that automation is actually sold in conjunction with specific reagents; that is, by the same automation manufacturer.

As a result, if you ask a laboratory can

you put on a different reagent set, then you're actually not mirroring the reagent set to the manufacturer. And in most cases, that would mean that you're ending up asking the laboratory to make a laboratory-developed assay because it's not necessarily what will be validated in the licensure of that payer.

For this situation, what ends up happening is that you have to then validate this assay with all the things that are expected of validation, including accuracy precision, lower limit of detection, et cetera, and that becomes quite an issue. That's in part why many clinical laboratories stay with one manufacturer, which may be contracted by their institution such that they don't even have the flexibility to choose which reagents they're necessarily going to use.

Then finally, the last point which I'll say is that if the reason to have that is in order to be able to accommodate different factor products that the patient's on, then you need excellent communication between the clinicians taking care of

the patient and the laboratory so that they can choose the right assay. Then the laboratory has to report the right assay the right way, and then it needs to go into the hospital electronic medical record in such a way that it's traceable.

All those connections, which are somewhat outside the laboratory but communicating between the laboratory and the clinicians and the patients, are all problematic connections.

DR. PIPE: I would also say that, at our laboratory, the precedent has already been set in other therapeutic areas, particularly in anticoagulants, where we have a product-specific anti-Xa assay for essentially every one of the anticoagulants that are used.

We had to set up all of those product-specific standards, and we demand, when those samples come to the laboratory, that the clinicians identify the product that the patient is on, and if it's not apparent on the order, our lab staff actually do the next step to make that determination.

There was motivation to have those internally and to be able to accurately report that out, and we were able to accommodate that. And we do way more anti-Xa testing than we would do for hemophilia applications. So as far as the practicality's concerned, I'm not sure that's the limitation. I think it's the internal motivation and the ability of clinicians to influence their individual labs to make this happen.

DR. MARLAR: I can understand that, and we have the same thing for the Doax [ph] as well, but it's the absolute communication because we will report out a wrong answer. If we don't get that, we have to spend time, which is money in our laboratory, to look into the medical record to find out what's going on. And if it's somebody outside of our hospital system, we have no idea, and that could possibly be the same way.

DR. OVANESOV: So one way to go around the need to introduce a brand new assay is to use a product-specific standard to pre-qualify or calibrate routinely used assays. Is it practical

for the labs in the United States to use product-specific reference standards similar to previously available ReFacto standard?

DR. DODT: Thank you. Before we start the discussion on the product-specific standards, I'd like to mention an important point. All products have been licensed based on an assay, which was the best assay for that product at the time of licensing. It is well described in the licensing dossier, and it is up to the companies to provide the users with information, which are the tests to be used and which are not suitable for that product.

So thinking about the comment from Kenneth, it is the interaction between the medical doctor and the lab to choose a test, and that, as I said, is a problem. So how can a product-specific reference standard be better communicated to a lab than the best method? What is your opinion on that?

DR. FRIEDMAN: Well, I guess my opinion on that is that when there were very few products, it

was a little bit easier to do. The laboratory that I direct, and probably other laboratories also, set up the ReFacto standard. And when we set up the ReFacto standard, we actually had the order set such that people were ordering a ReFacto factor VIII. And since the order was specific, then we knew what to do.

However, when I look at product-specific standards, theoretically, it sounds wonderful. The problem with product-specific standards are multiple; one, if you have a product-specific standard, you still have to validate the assay, and it's now, by almost definition, a laboratory-developed assay for that specific product standard.

In addition, you need to have materials to actually perform tests of accuracy. So you need, actually, materials that are provided by the manufacturer or by buying the actual products in order to calibrate your assay. You also need to participate in external quality assessments, and are there samples to actually do with a product-specific external quality assessment sample; so

there's that issue as well.

So you can see how the number of issues that come up with product-specific standards multiplies as the number of materials that come out there multiply. So I think, ideally, it sounds wonderful.

The last thing, which I'll say, is that if you have one patient who's on product X, but all the rest of the patients are on product Y, then you set up your assay for product Y. And then when the patient comes in on product X, that becomes a very expensive assay to run as a onesie for that one particular patient. So there are many logistic complications, is what I would say.

DR. MARLAR: One other point is that when you have patients on multiple products, which we've already had on two occasions, how do I measure that on two separate products, especially if they don't look like plasma or factor VIII? So that's another issue.

DR. PIPE: Richard, you brought up a point, because what we didn't really talk about in the

formal presentations is, for the first time, we are mixing therapeutic agents, both of which affect clinical assays. And if we talk about emicizumab being used for the routine prophylaxis, and then on top of that, they come in for acute surgery or need breakthrough bleeding management, if the clinician wants to monitor that patient, this adds a whole new complexity that wasn't anticipated.

DR. OVANESOV: I think that brings us nicely to our third question. What do hemophilia-treating clinicians want to achieve with factor activity testing? There are different scenarios, obviously.

DR. PIPE: I guess I tried to highlight a few of these. I think in prophylaxis, you certainly can get away without having to do routine monitoring. Some patients sort of find their sweet spot of dosing and interval based on the clinical feedback. But maybe getting back to Marilyn's point at the very beginning, you would hate to have to use the trial-and-error approach early on in life with a young pediatric patient, and have to

have bleeds be the readout for whether you've optimized their prophylaxis.

So I think the utility of having access to monitoring and then maybe application of population PK models seems to be a popular management issue. But definitely being able to understand why a patient is having a breakthrough bleed and monitoring for surgery, these have all been proven to be critical areas where, if you tell the clinician that they will not have access to those monitoring tools, they become quite anxious, actually.

DR. OVANESOV: Thank you.

DR. FRIEDMAN: Can I just go back to one -- I feel like I've been the naysayer about everything, and I'm sorry to do that, but I actually want to also make one potential suggestion related to the last question, which is that the labs that participate in the field studies get an idea of how their particular reagent responds to a particular engineered product. And the availability of testing those things going forward

after something is licensed is something that -- when you talk about postmarketing-type things, I don't know if FDA would consider that postmarketing issue.

DR. OVANESOV: We did consider that in some situations. It's obviously risk based. In some cases, we've worked with the company, and the company proposed to maintain a hotline that clinical labs can call, and they will be guided through the difficulties within assay standardization and calibration, and in some cases, the company might provide the material that is representative of the product.

This is not something that is done consistently, meaning that we don't require every company to have that, but some companies opted to have that in place.

But to put things into perspective, we have 18 licensed BLAs for factor VIII products and 9 factor IX BLAs. Not every product would require product-specific standards, but if you add a couple of gene therapy products to the creation, it's

going to be very challenging to have.

DR. MARLAR: Yes, I agree with you on that, that it is going to be challenging. And I also think that the laboratory community needs to have some information that's out there for every laboratory to assess, to know that, well, this product, you need to do this with, and this product, you need to do that with, that's available for everybody, rather than having to go through every product insert and through the original data to get that out. I mean, a summary of what's there is something that should be used or available.

DR. OVANESOV: Thank you.

DR. GRAY: I think that is important to remember that when we talk about a product-specific standard, the usage can still be discussed because you don't have to -- your lab, if you want to have a look to see exactly how your own assays behave, that's where the product-specific standard would be useful.

At the moment, the way I see it being used in the clinical lab is, really, for the clinical

labs to understand how their reagent behaves, and I think that's important. So you don't have to use it in every single assay. I think it's understanding the characteristic that's important.

But I think with the product-specific standard, it's also important from a manufacturing point of view because if we have something that's stable, we know that it's there. It pins down the unitage that's related to that product.

As someone who makes an international standard, when I replace a standard, I worried because those standards, the products right now are so intrinsically linked to the international standards that have been calibrated against using a specific set of reagents. If I decided not to make SynthASil anymore, what would happen?

So I think we have to think about it from several different angles about the usage of a product-specific standard.

DR. OVANESOV: Thank you. I think it's time for us to move closer to the surrogate endpoints, and I will read these two questions.

What would you consider a clinically meaningful assay discrepancy, and what are the safety risks that can arise from factor assay discrepancies to patients on replacements or gene therapies?

DR. PIPE: I think right before this, you had what degree of variability do we have even within even the individual assay; is that correct?

DR. OVANESOV: That's right.

DR. PIPE: I think they're both related. We're already starting with a variability that could be as high as, certainly, 5 to 10 percent, but maybe also pushing above 10 percent for some assays for variability. And then now you're laying on top of that a discrepancy, where there could actually -- you're overlapping with those interassay variabilities.

So as far as what's clinically meaningful, I don't actually believe that that's been sorted out even with the original discrepancies that I pointed out. We do not know -- even from the mild hemophilia patients with the 1-stage/2-stage

discrepancies, it's not clear that you could define one assay for those patients and say that is the truth. It's just an observation that illuminates a molecular mechanism that's a problem in that particular molecule.

I think you could say the same thing with the 1-stage/2-stage discrepancies with the clinical management with replacement therapy. We identified this problem in our laboratories, but we haven't done sufficient work to be able to say that one particular readout of those assays is truth as far as representing a clinical outcome. And I don't think we're any further ahead today than we were probably 25 years ago, when this first became an issue in replacement therapy.

So to answer your question, I would say I don't know how we could know that information at this point.

DR. OVANESOV: Thank you. It makes a lot of sense to me, but we need to get closer to gene therapies. So if a discrepancy is found, how do we pick the assay and threshold to measure factor

activity as a surrogate marker? And remember, we use a surrogate marker for the accelerated approval pathway, but are going to approve a product, hypothetical product, on the basis of the presence of a certain level of factor activity in blood of gene therapy patients.

Is it even valid to use this approach, given all the issues with assay discrepancies, with clinical lab issues, and what just Steve said, that we don't know what we're measuring, basically. Well, we know what we measure; we measure factor activity, but how does it relate to normal activity? That we don't know for sure.

DR. PIPE: I guess what I would say is, the continuum is always going to remain true. More factor activity is always apt to be better than less. So I don't think we can discount that the factor activity is absolutely useful and has proven to be a valid surrogate marker for decades, from diagnostics to replacement therapy, and now will also prove true in the gene therapy era.

Where we're maybe running into issues is

when those assays are straddling key decision treatment triggers. So if you're doing prophylaxis and you're running someone close to the wire with a trough of 1 percent, you're really putting a lot of stock in the ability for your lab to actually measure that 1 percent and to be making laboratory adjustments accordingly.

