FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH

PEDIATRIC ONCOLOGY SUBCOMMITTEE OF THE

ONCOLOGIC DRUGS ADVISORY COMMITTEE (pedsODAC)

Virtual Meeting

Day 2

Thursday, May 12, 2022

10:00 a.m. to 2:52 p.m.

Meeting Roster

ACTING DESIGNATED FEDERAL OFFICER (Non-Voting)

Joyce Yu, PharmD

Division of Advisory Committee and

Consultant Management

Office of Executive Programs, CDER, FDA

ONCOLOGIC DRUGS ADVISORY COMMITTEE MEMBERS (Voting)

Mark R. Conaway, PhD

(Participation in Day 2 Only)

Professor and Director of Translational Research

Division of Translational Research and Applied

Statistics

Department of Public Health Sciences

University of Virginia

Charlottesville, Virginia

David E. Mitchell

(Consumer Representative)

Founder, Patients for Affordable Drugs

Bethesda, Maryland

Alberto S. Pappo, MD

(Chairperson, pedsODAC)

Member and Head, Division of Solid Malignancies

St Jude Children's Research Hospital

Professor of Pediatrics

University of Tennessee Health Science Center

Memphis, Tennessee

ACTING INDUSTRY REPRESENTATIVE TO THE COMMITTEE

(Non-Voting)

Albert L. Kraus, PhD

Global Regulatory Portfolio Lead, Oncology

Pfizer, Inc.

Guilford, Connecticut

TEMPORARY MEMBERS (Voting)

Rochelle Bagatell, MD

Professor of Pediatrics

Perelman School of Medicine, University of Pennsylvania

Solid Tumor Section Chief, Division of Oncology

Department of Pediatrics

The Children's Hospital of Philadelphia

Philadelphia, Pennsylvania

Steven G. DuBois, MD

Director, Experimental Therapeutics

Dana-Farber/Boston Children's Hospital

Associate Professor of Pediatrics

Harvard Medical School

Boston, Massachusetts

Ira J. Dunkel, MD

Professor of Pediatrics

Weill Cornell Medical College

Attending Pediatric Oncologist

Department of Pediatrics

Memorial Sloan Kettering Cancer Center

New York, New York

Julia Glade Bender, MD

Vice Chair for Clinical Research

Department of Pediatrics

Memorial Sloan Kettering Cancer Center

New York, New York

Richard Gorlick, MD

Division Head and Department Chair, Pediatrics

Professor of Pediatrics

H. Grant Taylor, M.D., W.W. Sutow, M.D. and

Margaret P. Sullivan, M.D. Distinguished Chair in

Pediatrics

Department Chair *ad interim*

Sarcoma Medical Oncology

University of Texas MD Anderson Cancer Center

Children's Cancer Hospital

Houston, Texas

AeRang Kim, MD, PhD

Director of Clinical Research

Division of Oncology

Children's National Hospital

Associate Professor of Pediatrics

The George Washington School of Medicine

Washington, District of Columbia

E. Anders Kolb, MD

Director, Nemours Center for Cancer and Blood Disorders
Nemours Children's Health
Wilmington, Delaware
Professor, Department of Pediatrics
Sidney Kimmel Medical College at
Thomas Jefferson University
Philadelphia, Pennsylvania

Theodore W. Laetsch, MD

Associate Professor of Pediatrics
University of Pennsylvania/
Abramson Cancer Center
Director, Developmental Therapeutics and Very Rare Malignant Tumor Programs
Children's Hospital of Philadelphia
Philadelphia, Pennsylvania

Gianna McMillan, D.Be

(Patient Representative; Participation in Day 2 Only)

Los Angeles, California

D. Williams (Will) Parsons, MD, PhD

Associate Professor of Pediatrics

Baylor College of Medicine

Deputy Director, Texas Children's Cancer and Hematology Centers

Houston, Texas

Nita Seibel, MD

(Participation in Day 2 Only)

Head, Pediatric Solid Tumor Therapeutics

Clinical Investigations Branch

Cancer Therapy Evaluation Program

Division of Cancer Treatment and Diagnosis

National Cancer Institute (NCI)

National Institutes of Health (NIH)

Bethesda, Maryland

FDA PARTICIPANTS (Non-Voting)

Gregory H. Reaman, MD

Associate Director for Pediatric Oncology

Oncology Center of Excellence (OCE)

Office of the Commissioner (OC)

Associate Director for Pediatric Oncology

Office of Oncologic Diseases (OOD)

Office of New Drugs (OND), CDER, FDA

Martha Donoghue, MD

Acting Associate Director for Pediatric and Rare Cancer Drug Development

OCE, OC

Deputy Director, Division of Oncology 2 (DO2)

OOD, OND, CDER, FDA

Diana Bradford, MD

(Participation in Day 2 Only)

Cross-Discipline Team Leader

DO2, OOD, OND, CDER, FDA

Anup Amatya, PhD

(Participation in Day 2 Only)

Acting Lead Mathematical Statistician

Division of Biometrics V

Office of Biostatistics

Office of Translational Sciences, CDER, FDA

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)

P R O C E E D I N G S

(10:00 a.m.)

Call to Order

DR. PAPPO: Well, welcome to day 2. I hope you all had some rest and a nice dinner.

Good morning and welcome. I would first like to remind everyone to please mute your line when you are not speaking. For media and press, the FDA press contact is Chanapa Tantibanchachai, and her email and phone are currently displayed.

My name is Alberto Pappo, and I will be chairing today's meeting. I will now call the May 12, 2022 meeting of the Pediatric Oncology Subcommittee of the Oncologic Drugs Advisory Committee to order. Dr. Joyce Yu is the acting designated federal officer for this meeting and will begin with introductions.

Introduction of Subcommittee

DR. YU: Good morning. My name is Joyce Yu, and I am the acting designated federal officer for this meeting. When I call your name, please introduce yourself by stating your name and

affiliation.

We'll start with Dr. Conaway.

DR. CONAWAY: Mark Conaway, University of Virginia.

DR. YU: Mr. Mitchell?

MR. MITCHELL: I'm David Mitchell. I am the consumer representative to the ODAC. I am president of Patients for Affordable Drugs, and I am a multiple myeloma patient.

DR. YU: Thank you.

I just want to remind everyone to please keep your line muted when you're not speaking.

Dr. Pappo?

DR. PAPPO: Good morning. I'm Alberto Pappo. I'm a pediatric oncologist at St. Jude Children's Research Hospital, and I'm the chairperson for the Pediatric ODAC.

DR. YU: Dr. Bagatell?

DR. BAGATELL: Hi. My name is Ro Bagatell. I'm a pediatric oncologist at the Children's Hospital of Philadelphia.

DR. YU: Dr. DuBois?

DR. DuBOIS: Hi. This is Steve DuBois, a pediatric oncologist at Dana-Farber Boston Children's.

DR. YU: Dr. Dunkel?

DR. DUNKEL: Good morning. This is Ira Dunkel. I'm a pediatric neuro-oncologist at Memorial Sloan Kettering Cancer Center.

DR. YU: Dr. Glade Bender, please.

DR. GLADE BENDER: Good morning. I'm Julia Glade Bender. I am a pediatric oncologist also at Memorial Sloan Kettering Cancer Center in New York.

DR. YU: Dr. Gorlick?

DR. GORLICK: Good morning. I'm Richard Gorlick. I'm a pediatric oncologist at MD Anderson Cancer Center in Houston, Texas.

DR. YU: Thank you, Dr. Gorlick.

My apologies, again. I just want to remind all of our participants today to please be mindful of the advancing of the slides. Thank you.

Dr. Kim, please.

DR. KIM: Hi. This is AeRang Kim from Children's National in DC. I'm a pediatric

oncologist there.

DR. YU: Dr. Kolb?

DR. KOLB: Yes. Hi. Andy Kolb. I'm a pediatric hematologist/oncologist at Nemours Children's Health.

DR. YU: Dr. Laetsch?

DR. LAETSCH: Hi. I'm Ted Laetsch. I'm a pediatric oncologist at Children's Hospital Philadelphia at University of Pennsylvania.

DR. YU: Thank you.

Dr. Laetsch, your audio is a bit low on my end. Could you introduce yourself one more time, please?

DR. LAETSCH: Sure. Hi. Is this better?

DR. YU: Yes.

DR. LAETSCH: Hi. I'm Ted Laetsch. I'm a pediatric oncologist at the Children's Hospital of Philadelphia at University of Pennsylvania.