If your gene therapy outcome, on one assay, your median is, say, 7 percent, but on the chromogenic, those patients' median is down around 3 or 4 percent, that's putting clinicians at an awkward interface because they would make maybe clinical assessments of outcome differently based on where that straddle occurs.

But the further we move up the continuum, these discrepancies become less and less relevant to us clinically. It's hard to imagine, from anything that Marilyn showed us today, that we would really be making a different clinical decision for a patient who sits at 40 percent versus a patient who sits at 27 percent. I just can't imagine how I would manage that patient much

differently with that kind of a differential.

So is it clinically meaningful at that level? I would say no. But at the low end, it definitely could be. So related to these gene therapy trials as a surrogate marker, I guess it really does depend on where they are on that continuum.

DR. OVANESOV: I think I can refer back to the discussion we've had on the instruments on whether we have evidence to say that the difference that was measured by a particular, say, quality-of-life measure is meaningful.

We actually have the same problem here. I understand that a 20 percent increase or an increase in 20 percent of factor activity may seem meaningful, but where is the evidence that supports this statement?

Maybe there is evidence, and that's actually the question that is represented here, and we described that in our guidance for gene therapy and hemophilia. But the issue is the kind of evidence that is available to us to say that this

is enough.

I think Steve already responded to this question; does factor activity level post-gene therapy have equivalent meaning to prior levels achieved with exogenous factors? Probably, not always.

So considering the discrepancies between assays and reagents, can we predict the correlation of factor activity and bleeding in a particular case? In general, yes, we can agree more factor is better, but when we are presented with a particular gene therapy, how do we predict that correlation? What kind of evidence would we need from the company?

Maybe the companies can respond if they want.

DR. PIPE: I guess I'm somewhat fixated on the fact that we have had traditional clinical decision-making triggers that are benchmarked against certain thresholds of factor activity. But once we get anything above 10, 15 percent, we're on very shaky ground as far as being able to

distinguish clinically meaningful differences across patients.

So when you say, can we predict the correlation of factor activity bleed, I think there's plenty of evidence that after you cross a certain threshold, spontaneous joint bleeding stops, traumatic bleeding becomes much, much less frequent, and at some point, clinicians will probably even choose not to recommend additional hemostatic replacement therapy or even coverage for surgery based on a particular factor level.

So as long as critical thresholds are surpassed, it may not be important to be able to make a clear predictor between these. So your 20 percent example is sort of an interesting one because I think anybody looking after hemophilia would say you would not expect spontaneous bleeding at that level. Almost all traumatic bleeds would probably be prevented for the most part. And you're probably talking limited to need for replacement therapy with certain types of major surgery.

So I would be hard-pressed to demand that there be a clinical correlate with that 20 percent activity. As a secondary outcome, almost certainly, it would be obtained in the course of the trial, but there would be no reason to doubt the utility of that 20 percent in that patient.

DR. OVANESOV: Thank you very much.

DR. MARLAR: Steve, I have just a question to follow up on that. Do you think that the products are going to have a different 20 percent level when you start working with that?

DR. PIPE: If there's an alteration of the biology of the molecule, codon optimization, which is a hypothesis at this point, of course, or Padua, where actually there's clearly an alteration of the biology, there may be not reagent issues that are at the root of that, but actually, the biology of how the molecule gets activated and how it initiates in early components, for instance, in a 1-stage assay.

So what you're going to be challenged by there is that may not only be an in vitro

manifestation of the biology of that molecule. That benefit, if you like, that altered activity advantage of that molecule could also be relevant in vivo.

So 20 percent at a non-bioengineered VIII versus bioengineered, the activity is still the activity. And if it's an alteration beneficially, if you want to call it that, for the molecule, you would think that that would probably be represented clinically. And I don't know how you would tease that out in the levels that we're talking about here.

When we were down at, say, 1 to 5 percent, these would have been absolutely critical ideas to try to wrap our minds around, but as soon as we get across some critical threshold levels, I think this becomes kind of noise.

DR. GRAY: But then the problem becomes that your assay discrepancy, say within 2 APTT reagent, could be 40-fold difference, so --

DR. PIPE: Did you say 40?

DR. GRAY: -- yes, which happens with one

of the, say, pegylated factor VIII molecules.

So if you really chose -- well, I don't know what is right and what is wrong, but it tells me that the reagent that gave lower activity, obviously, is not quite right in some way.

So I think it goes back to the point that it's very important that the information for these products, about how these products are potency labeled, the assay that's being used should be information that should be accessible because in those types of situations, you really don't want people to use a certain reagent and then think the company should come straight out and say you shouldn't be using those reagents.

DR. PIPE: To that exact point, I think from Mikhail's example in gene therapy, knowing what the distribution of those factor VIII levels or IX levels are across a variety of different reagents should be a critical part of the learning from these trials because, then, that information is available to the clinicians.

It won't be 40-fold for any of the gene

therapy of course, but it could be 1.5-fold differential, maybe even up to 2-fold differential if you want to talk chromogenic and certain specific 1-stage. But I think, as long as the clinicians know that and they know what that differential is, I think we would all be comfortable in the day-to-day management of these patients.

DR. GRAY: I think that it may also help for the gene therapy product if the in vitro produced expressed protein. If you do a characterization of that with a different reagent and follow up looking at patient sample from that gene therapy to see whether they follow the same pattern or not in terms of the reagent characteristic, I think that would be helpful to help us understand a little bit more whether you can predict what reagent you should be avoiding.

DR. OVANESOV: Thank you very much for this excellent discussion. We ran over our time, and I want to thank our panelists for their time they spent with us today. Thank you.

(Applause.)

DR. LOZIER: Let's go ahead and take our recess.

(Whereupon, at 2:38 p.m., a recess was taken.)

Session 5

Moderator - Jay Lozier

DR. LOZIER: I am a medical officer in the Center for Biologics and Review, among other things, gene therapy and various factor concentrates.

In this session, we're going to talk about clinical trial design, and we'll be talking about a couple of things that are of particular importance to us, one of which is when do we move from adults to kids, however carefully, and we'll have a couple of presentations addressing that. And then we need to address some of the issues about long-term surveillance and focus on a particular risk that's been identified in the preclinical animal models.

The first question up here for your Slido polling is at what age is the human liver

essentially an "adult" organ? And your options are 13 to 14 years, 15 to 16, 17 to 18, and 10 to 12. I notice about 30, 35 people have been responding to the morning session, so I hope you won't slack off and we'll get a good response on this. And there's no right answer, I don't think.

(Audience responds.)

DR. LOZIER: Why don't we go ahead and close this down? It looks like there's a sliding scale here around 13 to 14 years.

Let's go to our second thought-provoking question. How long should factor VIII or factor IX levels be demonstrated to be stable in adults before treating adolescents with gene therapy; that is, what sort of a track record do you want to see with adults before you move to children, whether they're older adolescents? Let's just assume that and not young children.

This one gets a little more activity a little quicker.

(Audience responds.)

DR. LOZIER: We'll give that a pause. It

looks like quite a few people are wanting at least 1 year and some are wanting 5 years. Let's put the questions down now, and I'll go ahead and introduce our first speaker.

Amy Shapiro is the founding member and the medical director and CEO of the Indiana Hemophilia and Thrombosis Center in Indianapolis and has been a leader in hemophilia treatment for many years. She's also an adjunct professor of pediatrics at Michigan, where she, I think, administers a coagulation fellowship with Steve Pipe. She is going to talk to us today about the duration of the gene therapy response.

Amy?

Presentation - Amy Shapiro

DR. SHAPIRO: Thank you very much for inviting me today. Dr. Lozier asked me some very difficult questions. Here are my disclosures. The questions that Jay posed to me include this set of 4 basic questions: how long data would be required in adults for duration of response before trials in children could be initiated; the duration of the

vector-sustained expression in children, would it be different in children compared to adults, what would we want to achieve; and can we define the target factor level in children based upon the age treated to achieve a reasonable level as an adult, so what would be the end target level and what would you be starting with based upon the age that the child received that therapy; and how do we proceed in children? Do we consider age cohorts?

In order to approach those questions, I broke this down into a few areas, the data on the duration of response that we have so far; specific pediatric concerns, including the age of the patient, pulling out what I might call special populations, where the risk-benefit ratio for specific therapies could be considered slightly different than the general pediatric population; and then those unknown issues, the things that we don't really have enough information about at this time, and do we need further information as we approach pediatrics; and then looking at the overall risk versus the current burden of therapy

based upon what do we have available at this point in time to treat patients.

In terms of duration of response, the most information we have at this point in time is regarding factor IX gene therapy. The St. Jude Children's Research Hospital and University of College of London project was originally published approximately 8 years ago and stills shows continued sustained factor IX activity in the 3 to 5 percent range. It was present in a dose-dependent manner with no long-term safety issues for the duration of follow-up at this time.

Subsequent trials by Spark and other companies have used factor IX Padua and have achieved higher factor IX levels of approximately 30 percent with lower vector doses, with a follow-up that's shorter since that is a newer innovation, lasting approximately 2 to 3 years.

Then newer trials, including one recently discussed at ASH by Dr. Nathawani, looking at a different vector achieving levels of approximately 90 percent with a Padua variant, and then other

modalities where we don't even have clinical data as yet; for example including gene insertion in the safe albumin harbor of the albumin gene from Sangamo, and here we don't even have any data on the levels achieved or the response duration.

So we have quite a big range in terms of what we have available and how gene therapy is moving forward.

For factor VIII, the most mature data we have is from BioMarin. This used, in the original study, 2 dose cohorts. There was not a linear dose response. The higher dose cohort, which consisted of 7 patients, achieved levels that varied between 19 percent to 164 percent.

Interestingly, in this study, there didn't appear to be a clear connection between the elevated ALTs and the anti-capsid T-cell response, and then the steroid use in factor VIII activity to ameliorate the elevated liver enzymes. Four of the 7 patients with steroids did not halt the increase in ALT, and the question is then raised, is this an immune response versus actual hepatotoxicity? Is

there a difference?

This is the data that was published from the BioMarin trial looking at the high-dose cohort with 7 patients. For the first 52 weeks, there is further data that is now available, but not yet published in a manuscript. The lines show the median levels, the little areas -- these are the mean levels, and this is between the 25th and 75th percentile. But you can see that the majority of these patients are within the normal range, although as I said before, there was quite a bit of variability in the levels that were achieved within the same dose cohort.

We have quite a bit of information in terms of development of this technology, including the AAV as a vector capsid and lots of different things that have been performed over the years in order to try to achieve where we are today and the success that we have achieved. As you can see, we have a lot more data with factor IX gene therapy as we do with factor VIII at this point in time.