DR. YU: Thank you so much.

Dr. McMillan?

DR. McMILLAN: Good morning. I'm Gigi McMillan. I'm a bioethicist at Loyola Marymount

University in Los Angeles, and I'm a patient representative.

DR. YU: Thank you.

Dr. Parsons?

DR. PARSONS: Good morning. This is Will Parsons. I'm a pediatric oncologist at Texas Children's Hospital and Baylor College of Medicine in Houston, Texas.

DR. YU: Dr. Seibel?

DR. SEIBEL: Hi. Good morning. I'm Nita Seibel, pediatric oncologist in the Clinical Investigations Branch at CTEP.

DR. YU: Dr. Kraus?

DR. KRAUS: Good morning. Albert Kraus. I work in research and development, for decades, in oncology therapeutics. I'm currently with Pfizer Corporation, and I'm the industry representative. Thank you.

DR. YU: I'll now introduce our FDA participants for today, starting with Dr. Reaman.

DR. REAMAN: Good morning. I'm Greg Reaman, associate director for pediatric oncology in the

FDA's Oncology Center of Excellence and the Office of Oncologic Diseases in CDER.

DR. YU: Dr. Donoghue?

DR. DONOGHUE: Hi. Good morning. I'm a pediatric oncologist at the FDA, and I work in one of the review divisions that oversee the development of oncology products.

DR. YU: Dr. Bradford?

DR. BRADFORD: Good morning. My name is Diana Bradford. I'm a pediatric oncologist and cross-discipline team leader in the Division of Oncology 2 in CDER at FDA.

DR. YU: Dr. Amatya?

DR. AMATYA: Good morning. I'm Anup Amatya. I'm a statistician in Biometrics Division V at CDER.

DR. YU: Dr. Pappo, please?

DR. PAPPO: Thank you very much, Joyce.

For topics such as those discussed at this meeting, there are often a variety of opinions, some of which are quite strongly held. Our goal is that this meeting will be a fair and open forum for

discussion of these issues and that individuals can express their views without interruption. Thus, as a gentle reminder, individuals will be allowed to speak into the record only if recognized by the chairperson. We look forward to a productive meeting.

In the spirit of the Federal Advisory Committee Act and the Government in the Sunshine Act, we ask that the advisory committee members take care that their conversations about the topic at hand take place in the open forum of the meeting. We are aware that members of the media are anxious to speak with the FDA about these proceedings, however, the FDA will refrain from discussing the details of this meeting with the media until its conclusion. Also, the committee is reminded to please refrain from discussing the meeting topic during the break. Thank you.

Now Dr. Joyce Yu will read the Conflict of Interest Statement for the meeting.

Conflict of Interest Statement

DR. YU: The Food and Drug Administration,

FDA, is convening today's meeting of the Pediatric Oncology Subcommittee of the Oncologic Drugs Advisory Committee under the authority of the Federal Advisory Committee Act, FACA, of 1972. With the exception of the industry representative, all ODAC members and temporary members of the subcommittee are special government employees, SGEs, or regular federal employees from other agencies and are subject to federal conflict of interest laws and regulations.

The following information on the status of this subcommittee's compliance with federal ethics and conflict of interest laws, covered by but not limited to those found at 18 U.S.C. Section 208, is being provided to participants in today's meeting and to the public.

FDA has determined that ODAC members and temporary members of this subcommittee are in compliance with federal ethics and conflict of interest laws. Under 18 U.S.C. Section 208, Congress has authorized FDA to grant waivers to special government employees and regular federal

employees who have potential financial conflicts when it is determined that the agency's need for a special government employee's services outweighs his or her potential financial conflict of interest, or when the interest of a regular federal employee is not so substantial as to be deemed likely to affect the integrity of the services which the government may expect from the employee.

Related to the discussions of today's meeting, ODAC members and temporary members of this subcommittee have been screened for potential financial conflicts of their own as well as those imputed to them, including those of their spouses or minor children and, for purposes of 18 U.S.C. Section 208, their employers. These interests may include investments; consulting; expert witness testimony; contracts, grants, CRADAs; teaching, speaking, writing; patents and royalties; and primary employment.