What are our concerns in pediatrics? Well,

age is an important concern. If we have an episomal vector, it's going to be diluted as time goes on with liver growth, so what level you initially require to achieve is going to have to be targeted at a different level to achieve an adult liver size and as the vector dilutes over time.

So we have to think about what we want to achieve as an adult and then work backwards in terms of what we need to achieve based upon the age of the child that we treat.

The answer may be different for factor IX deficiency as compared to factor VIII. Consistent levels of factor IX of about 30 percent are likely better than anything right now that we can achieve with current available therapy, and consistent factor VIII levels of 40 percent are likely better than anything we are likely to achieve right now with current therapies with factor VIII, including novel therapies.

In terms of durability of response, it's clear that as you transvect to youngest patients, we're going to want the longest durability of

response. So if you treat someone who's 50 years old, they have a shorter life expectancy in terms of what you want to achieve in terms of durability response as compared to treating someone who is 10 years old, where you want a much longer durability of response.

Pediatric patients represent a vulnerable population in terms of participation in clinical trials and consent, so we have to be very careful as we approach this population because the parents are essentially consenting for these young patients.

Safety data and long-term durability are required if other reasonable therapies are available, so we really have to think about what's the burden of disease and what is reasonable to treat our patients with, when we take risks with young patients.

We also have to think about data about overcoming development of neutralizing antibodies, If a second vector infusion is required later in life, if durability of response is not what we want

for a lifetime, how are we going to overcome that, and we need to think about that and plan for that as we approach children.

We might pull out what I would call a special population of children. I'm using this as an example and not saying that this would represent the special population of children, but patients with inhibitors are clearly more vulnerable patients, as we've heard before. Gene therapy could provide the ability to tolerize these patients without costly, burdensome infusion therapy, and they may represent, therefore, younger candidates for gene therapy due to the burden of care and the sequelae experience.

With the advent of emicizumab to at least control bleeding in factor VIII inhibitor patients, it does not tolerize them, but at least we get better bleed control. This is not available at this point in time for factor IX inhibitor patients, which are far more difficult to tolerize and difficult to treat. So you might even categorize a factor IX inhibitor patient different

than you would categorize a factor VIII inhibitor patient in terms of risk and risk-benefit ratio.

Then there are a whole group of what we would consider to be those unknown issues. Does the cell line for vector manufacturer result in different pathophysiology of the elevated liver enzymes that we see in patients post-infusion? Some of these cell vectors are produced in mammalian cell lines and some in insect cell lines. And does, perhaps, one create a cellular immune response versus the other actual hepatotoxicity?

The seroprevalence of immunity to AAV serotype is likely based on age, so that if you got a younger patient population, you might have eligible a larger number of patients for this therapy, so you have to try to figure out what's your optimal age to reach the most eligible patients while considering and balancing the risks at that point in time.

Then you have to consider about overcoming immunity to AAV serotype positivity, whether it exists in the patient before due to some natural

exposure or whether the individual has been exposed to that vector in the past and would require retreatment later in life.

Longer-term outcomes and unanticipated events need to be thought about. Apoptosis of transduced cells due to protein overload and loss of efficacy over time can occur, so some degree of prolonged observation with some of these particular technologies should be considered; a potential for malignant transformation later in life; for example, hepatocellular carcinoma. This may depend upon the age at which the patient was treated. It may depend upon their prior viral exposure. It may depend upon their stage of liver development or insertion of the vector, even if it's episomal off site or off target.

We need to think about this and know how to monitor our patients who undergo this therapy: how often do we see them and what's the optimal tool for monitoring them for long-term sequelae related to unanticipated events?

If you think about patients from birth to

adulthood, we think about their growth of their liver over a certain period of time, at which point we can consider their liver to be near mature in size. We think about the prevalence of the vector serotype, which can be perhaps very low at birth and then increase with increasing age. And then we think about, for example, special populations, including inhibitor populations.

So when do we pick the best opportunity to increase eligible patients in terms of seroprevalence of a vector; in terms of the optimal level when the adult liver size can be near achieved and you don't worry about dilution of the vector; and when we call out specific patient populations that we think the risk-benefit ratio would warrant perhaps earlier therapy; and then we need to create a stepwise approach to including pediatric patients as we move forward into gene therapy for children.

So we need a balanced approach to pediatric patients. The benefits for gene therapy obviously are consistent levels, bleed protection, decreased

burden of care, improved quality of life, and tolerance even for some patients with inhibitors. The risks include perhaps a waning level over time, a need for reinjection in the presence of positive antibodies, consideration for hepatotoxicity, and some late effects, including malignancy, and then their very long life expectancy; how do we monitor these patients? What are our care plans for follow-up of these patients? What are the best modalities for following them?

Against that, we have to balance new agents that have come to market, including novel agents such as emicizumab and those in clinical study, for example anti-TFPI inhibitors; and then also extended half-life products. And I've highlighted factor IX here because, clearly, what we've been able to achieve with extension of half-life for factor IX has been much better than as compared to factor VIII, although at ASH we heard about a new factor VIII engineering that extended the half-life at a higher dose of up to 7 days. So there are some nice things that are coming along the pipeway

as well.

We need a balanced approach and a stepwise population approach to pediatric patients. We need to determine the durability of the response, especially for the less mature trials; determine the optimal level required based upon the age of administration; and we need probably a better idea of a dose-response curve as we're treating these patients so that we know exactly what we're going to get when we expose a patient to gene therapy; determine the need for further data based upon the deficiency itself in the vector; evaluate the risk in children based upon the current therapies and the current burden of care in populations that could represent increased need such as inhibitors.

So going back to Dr. Lozier's questions, how long is data required before we proceed in children -- and I didn't mean this in a facetious standpoint -- really, the longer the better for response duration in safety, especially as you approach children.

The duration of vectors sustained expressed

in children; is it different in adults? Yes. Their life expectancy is longer. We need a longer duration to assure that what we're doing is safe and beneficial and really exposes them to a risk.

What's our risk-benefit ratio in terms of the burden of care, and can we define the target level in children based upon the age treated to achieve a reasonable level as an adult?

Well, we'd have to work backwards. These are just guesses, but if we got a level of 30 percent or above for factor IX, and if we got a level of above 30 to 40 percent for factor VIII, that's likely better than what we're achieving with current therapies, including novel agents. That would consider perhaps a different weighing of risk versus benefit and burden of care for patients.

How do we proceed in children, and do we consider age cohorts? I think, yes, we would have to work backwards unless we found a population that was of extraordinary need in a particular pediatric group, where current therapies are clearly not as good and the patients are suffering more sequelae.

And I would say that perhaps factor IX-deficient inhibitor patients represent one of those groups, although very small. I think that's it.

(Applause.)

DR. LOZIER: Thank you, Amy.

We'll be holding the questions until after our speakers have finished their presentations.

Dr. Stacey Huppert is an associate professor of gastroenterology, hepatology, nutrition at Cincinnati Children's Hospital Medical Center and at the University of Cincinnati College of Medicine. Her research specifically focuses on hepatic cell plasticity commitment and therapeutic potential of differentiating hepatocytes. She also works on the molecular regulation of hepatocyte differentiation via transcriptional networks in the epigenetic landscapes.

I thought she would be very well positioned to give us a talk on the development of the adolescent liver. Stacey?

Presentation - Stacey Huppert

DR. HUPPERT: Good afternoon. So this is

definitely a different type of meeting than I normally go to, but it's been very enlightening. Jay had given me three areas to talk about considerations for hemophilia gene therapy treatment. They're listed here, basically talking about hepatocyte, differential gene expression, and physiological function that evolved from a neonatal period to adolescent stages.

I added in models for molecular regulation and hepatocyte differentiation, where the field is at this point in time, what we know about it, and then finally liver growth, which has come up a lot so far. This is my funding.

As Jay said, really, the bread and butter of my group is really looking at molecular factors involved in regulating cell identity and commitment in the liver. For this group, really, the important things are in the orange box down below. As we all know, the liver is alone in solid organs and its ability to regenerate mass, so we need to think about that all the way through life. And mouse studies in the last couple years have really

shown us in cell fate tracing studies, that there's no evidence of a contribution of a reserved stem cell population.

I've diagrammed that in the right side. You can see that hepatocytes and cholangiocytes, which make up the bile duct epithelium in the liver, are really in states of transition when you are replacing mass of either this population or this population. So you need to think about, in states of liver disease, that cells are continually in flux, and this makes a difference when you're trying to find vectors that hit a specific cell identity.

Just to set you up about hepatic architecture, I think when we're talking in this group about trying to target hepatocytes to express different factors, we need to think about all hepatocytes are not the same.

I'm just showing you this diagram here where the hepatocytes in zone 1 do very different functions from hepatocytes in zone 3. They produce substances and metabolites that are secreted into

this canalicular membrane and go into the bile duct. Then in this structure here, you can see are then exported out of the liver. Hepatocytes also dump on their basal lateral side substances into the blood that is carried out of the liver.

A liver-centric view is really that hepatocytes perform a very specialized function, yet they remain very plastic in adults and in children. The other issue is that the absence or low expression of many hepatocyte-produced enzymes at birth is thought to be responsible for the differences in pharmacokinetics and toxicity between pediatric and adult populations.

Here, two extreme examples are glutamine synthetase. The hepatocytes that do a lot of this function are in zone 3, and cholesterol synthesis, the hepatocytes that do that function, are mostly in zone 1. So there are very diverse populations of hepatocytes in the liver.

These are images from an experiment that Abby [ph] in the lab performed just to show you visually the changes of hepatocytes and some of

their functional enzymes where they're expressed.

On the left-hand side is in a mouse embryonic liver at 14 and TBX3 is a transcription factor. You can see red in the nucleus. Glutamine synthetase that I told you in adults is in zone 3 hepatocytes. You can see that all hepatocytes expressed both of these markers early postnatally, so 3 days after birth in a mouse, you can see the glutamine synthetase is mostly located in zone 3, whereas TBX3, the red, is still diverse in its expression pattern, but it's starting to resolve.

At 4 months of age, you can see glutamine synthetase is tightly correlated with a central vein area of zone 3, and now TBX3 is localized there. These are just background because we have to amplify to see that signal. So there's really a chance in the expression pattern across the liver.

The other thing that we want to think of and, especially bringing up hepatocellular carcinoma or liver cancer, are these factors at the top. I'm showing you 3 factors, delta like 1, alpha-fetoprotein, and glucagon 3, that are

expressed highly in early postnatal liver and also in hepatoblast or embryonic liver. But as the mouse ages from 15 days, 21 days, 28 days, that gene gets shut down, and these are all factors that get re-expressed in hepatocellular carcinoma.

Here on the bottom art, I'm showing you just a few markers, which are known to be involved in the canalicular membrane or forming that secretion level. So as hepatocytes start to mature, they start to up-regulate expression of these genes and functional genes within the liver.

I think the clearest example of the changes that happened; here, I'm showing you our cytochrome P450, which was the example earlier, that are really phase 1 enzymes that are involved in metabolizing many different chemical compounds in the liver. You can see here that in mouse doing RNA sequencing and of all the genes expressed in the liver, that there are two surges. There's one that happens a few days before birth, and then a few days after birth, you can see the surge of a few P450 genes.

Then there's another surge that happens between 10 and 20 days, and that's really still core to this peak and liver volume or growth in mouse liver, and also at the time when weaning and changing of food diet happens in the liver.

At the bottom, I'm not going to go through it, but you can see that these P450 genes can really be classified into 4 different groups, ones that are very early in the neonatal liver that reach peak and then decrease, and then adult. Over here, you can see that they don't become expressed until about mid-gestation out a few days, and then they level off and peak out here.

In this slide, I wanted to show you that this is just a visual representation of specific cytochrome P450s. These labels are all incorrect here.

This is at day 10 and this is at day 20. This CYP2D1 is not expressed if this pie graph would have showed up here at neonatal times, and then starts to increase, whereas a few other cytochrome P450s are not expressed at all in

neonatal and then start to be expressed in adults. So there's a big switch.

This is also observed in humans when you're looking at proteomic profiling of P450s, that you can see that some of these cytochrome P450s are expressed at a low level no matter what age, then up here is late first trimester, all the way up to adult. Some are expressed at fairly high levels no matter what age, the hepatocytes are. Then there are some that are very low expressed in the early liver and hepatocytes, but them become upregulated. So you can see there's definitely a transition of the liver and the hepatocytes as they mature.

What's the molecular regulation of this? There really have been found 6 key master regulators or liver-enriched transcription factors that are expressed in the liver, both at embryonic and adult times. One of the areas that we're really interested in is how do these master regulators, which are expressed at both these times, really coordinate the transcriptional changes that happen and are necessary for organ

maturation and also to mature hepatocyte physiology.

There are two different models that are thought about. One is progressive assembly of transcription factors, that you may just have a couple on gene X, but in adult hepatocytes, you have 4, 5, or 6 of these master regulators that are sitting on the promoter.

This, you can see in mouse, looking at embryonic day 14 to postnatal day 45. If you focus in on Hnf4, which is this center circle right here, you can see that the number of arrows pointing in on Hnf4 increases with age, meaning that many more of these liver-enriched transcription factors are sitting on the promoter.

In human, chip sequencing has been done on the genome, and many genes you can see have 2 regulators, 4 regulators, and 6 regulators, so there's this reinforcement and progressive assembly on the promoters.

The second model is really differentiation-dependent enhancer switching, and this is from

Pamela Hoodless' group, where she's shown, if we're focusing on gene X in a neonatal stage, you can see that it's bound by these couple transcription factors, but if you look in adult hepatocytes, it's no longer bound. But if we look at gene Y in neonatal hepatocytes or hepatoblast, gene Y has no transcription factors, only enhancers, but in adult hepatocytes, now you see occupation. So there's this switch of what's regulating hepatocytes.

The other thing we need to think about is also epigenetic regulation. If you take public data from in ENCODE and look at H3K4 monomethylation, you can see that there's differences in changes in the pattern of where the peaks are, and the binding of these different histomodifications receive bimodal distribution in adults where you have enhancers bound inside a promoter, and monomodal if you don't have binding.

I also had one of the bioinformaticists in our division to look at the ENCODE database to see DNA sequencing if you look at hepatocytes at birth and hepatocytes in the adult. And this is just a

region where we knew there were adult-expressed cytochrome P450s but were zoomed quite a bit out. You can see that whole region of the chromosome is regulated so that, at birth, it's completely closed down, and in adults, it's opened up for expression.

There are really dynamic and epigenetic changes that occur in the postnatal liver as it's maturing, and these hepatic master regulators obviously play a very important role.

I just pulled out a few genes that would interest this audience, and it's not an in-depth bioinformatics that were done, but just to look at RNA sequencing, you can see that, some of those, yellow means higher expressed postnatally at 28 versus day 7. Some of them get up-regulated. Some of them get down regulated.

If we look at the promoters of factor VII and factor IX, which are expressed in hepatocytes, you can see that they're all bound by Hnf4, one of these master regulators, but all the work has been done in a very minimal promoter situation. So we don't know anything about epigenetic regulation.

My last topic is liver growth. In a mouse, the peak of liver growth and proliferation is around postnatal day 10 and 20, and reminder that this is when cytochrome P450 transcripts are really having that high surge that they're changing into postnatal differentiation.

This correlates with humans in body growth. The liver is really tied to the metabolic requirement of the organism. This is showing you liver growth. This is a group at Cincinnati Children's that has really looked at bone mineral content with size and height growth of normal children, both African-American and non-African-American. The girls are these solid lines here, that peak in their linear body growth around age 11, and boys here are peaking around 13. This really matches the CDC stature for age and weight growth.

When we look at liver volume by micro-CT, really, the conclusion, just to cut the story short, is that there is significant change in liver volume in these ages when you look at a couple

months of age down to 18 years of age.

When you get here to age 13 to 18, there's not a significant difference, but that's with no liver disease going on in these kids. There also is a decrease, when you look at liver to body weight, from a couple months old into 18 years. So there's really a rapid increase in infants, there's gradual increase in liver volume in school children, and there's not so much in adolescents in normal kids with no disease.

This is just one study for your reference that went through all of the micro-CT studies at the time to look at combining all the different reference sets from different ethnicities. It basically comes down to the same conclusion, that the liver is about 4 percent of the body weight in infants compared to adults, where it's around 2 to 3 percent. Really, the best correspondence is body surface area to liver volume versus looking at weight and height.

This is one of the more recent studies, which was done in 2011, which really was trying to

get at very neonatal early liver size, and it predicts a little bit better than some of the early studies that were done. What they show is that there is this difference once a child hits 20 kilograms in the slope of the curve. and the other magic point is 110 centimeters in growth height.

Just to finish up, as far as hepatocyte differential gene expression and physiological function from neonatal to adults, there is this spatial and temporal changes that happen with age and that hepatocytes remain plastic even as cells with specialized function. That's very important as you're targeting in a non-diseased versus disease state, if there's any underlying liver disease that the vectors may be targeting different cells.

Also, models for molecular regulation of hepatocyte differentiation really begins to basically lay out what the impact might be if targeting specific cells and the impact of choice of promoter for gene therapy, and also, really,

what's open epigenetically if we start discussing integration of some of these vectors.

Finally, liver growth, it increases basically with human linear body growth. and liver volume seems to subside around 11 to 15 years of age, and this may impact the timing of vector delivery. That's it. Thank you.

(Applause.)

DR. LOZIER: Thank you very much, Stacey.

Our next speaker is Dr. Mark Sands, who is an NIH-funded investigator in genetics at the Washington University of St. Louis and studies various lysosomal storage diseases. In the course of his experiments with AAV gene transfer, he made some very critical observations about the incidence of hepatocellular carcinoma in mouse models, so we thought he would be a very good speaker to tell us about what some of the preclinical animal data are for this risk factor.

We're also grateful that you broke away from a site visit for child health and development to come here.

Presentation - Mark Sands

DR. SANDS: Thank you, Jay. Actually, I appreciate you tearing me away from a site visit. This is better.

What I'm going to do this afternoon is tell you about a rather troubling finding that we had a number of years ago, and this association of AAV-mediated gene therapy and hepatocellular carcinoma in our mouse models. I have no conflicts of interest to disclose at this point.

back in the mid- to late 1990s, we did a number of experiments using AAV-mediated gene therapy to try to treat our mouse models of lysosomal storage disease. Since these diseases are progressive, the question we were asking is if we deliver this vector during the neonatal period, when they're pre-symptomatic, can we prevent the onset of the disease?

To summarize 10 or 15 years' worth of work, the answer is, yes. If we deliver these vectors very early on, they have a much better impact. But as part of those studies, we did several lifespan

studies. And what we discovered is that animals that lived a very long time -- and when I say very long time, a year or more, what we discovered is that there was really quite a high frequency, about 40 percent, of our AAV-treated animals that developed hepatocellular carcinoma. Now, again, we didn't see it very often before 1 year of age, but again, between 1 year and 18 months of age, there's relatively high frequency of hepatocellular carcinoma.

In fact, the average age that we saw this was about 16 months. And one thing that was very puzzling was when we were analyzing these animals to try to determine if AAV might be the causative factor, we hypothesized that if it was, we should see about 1 AAV vector genome per cell in the tumor tissue. Interestingly, what we saw was very much less than 0.1 vector genomes per cell.

So this actually suggested to us that it might not be AAV. But we had been studying this particular mouse model for the last 10 or 15 years, and what was very troubling to us was simply the

presence of hepatocellular carcinoma. We had never observed that before in any of the studies we had done, and we developed other therapies that would make these mice live a long time.

So this raised a number of questions to us. The first question, and what we actually had hoped for, was that perhaps there was some contaminant, either infectious agent or a chemical agent, in the AAV prep that would ultimately lead to hepatocellular carcinoma.

Also, another question, is the hepatocellular carcinoma disease specific? So is it a feature of mucopolysaccharidosis type 7? Is it mouse-strain specific? Is it transgene specific, dose dependent, age dependent? Is it AAV-serotype specific? And again, this question that really bugged me for a long time was why do we have very much less than 1 vector genome per cell in the tumor tissue?

So the first thing that we needed to do was to try to replicate this finding. Again, this potentially could have been a one-off observation,

never seen again, so we tried to replicate it. And to make a long story short, we were able to replicate it. And over here, on the lower right, that shows a typical liver from an aged animal treated with AAV. And what you see is a little bit of normal-looking liver tissue and then usually multiple tumors within that liver.

So we did the exact same experiment we did the first time. The mice received an intravenous injection of an AAV-2 vector the day they were born, during the neonatal period.