Today's agenda involves consideration and discussion of the potential utility and steps to validation of an intermediate clinical endpoint,

response to induction therapy, in the development of new drugs for the first-line treatment of patients with high-risk neuroblastoma. The European Medicines Agency has also been invited to present.

This is a particular matters meeting during which general issues will be discussed. Based on the agenda for today's meeting and all financial interests reported by the ODAC members and temporary members of the subcommittee, no conflict of interest waivers have been issued in connection with this meeting. To ensure transparency, we encourage all ODAC members and temporary members of the subcommittee to disclose any public statements that they have made concerning the topic at issue.

With respect to FDA's invited industry representative, we would like to disclose that Dr. Albert Kraus is participating in this meeting as a non-voting industry representative, acting on behalf of regulated industry. Dr. Kraus' role at this meeting is to represent industry in general and not any particular company. Dr. Kraus is

employed by Pfizer.

With regard to FDA's guest speakers, the agency has determined that the information to be provided by these speakers is essential. The following guest speakers have reported interests which are being made public to allow the audience to objectively evaluate any presentation and/or comments made by the speakers.

Dr. Dominik Karres has acknowledged that he is employed by the European Medicines Agency, EMA. Dr. Navin Pinto has acknowledged that he is an unpaid scientific advisor for Y-Mabs Therapetics. Dr. Maja Beck Popovic has acknowledged that she is employed by the University Hospital in Lausanne, Switzerland. As guest speakers, Dr. Karres, Pinto, Beck Popovic, and Ms. Knox will not participate in subcommittee deliberations, nor will they vote.

We would like to remind ODAC members and temporary members of the subcommittee that if the discussions involve any other topics not already on the agenda for which an FDA participant has a personal or imputed financial interest, the

participants need to exclude themselves from such involvement, and their exclusion will be noted for the record. FDA encourages all participants to advise the committee of any financial relationships that they may have regarding the topic that could be affected by the subcommittee's discussions. Thank you.

DR. PAPPO: Thank you very much, Joyce.

We will now proceed with the FDA introductory remarks.

DR. DONOGHUE: Thank you, Dr. Pappo. I think Dr. Reaman may want to start things out.

Dr. Reaman?

DR. REAMAN: Thanks, Martha.

FDA Introductory Remarks – Gregory Reaman

DR. REAMAN: Good morning. This is Greg Reaman, and I again want to welcome the members of the committee to the second day of our advisory committee meeting, and a special welcome to those of you who weren't here yesterday, and just to remind you of the importance of the discussion and the deliberations as it relates to informing FDA

regulatory decision making. So we very much appreciate the time and effort that you're putting into this.

I'd like to, again, acknowledge, in the spirit of international collaboration, a special welcome to my colleague, Dominik Karres, from the Pediatric Medicines Office at the European Medicines Agency, and a special welcome to our European patient advocates and investigators who will participate as presenters in this session.

With that, back to you, Dr. Donoghue. Thank you.

DR. DONOGHUE: Thanks, Greg.

I'd like to echo Dr. Reaman's welcome, and thank all of those on the committee and guest speakers for devoting their expertise and time to today's meeting, which will focus on important topics relevant to the development of drugs for the treatment of patients with high-risk neuroblastoma.

I'd also like to extend a warm welcome to all stakeholders who are attending today's session for your interest, and being present here to help

discuss how we can align and collaborate together to help advance treatment of pediatric patients with cancer. We at the FDA value very much your collaboration and support.

The topics of today's session is a bit of a shift in focus compared to yesterday's higher level broad discussion, which was aimed at developing a framework to inform FDA decision making regarding pediatric development plans when there are multiple same-in-class molecularly targeted products. We're here today because we all recognize that children with high-risk neuroblastoma have a high unmet medical need, and we have a vested interest in working together to develop new treatments that are safe and effective for pediatric patients with cancer as efficiently as possible.

Together today, we'll consider and discuss the current use and potential future validation of a biomarker, which we also refer to sometimes as an early or intermediate clinical endpoint, end-of-induction response, for clinical decision making and development of new drugs for the

treatment of patients with high-risk neuroblastoma, with a focus on the frontline setting.

We will hear perspectives from a variety of stakeholders from the U.S. and abroad. First, we'll hear from Drs. Bradford and Karres, who will provide regulatory insights on the current treatment approaches and ongoing efforts in the development of new drugs for the first-line treatment of pediatric patients with high-risk neuroblastoma.