In this particular experiment, exactly half of the animals treated with AAV developed hepatocellular carcinoma. Now, what we did determine here was that it was not disease specific because the MPS 7 animals, half of those had hepatocellular carcinoma and half of the wild-type animals. And these were littermates, so there's no differences in the genetics here.

Same thing; we saw a rather protracted phenotype. The hepatocellular carcinoma showed up between 54 and 72 weeks. We also asked the

question, is it an infectious agent, and the most likely culprit would be Helicobacter hepaticus, which is known to infect mice, and ultimately result in hepatocellular carcinoma. All of our mouse colleagues are Helicobacter hepaticus negative.

Also, the strain of mouse that we use; all of our disease models are on the C57 black 6 background. And if you go to the Jackson lab website and you look, part of their website is a table of tumor susceptibility in various strains, and C57 black 6 are relatively resistant to hepatocellular carcinoma.

I'm not going to go through this in detail, but this is a table showing the breakdown on the various animals. This was the original observation, which we replicated, and you can see about half of those animals developed hepatocellular carcinoma.

Importantly, the wild-type animals, you see the same proportion, but we answered a couple of other questions here as well. These MPS-7 animals;

if we treat them with bone marrow transplant, which extends their lifespan, or if we treat them with radiation to try to bring out this phenotype if it's a function of the disease, very few of those animals developed hepatocellular carcinoma.

There is some low frequency of hepatocellular carcinoma in the untreated wild-type animals, but it's less than 10 percent. And this is an important group right here as well. These are untreated transgenic animals. So we have a transgenic animal that harbors the same transgene as our AAV vector, and this animal produces about 20-fold higher than normal levels of beta glucuronidase, and you can see no hepatocellular carcinoma.

Now, the truly striking finding, though, from this replication experiment was when we tried to pull out junction fragments -- in other words, insertion sites from the AAV vectors -- we were able to isolate 4 junction fragments from 4 individual mice, and these junction fragments are represented here, here, here, and here. And again,

the really striking finding is, all 4 of these landed in essentially the same spot. It's within a 6,000 base-pair region of what's referred to as the Rian locus, which is on the distal end of mouse chromosome 12.

So all of these fell right into this little bitty area here, and when we analyzed the transcription of downstream genes and micro-RNAs, all of these were dysregulated.

So we answered a couple questions with this replication experiment. First of all, we had no Helicobacter hepaticus in our mouse colonies. This doesn't completely eliminate an infectious agent, but this is the most likely candidate here.

Wild-type animals had the same frequency of HCC as the MPS-7 mice did. All of our mice are on a C57 black 6 background, which are relatively resistant. Is this transgene specific? Well, probably not; at least our transgenic animal would suggest that the presence of that transgene and dramatic overexpression is not a problem.

Is this dose dependent? We don't know yet.

Age dependent? Don't know yet. Why is there less than 1 vector genome per cell? When we originally did our analysis on the first observed hepatocellular carcinoma, we were trying to quantify the vector genomes by using primers within the transgene, and all of those insertion sites are rearranged AAV vectors, and the transgene is gone. And that's typical for what people are finding when these things integrate. So that explains why we had this strange number initially.

Then quite a bit of time went by, and no one else had ever replicated this finding until, in 2013, a group in Pennsylvania was working with gene therapy for ornithine transcarbamylase deficiency, and they had earlier published a paper where they had injected AAV vectors in the neonatal period, and they discovered a high frequency of liver tumors.

Their initial conclusion was that it was caused by something else and not AAV. But once we published our data and then they went back retrospectively and reanalyzed those tumors, they

actually did find a number of tumors with AAV integrations within the Rian locus, very much like what we saw. In fact, on that chromosome, it was very near where our integration sites were as well.

Then in 2015, there were two papers that came out almost simultaneously. One was from a group in Canada that was studying Sandhoff disease, which is another lysosomal storage disease. They did the same thing; IV injection at birth to try to prevent the onset of the disease. Their mice were also on C57 black 6 background. They saw 80 percent of their AAV-injected animals develop hepatocellular carcinoma. Again, it's a rather protracted phenotype. They also saw high frequency of AAV integration in the Rian locus.

At the same time, Chuck Venditti's group, who's at the NIH, who studies methylmalonic acidemia, again, did the same experiment; IV injection, newborn animals, and about 50 percent of his animals also developed hepatocellular carcinoma. He used several different serotypes, same thing; high frequency of AAV integration

within the Rian locus, and also just like we saw, dysregulation of downstream genes.

I'm not going to go through this whole table because it would take me too long, but this highlighted region, Chuck was able to answer several other burning questions that we had. These two groups here, this AAV vector had a very strong promoter, the CBA promoter. But they were injected with a relatively low dose of virus, 10 to the 10th vector genomes, and you can see the frequency of hepatocellular carcinoma is quite low.

In contrast, all these groups here in green, same promoter with one exception, serotype 8, but they were injected with a dose vector 10-fold higher, so 10 to the 11th vector genomes per mouse. This is where you see all the hepatocellular carcinoma.

He had two more groups, same serotype, same dose, 10 to the 11th vector genomes per mouse, but in this case, he had a much weaker promoter. This is the human alpha 1 antitrypsin promoter, which is much weaker than either the TBG or the CBA

promoter. And you can see down over here, no hepatocellular carcinoma.

The other thing that Chuck did, which is really helpful for this analysis, he pulled out the sequences for a portion of the Rian locus from multiple species; mouse, rat, importantly human. He got elephant DNA, too. I'm not quite sure where he got that, but he directly compared these sequences. And what he discovered is that there's about a 65 base-pair region in the rodent genome that is unique to the mouse and rat. It's not present in any of these other species, and importantly, it's not present in human.

Then what he did is he superimposed all of these integration sites that were identified by multiple groups; Chuck's group, our group, another group. And you can see that a number of these integration sites fall right within this unique region.

Now, I will caution you at this point. This is a little bit misleading in that these integration sites, this only represents about

60 percent of AAV integration sites. About 40 percent of the integration sites that people have pulled out are outside of this unique region, so in regions where there's nearly perfect homology between the mouse and human.

So what about this issue of newborn versus adult? There are two studies here, both groups studying hemophilia. This is Kathy High's group here. This is a group from Japan working with a Padua mutation. They injected young adult animals with high doses of AAV, and then asked the question, do they develop tumors?

What you can see is when you postpone the injection to young adults, you see the frequency of hepatocellular carcinoma decreases dramatically. It doesn't drop to zero if you look. If you read the entire paper, it does look like there's still some propensity towards hepatocellular carcinoma, but it's dramatically reduced.

Finally, for the data slides, this was published just recently, in 2017. This is a study where a group did in utero IV injection into fetal

macaques. So it's a primate model, IV injections in a fetus, what would be closest modeling to a newborn mouse, I guess.

What you can see is there are two animals here at least that have an enormous number of unique integration sites within the genome. Keep in mind this is from a needle aspirate, so it's not a big chunk of tissue. And again, many, many thousands of unique integration sites, which is troubling. In fact, if you read this paper, they're troubled by this as well.

But what I will say, this is a 6-year follow-up from these animals, so it was 6 years ago that these animals were injected intravenously in utero, and so far, there have been no adverse events noted from any of these animals, so it's not clear it's a problem.

So at this point, what do we know? We know that AAV integration in and disruption of the murine Rian locus can cause hepatocellular carcinoma. It seems to be independent of disease model. There have been lysosomal storage diseases

and methylmalonic academia and ornithine transcarbamylase. They all develop hepatocellular carcinoma.

It seems to be independent of serotype. It is age dependent, and newborn animals seem to be much more susceptible to hepatocellular carcinoma development than do young adult animals. It seems to be promoter dependent. Strong promoters have a greater propensity for developing hepatocellular carcinoma than we weak promoters.

There's a high frequency of AAV integrations in a rodent-specific region of Rian. And again, there's a large number of unique AAV integration sites throughout the genome, independent of Rian, in this primate study. But again, I'll point out, so far, there's been no hepatocellular carcinoma or any other adverse events noted in those animals.

Finally, what don't we know? Well, this is a really short list. There's a lot we don't know about this yet. But first and foremost, what we don't know is AAV-mediated hepatocellular carcinoma

problematic for human gene therapy? And we really don't know at this point, and it's extremely difficult to accurately model.

Are other tissues also susceptible to malignant transformation? There hasn't been a lot of reports. There is one report where other types of tumors have arisen, but it's not widely known at this point.

Are there other consequences, either acute or chronic, of AAV-mediated gene therapy? And again, importantly, can the AAV vectors be redesigned to be safer? Chuck Venditti's data would suggest that that may be possible, but what you may be doing is trading efficacy for safety, and trying to find some balance there.

With that, I'll stop, and I guess you're holding questions until later. Thank you.

(Applause.)

DR. LOZIER: Thanks, Mark.

Our next speaker is Theo Heller, who is the chief of translational hepatology in the liver diseases branch, in the NIDDK institute at NIH.

His work includes studying factors that cause progression of liver disease and rare liver diseases. They've got liver access in the microbiome, and we thought he would wrap things up and tell us what are we going to do with these safety signals. Theo?

Presentation - Theo Heller

DR. HELLER: Thank you, Jay.

Thank you, everyone. I thank you all for being at the end of a long day. I'm impressed that so many people have stayed, and that's why I'm particularly grateful to Jay for telling me I have 2 hours to review the literature, including the molecular aspects of hepatocellular carcinoma. I refused. I said, "I'm not going to do it. I'm going to stick to five minutes."

So I'm going to do a very conceptual talk. I'm going to try and fill in thoughts as we go through concepts, and I'm going to tell you how I think about these things and how I approach things.

The first thing we should talk about is just some definitions. Screening is when you look

once. All the baby boomers are now recommended to have hepatitis C testing 1 time. Surveillance is where you look repeatedly. Someone is at risk for carcinoma of the cervix, pap smears are done repeatedly. You wouldn't just accept one. An objective for both is to reduce disease-specific mortality.

There's a paper that I would recommend from the American Journal of Pediatrics Hematology and Oncology in 1992 because I think it's really a good approach to surveillance and how we should think about surveillance. There are a couple of points that I'll make, and I'll fill in as we go.

First of all, you have to have a common disease with morbidity and mortality, so if we think about hepatitis B and liver disease, once patients developed cirrhosis, the risk of cancer is 3 to 8 percent per year. That's significant morbidity and mortality, and it's relatively common in that population.

Easily identifiable target population; yes, hepatitis B, hepatitis C, Wilson's disease

hemochromatosis, these are diseases you can identify, diagnose, and follow. As a subtext in that category, surveyors have to agree -- in other words, the physicians or the mid-level providers whose job it is to survey the patients have to agree that this is something we should do. And the target population, the patients, have to agree that we want this sort of screening.

The test has to have low morbidity -- in other words, we're not going to take off your right leg to see if you have a clot inside it -- high sensitivity, and specificity. There is to be a standardized recall. In other words, what do you do if you do an alpha-fetoprotein and it comes back high?

We have to have a standardized approach to follow through on that. It can't be that some people say, "Let me re-check in 3 months," some people say, "You're probably flaring; let me check the ALT," and other people do further imaging.

There has to be a test acceptable to target population. If we recommended colonoscopy every

3 months, I don't think we'd get everyone agreeing. The fact that we suggest it every 10 years as gastroenterologists makes it palatable. I think gastroenterologists would like to do it more.

There has to be an acceptable and effective therapy. So for hepatocellular carcinoma, in the early stages, now we have very effect therapies. Resection and transplant has dramatically changed the landscape. Once tumors are advanced, the standard of care is palliative. That's an important thing, advanced disease, palliative care; early disease, possibly curative; even 60, 70 percent range.

This is not in that paper, but this is something I added. There's an important concept of competing mortality. We'll come back to that in the guidelines, but someone with metastatic lung cancer doesn't need to have a colonoscopy to check if they have polyps. So we need to bear in mind what the patient looks like, who the patient is. We can't just stay this is the test you should have, this is the guidelines. We need to think

about who we're dealing with.

This is something that I get asked a lot and comes up a lot, surveillance versus diagnosis. Once you have an abnormality, you're no longer surveying. And if we stick to the theme of hepatocellular carcinoma, if the alpha-fetoprotein is high, we don't do an ultrasound to follow up on it. Ultrasound is a screening test.

If you have an elevated alpha-fetoprotein, you would go to an MRI, or if you have an ultrasound that shows a nodule, you wouldn't then do an alpha-fetoprotein, you would go to an MRI or a CT scan. This concept of repeating another screening test is something we run into all the time and delays care.

Biology break. In general, hepatocellular carcinoma requires risk factors, and the most significant is cirrhosis. Eighty percent of hepatocellular carcinomas will occur in cirrhosis. That makes it easy, again, to define the population that should be screened.

These are general concepts. This is not

unique to hepatocellular carcinoma. This is by the economics people, the CMS-type people, and the people who sit in front of computer screens. Their outcome desired is that we should increase survival by more than 100 days, and the second is that it should be cost effective.

There's the concept of QoLies [ph] or year-of-life gained, and the cost should be less than 50,000 per year. That's for the whole population screened, not for the individual patient where you find something, and that takes in work, hours lost, and all sorts of things.

These are the guidelines. From this year, update is from this year, the American Association for the Study of Liver Disease puts out regular guidelines, and the recent most up-to-date guidelines say that in adults with cirrhosis, we improve survival by screening. That's without question.

What's recommended is an ultrasound with or without an alpha-fetoprotein. I'll get back to alpha-fetoprotein and why that says with or

without. And it's recommended to do it every 6 months.

That every 6 months is not a convenient time frame. It's based on biology. Given the doubling time of hepatocellular carcinoma, the optimal time for most patients would be 4 to 8 months. So if you screen every 6 months, you're less likely to miss tumors of significance. You're still likely to find small tumors.

Do not screen Child C. Child classification is how we think of cirrhotics. A is good. C is very bad. C is close to death, decompensated yellow with ascites. And the mortality there is so high once they reach Child C, that there's no point in screening for hepatocellular carcinoma because even if you find it, they're likely to die of the liver disease first.

Novel biomarkers; everyone is very excited. There are 186 gene profiles that have been looked at. There are all sorts of novel panels looking at different genes. They require further evaluation.

There are other biomarkers. There's AFP-L3 percent. There's DCP. You might have heard of all of these.

Some of these are FDA approved for risk stratification. Once you have something, but not approved for screening, the jury's still out. And if you look, CT or MRI is not recommended. That's because of cost, because of radiation, because of convenience. We're talking about ultrasound, a few hundred dollars, CT, MRI, a few thousand dollars. It really changes the equation.

There are exceptions. Patients who are very obese are very difficult to do an ultrasound that's high quality. Patients who can't go into a CT scan are allergic to contrast, you might come back to an ultrasound. Again, it's a matter of looking at the patient and not being fixated on guidelines.

What about gene therapy in our situation? It's not quite surveillance because we don't really know that adult humans getting gene therapy are at risk for hepatocellular carcinoma in this setting.

So it would be different. The risk is unknown. And surveillance is really defined in the context of prevalence. We have no idea what the prevalence is, so we can't talk about surveillance.

There are thousands, tens of thousands of patients who went into studies from which the guidelines were derived. We don't have any patients here.

There are other needs, the competing needs. As scientists, as an approval agency, as physicians taking care of the patients, and as patients, we want to know if cancer really occurs. How risky is the therapy? What is the percentage? And we want to know this with some rigor. We want as small a margin of error as possible. So how badly do you really want to know? Because if you want to know really badly, you would not screen with alpha-fetoprotein and ultrasound.

What are our options? The first option is to do nothing. The second is blood tests. Third is imaging. The liver biopsy always comes up. I spent the morning doing liver biopsies. I love

them, and the patients don't always; different story.

Symptoms are no longer surveillance. So once the patients have symptoms, we're no longer talking about surveillance. So the first option, do nothing. I don't think that's an option. The second, blood test.

These are best studied in regular cirrhosis. An alpha-fetoprotein, more than 20, is considered the cutoff. Normal range in my hospital at the NIH is 6.6, so 20 is more than 3 times that. Results vary at different labs. You were speaking about that earlier.

Sensitivity of 60 percent; that's not great. Specificity of 90 percent, and if hepatocellular carcinoma is 5 percent, it's a 25 percent positive predictive value. There are variances already mentioned the AFP-L3 percent, the DCP, for risk stratification.

What about novel tests? Well, there's even less known, and even less known in this setting. What about imaging? Ultrasound is the best

studied. Cost-wise, it's the most effective. It's very available. It is somewhat operator dependent, but I think that's less and less of an issue with higher-quality machines in academic centers.

Efficacy, it's pretty good, and we can go to MRI and I'll put that into context. MRI is a lower false-positive, 3 versus 5.6 percent. It is a greater sensitivity and specificity, 80 to 90 percent and 91 to 98 percent. And I use the recent reference, Kim in JAMA Oncology from last year, but there are many other references which show similar things.

The MRI has to be dynamic. That means they have to get contrast. It's about 45 minutes to an hour, and it requires a center that's comfortable and familiar with doing liver MRIs. We see a lot of MRIs from smaller community hospitals; they're not adequate.

CTs have to be 3-phase, again, with contrast, and that's a significant amount of radiation. So as hepatologists, we are doing less and less CTs because we don't like the abdominal

radiation. We're moving more and more towards MRIs, but again, that requires greater facility with it and also cost is greater. But CTs are easier to reproduce and are more generally available.

CT and MRI are very helpful. Actually, it's considered diagnostic in most instances. And in the liver, we find lots of other things, which is why I like being a hepatologist. We find focal nodular hypoplasia, we find hemangiomas, we find all sorts of things, and ultrasound can't distinguish that very accurately, but MRI and CT are very good. So that makes it easier and less likely that you'll go down a rabbit hole.

What about biopsy? I apologize for the small print, but I really wanted to include these concepts. It's invasive. There's risk. It's 150,000th of the liver, so to do a blind biopsy in someone with hepatitis C where the whole liver is affected, if you have an adequate biopsy, your risk of sampling error is less than 2 percent and 98 percent good; same for hepatitis B; same for

autoimmune hepatitis. But if you're looking for random hepatocellular carcinoma, is 1 in 50,000 adequate? No. And as I said, although it's good for me, going through it for the patient is not always pleasant.

This is from the guidelines. Biopsy may be required in selected cases, and this is for diagnosis, not screening. But its routine use is not suggested. Biopsy has the potential to establish a timely diagnosis -- and, again, diagnosis -- in cases in which a diagnosis is required to affect therapeutic decision making.

However, biopsy has a risk of bleeding -- it's a good thing there's no risk of bleeding in this patient population -- and tumor seeding -- in fact, some transplant centers won't do liver transplants in patients who have hepatocellular carcinoma and have had liver biopsies because of that risk -- and the possibility that a negative biopsy is attributed to the failure to obtain tissue representative of the nodule rather than a truly benign nodule.

Then I included something else from work from the NIH. We looked at our last 3 and a half thousand liver biopsies, and we looked at risk of complications, and we published this last year. Compared to viral hepatitis, biopsies performed of certain diagnoses had significantly higher odds of major complications: NRH, drug-induced liver injury, GBHD. And look at the odds ratio for hepatocellular carcinoma, 34, greater risk of complications compared to viral hepatitis.

So that's one of the reasons we don't like to biopsy hepatocellular carcinoma and one of the reasons we rely on CT and MRI criteria. And we do biopsy if we have to, but it's not just to be certain and because we're curious.

Furthermore, by multivariate backward logistic regression -- don't ask me any questions about that; I don't understand what that means -- platelets less than 100 and APTT greater than 35 were independent risk factors of post-biopsy bleeding. So I think we can put biopsy to rest.

Where does that leave us? So we have to make peace with the silver standard. Imaging is the core, and I don't mean core biopsy. I mean core. That's what we rely on as hepatologists. We don't rely on the FP for the reasons that I explained. We rely on ultrasound to screen. If you really have a high-risk population and you want to know with absolute certainty, for example, a transplant population where people are going to liver transplant and you cannot afford to miss an HCC, we would rather use an MRI.

The age is important. We heard discussion from Dr. Sands about what time people are exposed to risks. And it's true, in human disease, too, the earlier you are exposed to hepatitis B, the earlier you develop cirrhosis, the more time you have to develop cancer.

How long people have had the disease, so even if you were affected as an adult, your risk factor started at adult. It's not the same when you're 30 as when you're 60. And when to stop screening or when to stop surveying; 10 years after

gene therapy, 20 years after gene therapy, 5 years when the factor levels drop? I don't know the answers.

I would say that, for me, thinking about this patient population, this is not the same as surveying a patient's group with hepatitis B. I'd want to know with certainty. I would not be comfortable with a 60 percent sensitivity. I'd want to go to something a little bit more certain.