After that, we will hear important perspectives from Ms. Leona Knox, who's a patient advocate, who has research at Solving Kids' Cancer in the UK, as well as Drs. Pinto and Beck Popovic, who conduct research in high-risk neuroblastoma, and will provide their thoughts on how to best develop new treatments for these patients, including their perspective on use of end-of-induction response in patient care and drug development.

After some clarifying questions and a break for lunch, we'll shift gears a bit and hear

perspectives from statistical colleagues from the National Cancer Institute and the FDA, Drs. McShane and Dr. Amatya, respectively, on how we might formulate a path forward to validate end-of-induction response as an early endpoint for assessment of investigational drugs developed for patients with high-risk neuroblastoma.

Lastly, after the open public hearing, we look forward to a robust discussion on use of early clinical endpoints for the development, in general, in pediatric neuroblastoma, and in particular, the current strength of evidence for use of end-of-induction response, how it's being used currently for clinical decision making and trial conduct, as well as future steps to validation of this endpoint, if warranted.

Thank you very much again for your attention and input, and I look forward to a fruitful discussion, and I will turn the podium now over to Dr. Diana Bradford.

DR. PAPPO: Thank you, Dr. Donoghue and Dr. Reaman.

We will now proceed with an FDA and guess presentation, starting with Dr. Diana Bradford, followed by Dr. Dominik Karres.

FDA Presentation - Diana Bradford

DR. BRADFORD: Thank you.

Good morning, everyone. Again, my name is Diana Bradford. This morning I'll briefly provide some background for today's discussion, including the approach to initial treatment of patients with high-risk neuroblastoma, current trials and investigational strategies, and finally highlight some aspects of prior FDA approvals in this disease space, before turning the floor over to Dr. Dominik Karres from the European Medicines Agency.

Neuroblastoma is the most common extracranial solid tumor in pediatric patients with approximately 650 new cases per year in the U.S. This is primarily a disease of young children with a median age of diagnosis of 19 months, and 90 percent of all diagnoses of neuroblastoma occurring by 5 years of age.

This is also a very heterogeneous disease

with variable clinical presentations and biology, including molecular characteristics, and ultimately prognosis. Risk groups are based on patient age, stage, and molecular and histological characteristics of the tumor, and are used to determine appropriate treatment.

While some patients may require only surgical resection or even observation, patients with high-risk disease require intensive multimodality therapy. These patients are the focus of today's discussion.

Even with this intensive therapy, patients with high-risk neuroblastoma have a 40 to 50 percent chance of long-term survival, and survivors may have substantial long-term effects from their cancer therapy. At relapse, there are few treatment options, and patients face a very poor prognosis. Patients with high-risk neuroblastoma have an unmet medical need and additional therapeutic options are needed.

The general approach to the initial treatment of high-risk neuroblastoma is similar

between the U.S. and Europe. Treatment includes induction, consisting of multiple cycles of chemotherapy with surgical resection; consolidation, consisting of myeloablative chemotherapy with autologous stem-cell transplant with radiation therapy to primary and metastatic sites; and post-consolidation with anti-GD2 therapy, GM-CSF, and isotretinoin.

The specific number, frequency, and composition of induction cycles differ between the U.S. and Europe, as does the use of tandem transplantation. The specific anti-GD2 antibody also differs between the U.S. and Europe.

Of course, in pediatric oncology there is substantial participation in clinical trials, and one cannot describe the approach to treatment without discussing ongoing cooperative group trials, such as such ANBL 1531.

This is the current children's oncology group COG trial for high-risk neuroblastoma. This is a randomized trial consisting of five arms, with eligibility for different arms determined by MIBG

avidity and ALK positivity. Patients with ALK-positive tumors and those with ALK-negative MIBG-negative tumors are non-randomly assigned, and patients with MIBG-positive/ALK-negative tumors are randomized to Arms A or B. The initially opened Arm C is now closed to accrual.

This trial is evaluating primarily the addition of MIE [ph] therapy to induction and the addition of ALK inhibition throughout first-line therapy for patients with ALK aberrant neuroblastoma. The primary endpoint of this trial is event-free survival or EFS.