Biology is great. We can never have 100 percent certainty, but as close as we can get. Thank you all for putting up with me and listening to the last talk on what looked like a fantastic day.

(Applause.)

Panel Discussion

DR. LOZIER: So at this point, we'll open things up for some discussion and questions with the panelists. Stacey had to catch a plane, so she's not with us. Don't take it personally.

I had a question, I guess, first for Amy. And I would say people should be ready to ask

questions here at the microphone. I want to get some things into our panel specifically. But go ahead and come to the microphone, and we'll also look at Slido questions if they're pertinent to this session.

So you talked about special population, pediatric populations that might be, say, attractive targets for gene therapy, in particular inhibitor patients. Would you worry about either exacerbating a factor IX inhibitor titer, and then have continuing production of factor IX in those patients which could lead to complement-mediated disease?

For instance, if you get factor IX and have anaphylaxis, you don't do it. But once you give the gene therapy, you can't go back.

DR. SHAPIRO: Yes. That's a very good point. There was some very good work presented at ASH, looking at platelet-derived gene therapy with factor IX in a mouse model where, actually, the mice do get anaphylaxis when they're exposed, and they were tolerized using that method.

So there might be specific modalities that you could consider. But yes, if you were just using standard therapy and that patient had an anaphylactoid phenotype, you'd be very concerned about avoiding that or developing -- even if you were able to desensitize them, you'd be concerned about the longer-term effects of, say, nephrosis in those patients.

DR. LOZIER: Was there any evidence for complement-mediated problems with that, that you know of?

DR. SHAPIRO: It was just a 10-minute abstract, but it was Dr. Montgomery's group, who I think is gone now. But no, there wasn't anything that was presented.

DR. LOZIER: I guess a similar question would be, for factor VIII inhibitors, we think that continued exposure to factor VIII is usually okay because it's a non-complement fixing IgG4 antibody most of the time. But would you worry about something about gene therapy could change the subclass to one that fixes complement or causes

problems?

DR. SHAPIRO: It hasn't seemed to be a problem in those patients. Even those patients undergoing standard immune therapy with very high doses over very long times have not had that. So it's been a rare patient who's had what we'd call an infusion reaction in that category, whereas it's far more common in factor IX.

DR. LOZIER: I did have a question for Mark on the AAV story. You made the point that the promoter, the alpha antitrypsin promoter, you called a weak promoter. But it's a strong promoter in liver, is what I thought I understood. Or is that not really so?

DR. SANDS: I mean, everything is relative. Relative to the chicken beta-actin promoter, it's a weak promoter. Anybody who does this sort of work, the CBA promoter, if you wanted to direct very high levels of expression, that would be the promoter you would choose. And very much like Chuck, we've done some direct comparisons with CBA versus alpha 1-AT promoter. It's 5- to 10-fold weaker

than the CBA promoter.

DR. LOZIER: So I guess, if I was summarizing your talk, it seems like you've identified a signal that may be species specific, has a prototypical integration in the mouse, in the Rian locus, which is not found in humans or non-human primates, but there can be random integrations that, so far, are not associated with hepatocellular carcinoma that we know of, at least with 5 or so years of follow-up.

Is that about right?

DR. SANDS: Yes, that's correct. One thing that is pan species, if you will, for all the difference species that have been injected with AAV, the people that have looked have seen unique integration sites throughout the genome. The data in the mouse, in the Rian locus, it's the only example where there seems to be -- and I'm not even sure I want to call it directed, but there's a focal integration site. But if you look through the mice as well, in the young adults, the genome is littered with integration sites.

DR. GEORGE: Bindu George, FDA. I had a question for Dr. Sands. You mentioned that the vector was rearranged in I think it was the mice studies. Was that also observed in the non-human primate studies?

DR. SANDS: I don't know. They didn't evaluate it that carefully. They were simply looking for unique integration sites, but there were so many of them, they didn't do a detailed analysis on what the structure of the vector is.

Honestly, it's one of the major questions I have. I've never been able to get funding to look at it. But one question I have is, when we're seeing all these integration events, is it an acute event; in other words, immediately after or within a week or two after the injection, is that when these integrations occur? Or as these stable episomes sit around for 6 months, a year, 2 years, 5 years, is there some rate, continued rate, of integration? In other words, again, acute versus some continuous rate of integration as time goes on?

I think it's an important question. I don't know the answer. Nobody's done that experiment.

DR. GEORGE: I had a follow-up question to that. In terms of detecting insertional mutagenesis and using the PCR, what would be the implications of this information?

DR. SANDS: I'm not sure I understand your question.

DR. GEORGE: So if you're trying to look for these insertional mutagenesis, you're using a certain sequence, and you have a vector rearrangement here. How useful would the PCR probes be?

DR. SANDS: Well, it depends. I don't think there's enough information out there to give you a good, firm answer on that. In the mouse, all the junction fragments we've ever pulled out have been rearranged vectors, and primarily it's the 5-prime inverted terminal repeat that seems to get integrated, along with all the CIS-acting elements there.

Are we going to see the same thing in the dog, the primate, in humans? I don't know.

DR. GEORGE: Thank you.

DR. LOZIER: So we have a question at the microphone. Could you go ahead and identify yourself?

DR. BAFFI: Yes, Robert Baffi from BioMarin Pharmaceutical. I have a question for Dr. Sands. You didn't mention what production cell line system you used to produce your vector. And did you have a chance to evaluate if there was an impurity that might have facilitated the integrations you were seeing coming from the cell line that you used to produce the vector?

DR. SANDS: Sure. It's an important question. Our initial observation, again, we made back in the late 1990s, and we were making our own virus at that point. I don't know if you remember the technology from back then, but it was a transfection and then an infection with adenovirus, and then this very laborious purification process, which of course would increase the chance of some

sort of contaminant.

When we first reported this, that was the general consensus, that we had some sort of garbage in our prep, and it very well could have been. But since then, we've been having our vectors -- because it's cost efficient for me -- made either at the University of Florida, Vector Core, or at University of North Carolina Vector Core, which uses a column purification.

It's a mammalian system. What contaminants are in there? They do SDS page at the end, and it looks pretty pure. I'm sure there are things in there that we don't know what are in there. It's certainly not GMP-grade material. I'm not sure that helps at all. I think it's good quality material.

But the other reports that I've mentioned, Chuck Venditti's report, the Sandhoff mice, the ornithine transcarbamylase animals; all of those vector preps were made in different facilities. So if it is a contaminant, it may be a common contaminant. I don't know. But whatever it is, it

will come from multiple production facilities.

DR. BAFFI: If I could just follow up, are those other preps made from mammalian cell lines as well?

DR. SANDS: As far as I know, yes. I don't know exactly, but most of them are made from mammalian preps.

DR. LOZIER: We have a question over here.

MALE AUDIENCE MEMBER: I have a question to Dr. Shapiro. We shortly discussed about use of gene therapy in the patients with an inhibitor. Since inhibitor formation is really mediated by the T-cell responses, it's highly possible if in the liver cell -- factor VIII is produced in the liver cell. It's highly possible the T-cell really recognized factor VIII-producing hepatocyte, and it's a kind of undesired adverse cytotoxicity.

What do you think about that possibility, and what is your opinion about that one?

DR. SHAPIRO: If I understand you, you're asking, in patients with inhibitors who underwent gene therapy, could they suffer hepatotoxicity

because of the recognition of the T-cells against the hepatocyte?

MALE AUDIENCE MEMBER: Right, because the T-cell would recognize a factor VIII peptide, and a hepatocyte makes a factor VIII molecule. And through the ATC molecule, factor VIII peptide can be exposed, so that kind of situation.

DR. SHAPIRO: I guess I don't know the direct answer to that question, except that it's not expressed on the surface of the cell. It's secreted by the cell when you undergo gene therapy. And in the dog models that have had inhibitors who have undergone gene therapy, that has not been the case. They've had the typical type of transaminitis in the early period that's been steroid responsive.

DR. LOZIER: Do we have other questions? Yes, Dr. Pipe?

DR. PIPE: Steve Pipe from the University of Michigan. My question is for Dr. Heller, how the timeline for the evolution of a pathologic event like hepatocellular carcinoma would influence

the approach to surveillance.

If we're talking something that would be a 30- to 50-year timeline, something like that could never inform the current therapeutics that we're using today. So even if we pursued a pattern of surveillance, by the time we actually got an answer, we almost certainly wouldn't be using the current therapeutics that we are today.

So is there a window of time -- and I wouldn't limit this just to hepatocellular carcinoma. I would just take the data on a multiplicity of integration events and whatever pathologies could come from that.

Does there have to be some sort of practical timeline for which events have to happen for a focused surveillance program to really produce something that is really actionable?

DR. HELLER: I think there should be a timeline. If it's 50 years, that would be great, for something adverse to happen?

DR. PIPE: We're talking about bringing regulatory programs before a regulatory review, and

then also at a community level making decisions about embracing gene therapy. And I guess my fundamental question is, in what window of time would we have to find a pathology in order for it to actually inform what we're currently doing today?

We already have gene therapy programs that are 8 years out in humans. You mentioned some of the dogs. I mean, as far as we know, all the dogs that have undergone gene therapy have died of old age or have been put down because of old age with no known pathologies from integration events.

If we're going to impose postmarketing surveillance on gene therapy programs, does there have to be some window of time where these events have to occur? Or else it's just not going to be useful. How could it possibly change the course of what we're doing if it doesn't occur within a certain window of time?

DR. HELLER: Yes. So on my second-to-last slide, I had the word "time." I agree the time to develop something is important. If you don't see

it within 10 years, and by then, as you mentioned, technologies will be completely different -- this is something we'd answer as -- I'm not on any regulatory committee, and I'm not making any decisions. I'm a hepatologist. I would imagine that would be really important and would change your approach to surveillance, and you'd be far less concerned in humans if that's the pattern that you saw. But until you have some data, are we reluctant to say there's no concern?

DR. PIPE: And actually, I will get back to one of your points you made in your slide. You indicated that do-nothing was not an option, and I guess it depends on what the do-nothing is.

So we have longitudinal close follow-up, at least by our measures in hemophilia, through the comprehensive hemophilia treatment center programs, which have been in place for decades. That already is a mechanism of surveillance in our population. It's how we identify when new things that were unexpected occur in our population of patients, and then we can determine what actions are appropriate.

I wonder if -- it's not just gene therapy, but all of the new therapies we talked about today, it's very difficult to impose some sort of a window of postmarketing surveillance that is likely to capture all potential pathologies that could come from this paradigm shift.