In Europe, the HR-NBL2 trial is ongoing and is investigating different approaches to induction, consolidation, radiation, and administration. The primary endpoint of the HR-NBL2 trial is also event-free survival. Thank you to Dr. Lucas Moreno for sharing this slide.

In the induction, patients will be randomized to the rapid COJEC versus German pediatric oncology/hematology, or GPOH, regimen; and consolidation patients will be randomized to a

single transplant or tandem transplant; and finally, for patients with residual disease, randomization will evaluate the addition of a boost to residual tumor, in addition to the standard dose of radiation to the preoperative tumor bed. Patients with poor response to induction chemotherapy may enroll in the VERITAS trial, as shown here, which is a randomized trial of MIBG therapy plus autologous transplant versus tandem transplant.

In addition to the investigational age described in the prior slide, there's interest in adding agents to first-line therapy to improve outcomes in the EU, as we have seen in U.S. trials. The addition of ALK inhibition to frontline, high-risk therapy in the EU is forthcoming, and there's also future interest in augmenting initial therapy with chemoimmunotherapy.

I'll turn now to an example of a development program leading to FDA approval in high-risk neuroblastoma to illustrate some regulatory and development considerations. Dinutuximab was

approved in 2015 in combination with GM-CSF, IL-2, and 13-cis-retinoic acid for the treatment of pediatric patients with high-risk neuroblastoma in the frontline maintenance setting, who achieved at least a partial response to prior first-line multiagent, multimodality therapy. This was the first drug approved specifically for patients with high-risk neuroblastoma.

The basis for the approval was COG study ANBL 0032. This is a randomized, open-label, multicenter trial conducted in pediatric patients with high-risk neuroblastoma. All patients had received prior therapy consisting of induction chemotherapy; maximum feasible surgical resection; myeloablative consolidation chemotherapy followed by transplant; and radiation therapy to residual soft-tissue disease.

Patients were randomized between days 50 and 77 post-transplant. They were required to have achieved at least a partial response prior to transplant and have no evidence of disease progression following completion of frontline

therapy. A total of 226 patients were randomized. The major efficacy outcome measure was investigator-assessed event-free survival.

The study demonstrated a clinically meaningful and statistically significant improvement in event-free survival in patients randomized post-consolidation to receive dinutuximab plus IL 2, GM-CSF, and isotretinoin versus isotretinoin alone, with results shown here. The EFS hazard ratio was 0.57, the confidence interval shown here, and a p-value of 0.01. EFS results were supported by a trend in improvement in overall survival.

I wanted to highlight the timeline for the dinutuximab development program to illustrate some challenges. Anti-GD2 antibodies were initially evaluated in neuroblastoma in the 1990s. The original IND submission for dinutuximab was submitted to the FDA in 1991. Following the opening of ANBL 0032, due to the relevant rarity of high-risk neuroblastoma, it took 7 years to accrue the requisite number of patients on the randomized

portion of the study.

Randomization was stopped in 2009, after which a cooperative research and development agreement, or CRADA, was established between NCI and the commercial sponsor, United Therapeutics. Subsequent steps to establish manufacturing and comparability of products were needed, which were time-intensive. The BLA was approved in 2015.

The timeline for the development of dinutuximab, spanning more than two decades, can provide some insight and interest in earlier endpoints, as well as to the need to consider the potential for commercial development early in investigation to avoid delays in drug development.

Two products have been approved specifically for patients with high-risk neuroblastoma in the last 10 years, including one in the first-line setting, and these cases highlight some regulatory considerations for development to neuroblastoma. As just discussed, dinutuximab was approved on the basis of improvement in EFS, supported by a trend in improvement in overall survival. EFS and OS are

difficult to interpret in the absence of randomization. Here, randomization allowed isolation of the treatment effect of dinutuximab, plus GM-CSF, and Il-2.

Understanding the treatment effect of one component of treatment, as well as the contribution of each component within multimodality therapy, is one of many challenges in considering trial design in this disease space. Relapse and refractory disease is not the focus of today's discussion, but to briefly mention the recent approval in this disease.

Naxitamab is approved in combination with GM-CSF for the treatment of pediatric patients with relapsed or refractory neuroblastoma in the bone or bone marrow, who have demonstrated a partial response, minor response, or stable disease following prior therapy. It was approved on the basis of overall response rate as assessed by blinded independent review.