It may be that it's not actually doing nothing, meaning that we're not doing regular ultrasounds, et cetera. But it's at least something, that if these patients maintain engagement through what we call surveillance systems in our hemophilia treatment centers, which will continue hopefully in perpetuity, that's at least something and it's more than nothing.

DR. HELLER: I would say that that's not nothing.

DR. PIPE: Yes.

DR. HELLER: I would strongly argue that that's a very active process. Someone has to maintain that database. It costs money.

DR. PIPE: Yes.

DR. HELLER: It takes effort. Someone's

funding that, and patients are actively taking part in that. That's surveillance at the highest level. And if you're saying it goes on to perpetuity, that's incredible. So you're very actively surveying your patients in every single way; then you agree with me.

DR. PIPE: I do to a point. It's whether we're going to --

DR. HELLER: Do an MRI or put them in a database?

DR. PIPE: Exactly.

DR. HELLER: I understand what you're asking.

DR. PIPE: And we could cherry-pick assays, which may or may not be relevant.

DR. HELLER: Yes. So you could argue you'd come up with something that for the first 5 years, we'll do ultrasounds and the first 10 years, we'll do MRIs. I don't know. I was careful not to come down one way or the other. You can make that argument, and at a certain point stop, and then just follow your database.

That's reasonable. I think that in hepatitis B, you start to see cancers in childhood and people who were neonatally infected horizontally from their parents. So I think that if you have a reasonable window of time, which is a separate discussion, you can then say let's put them in this active surveillance in every single way, which has been carefully considered and adapt it to what we find, I think that's a very reasonable approach.

DR. PIPER Thanks.

DR. HELLER: I wouldn't argue with that.

DR. LOZIER: So this is the regulatory conundrum. We have products with a lifecycle, and we're talking about kids or older kids. And maybe if we follow the adults for 10 years, we're not going to use that vector. And that's the problem.

I think I would be very nervous about -- and this is just my own personal; this is not an FDA-approved opinion. But it seems reasonable not to think about AAV gene therapy for young children. And you can define that as

whatever you want; less than 4, less than 6. If you're 17 or 18, maybe that's a different discussion altogether.

But that's the problem we have. We do have people saying let's go do gene therapy in the older adolescents. And as you might guess, the number of patients available for adults who are willing to participate in a trial who aren't on 3 other trials already; there are not very many patients. They're not out there in droves, waiting to sign up for things.

So that's our problem. That's why we have these workshops, to discuss some of this.

I think, at this point, we can move to the wrap-up. We're running over time, but we don't have to spend the entire allotted time for the wrap-up. Ann Farrell couldn't be here, so Lori Ehrlich was going to come up and take her place.

Thanks to our speakers.

(Applause.)

Wrap Up

DR. LOZIER: I'm asked to make an

announcement that you should find a video replay of this conference, along with the speaker's presentations, in about two weeks on the workshop webpage.

I've been taking notes, and I have probably 40 or 50 slides here of things. We're not going to read through them all, but I just think we could sort of recapitulate some of the things that came out of the different sessions.

Since I've been taking the notes and you're filling on short notice for Ann, I can sort of lead this, and you can stop me if you see something that interest you.

I think, certainly, from Dr. Ragni's overview, we saw that newly approved drugs such as emicizumab offered the advantage of non-intravenous injection and infrequent dosing compared to standard factor treatment over conventional factor treatment with or without inhibitors.

Fitusiran and gene therapy, which are treatments in development, offer novel alternative pathways to hemostasis or at least a one-time

treatment in the case of gene therapy. And the cost of these treatments will all be high, but the cost of treating hemophilia by standard care is also high to start with.

We have to worry, as the FDA, about long-term toxicity, drug interactions, and particularly about hepatotoxicity, because the liver is our favorite organ, at least in hemophilia.

For session 2, I think Dr. Montgomery's talk was particularly critical because it pointed out what I would call the physiology of factor VIII, not just the synthesis. It's made, and it has a certain length, and it interacts with factor IX, but where is it stored; how is it released?

I think it does lead a little bit into the question of, the factor level associated with replacement therapy or gene therapy when it's made in a non-endothelial cell, is that going to have equal hemostatic efficacy to somebody with mild hemophilia who may have a mutation but has normal

stores of factor VIII that can at least translate increase under stress? I think that's an open question, but one we have to think about.

It's convenient that factor IX is normally made in hepatocyte, but we are talking about novel variants such as the Padua that has about an 8- or 9-fold specific activity increase over the wild type where we have other issues.

I think it's also important that there is the interaction with von Willebrand factor and collagen in the subendothelial matrix, where there may be, if not reserves, at least a local concentration of factor IX that occurs at the side of vascular disruption.

I think Dr. Manco-Johnson's discussion and presentation -- I think the analogy between the CRPR of oncology is actually an interesting one. What we would hope for in hemophilia is, just as somebody with cancer would hope for total eradication of a disease and all of its associated pathologies, we would hope with gene therapy or novel treatments, whether it's emicizumab, or

fitusiran, or any other product developed in either center, that we would not only restore a factor level, but also at some point prevent any joint damage.

I think earlier is better, but we have the conundrum and we don't want to take the current gene therapy approaches into young children. So the charge to the hemophilia providers is to take care of these kids with the best treatment you can. Preserve their joints until they can sign up for a trial at age 18, or 16, or whatever we decide is a reasonable thing to do.

I was struck by recent presentations at ASH talking about biomarkers relating to bone destruction and collagen markers that could be perhaps followed. It's speculative to say whether that's a necessarily useful thing that we will be asking people to do, but it's something to be thought of.

I think the subclinical bleeding is a major problem. It's interesting to see that ultrasound seems to be adopted by most of the hemophilia

providers much more. It was really unheard of when I was at UNC during training there. But it sounds like many of the providers are doing this on a very regular basis.

Then we get into discussions of what should be the trough levels, and this has obviously evolved over time. When I was writing papers about gene therapy, again, it was 1 percent and we've got something to hold on to and something to offer. Now, we would just say that's just not worth discussing.

Over time, the debate has shifted, in part facilitated by the fact that the vectors and the constructs in the gene delivery systems are so much better now. We're even now worrying about having supratherapeutic factor VIII levels, which is a good place to be in.

But I think the problem then comes back to the kinds of issues that we saw in session 4 about the factor activity assays because, at the FDA, eventually, we help sponsors write a package insert or label -- and there won't be a package insert in

a bottle, I don't think, but there will be a major instruction manual that goes with these products.

So the question is, how much vector do you get to get what target dose without getting too much? And I think, an interesting question is if we target 100 percent and we're getting some people at 200 percent because of variations in just the interpatient response to the vectors and then the question of the assays, we worry, then, will we have a problem where we are promoting thrombosis, at least in the long run? Because people in the highest deciles of factor VIII or factor IX are in increased risk for thrombosis.

We never thought we'd have to worry about that problem 15, 20 years ago, but that's of concern. And that's part of the issue with the factor assay discrepancy question that we have to think about, is if they're within 20, 30 percent, we really shouldn't bump up against any ceiling. I think, as Dr. Pipe says, it's much more important, what these troughs are, because troughs are what kill you.

I guess maybe a question I didn't want to pose at the time of the factor sessions, or the assay sessions was could we contemplate instead of looking at factor levels with 10 different standards and three different methods that all have to be cross-validated, just something to consider is whether some global assay for hemostasis like thrombin generation or old-timey things like thromboelastography could be considered.

I know that everybody says, "Not TEGs. Those are terrible," but thrombin generation might be something useful to think about. But we still have to work on getting these assays to the point where we think we know which is the right value, and particularly at these low levels.

I think, in the PRO session, I was particularly struck by the skepticism of many of the patients who fill out these PRO rating instruments about, well you know, maybe it's a bad day and I need to get out of here, or there's not enough time, or the question is not pertinent to my particular situation, or I have a joint and there's

not going to be any point in talking about pain in the replaced joint, that sort of thing.

Clearly, those instruments may need some work to make them more relevant to the hemophilia community. I think that's actually an interesting set of observations we had from our patients.

Regarding our last session, session 5 on the two main topics, when do we go to kids, kids being maybe older adolescents, and the question of what should we do about the theoretical risk for hepatocellular carcinoma, these are sort of our hardest questions as regulators.

Certainly, with going into kids, we have the ethical and regulatory question, but then there's a practical, are the 17-year-olds, 16-year-olds, are they practically adults? But do we know what the long-term outcome is going to be with respect to long-term toxicity, particularly hepatotoxicity and hepatocellular carcinoma?

This really is a question that makes it hard to know what to do. It makes it easy if you're talking about a 2-year-old, but if you're

talking about a 16-year-old, it's encouraging that we haven't seen hepatocellular carcinoma in any of the patients that had been treated with gene therapy, but I don't know that even 10 patients have been treated in all the trials. Somebody could look that up.

But we don't know what the risk is, and we don't know -- if we have no events out of a small denominator, it's very hard to set a risk rating, but that's something we have to bear with.

Lori, do you have any comments on any of the sessions? I'm sure you had some observations.

DR. EHRLICH: I think, instead of kind of rehashing each session, which I think Dr. Lozier did a good job of recapping all of those things, I just wanted to point out that a lot of these topics, we could have devoted a full day to or certainly a lot more time than we were able to devote to it. There were some questions, I know, on Slido that we weren't able to get to.

We hope to use this as a starting point for all of these issues, and kind of where can we go

from here, and how can we improve in hemophilia drug development. So we at the FDA look forward to further conversations with all of the stakeholders who are involved, where we can use the information that we learned today, and bring that back to our work, and hopefully some of your work as well, that we can improve the way that we are developing novel drugs in hemophilia.

Then lastly, I just want to thank everybody that was involved in the session, certainly the patients and the advocates that were able to come today and share their perspectives. I think they had an invaluable perspective on what we do here and where we're potentially missing the mark, and how we can improve things moving forward, but also the clinicians, and researchers, and commercial sponsors who were able to kind of come together and put forth some new ideas.

Adjournment

DR. LOZIER: I think we also need to thank Joan Todd and Valerie Vashio, who have been our support staff and have sent out thousands of

e-mails, literally, to many of the participants and kept the trains running on time here, and making sure that everything was arranged, and the people arranged travel. I also want to thank the Oncology Center of Excellence for sponsoring this workshop.

I think at this point, we can conclude, and everybody can try to catch their flights to get out of here. Thank you very much.

(Applause.)

(Whereupon, at 4:30 p.m., the meeting was adjourned.)