The FDA considered that a randomized trial in this setting could be challenging given the lack

of approved therapies, which might serve as a comparator, and given the rarity of the disease. Overall response rate may be an appropriate endpoint when responses can be objectively measured and may be used to support an approval. If the overall response rate is substantial in the context of available therapies, the duration of response is substantial, and together these can be considered likely to be predictive of clinical benefit.

I think it is important to point out the unique development considerations in pediatric oncology as illustrated through these approvals. Investigation of both these products were initiated not by pharmaceutical companies but by academic investigators and cooperative groups.

To briefly summarize, patients with high-risk neuroblastoma have a high unmet medical need. Few drugs have been approved specifically for patients with high-risk neuroblastoma, and there is a need for improved outcomes. Recognizing this, and the worst prognosis, and limited options upon relapse, the exploration of additions to

frontline therapy are the focus of both U.S. and EU frontline trials, including the additions of drugs to induction regimens such as the addition of ALK inhibition for patients with ALK-positive tumors and the addition of MIBG therapy early in therapy in the COG study, ANBL 1531.

In Europe, changes to first-line therapy are also being evaluated. ALK inhibition is being explored in the first-line setting as well, with future interest in the addition of immunotherapy earlier in therapy. Prior approvals in this space have been founded on trials initially driven by the research of the pediatric oncology academic and cooperative group community with involvement of pharmaceutical companies at various points in development. Investigators, pharmaceutical companies, regulatory agencies, and patients all bring unique perspectives and insights to this process, highlighting the need for early multistakeholder collaboration to bring new therapies to patients as efficiently as possible.

Drug development in high-risk neuroblastoma

faces several challenges, including designing trials adequately to isolate the treatment effect of a given therapy and the time required to conduct larger trials given the rarity of the disease. Survival-based endpoints, specifically EFS, have been used to support approval in the first-line setting.

If one considers potential delays as observed with the dinutuximab development program, in addition to a long timeline based on accrual rate, it is easy to understand the interest in exploring the use of intermediate clinical endpoints such as end-of-induction response to support drug development.

As we will be discussing at length today, and depending upon the extended validation, intermediate clinical endpoints may permit earlier assessment of efficacy of a given treatment but also can be used in other ways to inform development.

Now, I will turn to my colleague, Dr. Dominik Karres at EMA, for his thoughts on the

development of intermediate endpoints. Thank you for your attention.

Guest Speaker Presentation – Dominik Karres

DR. KARRES: Thank you very much, Dr. Bradford, and thank you very much again for the invitation an opportunity to provide a general EMA perspective on the development and utility of intermediate endpoints such as end-of-induction response to support drug development considerations for the treatment of patients with high-risk neuroblastoma. This is my usual disclaimer.

I would like to start re-emphasizing that regulatory approval of a new drug is based on the robustness of evidence, demonstrating clinical benefit, for example, by means of clinically meaningful survival improvements balanced against identified risks. This includes the need for considerations related to the actual individual contribution of a new drug to benefit and risks, and context of its use within multimodal treatment regimens.

With that in mind, mature data generation

with the objective to informing benefit-risk considerations in front-line high-risk neuroblastoma takes time, as we all know and have experienced, and it's not even feasible for all potentially available suitable novel agents.

Appreciating now the high unmet medical need, as we've heard, taking into account the prognosis, limited treatment options at relapse, but also toxicities of current treatments, there's a clear need, as mentioned by Dr. Bradford, for early multistakeholder collaboration to finding new ways to timely bring novel agents to patients with newly diagnosed high-risk neuroblastoma, for example, by means of intermediate endpoint considerations to support guiding decision making and priorities to accelerate drug development in the interest of the patient.

I will now very generally reflect on what potential purposes an intermediate endpoint like end-of-induction response could have, all having its value potentially able to supporting regulatory decision making, and I'm sure subsequent speakers

will provide more detailed insights later today.

Depending on the level of evidence available and needed to justify its use, this could, for example, be to support guiding patient care as considered in the SIOPEN study, where parents with poor response to induction chemotherapy may enroll in the VERITAS trial, as described earlier by Dr. Bradford; or guiding prioritization discussions where one could see the value of an early assessment of efficacy guiding further contextualized development discussions potentially supporting go or no-go decisions; and lastly, to ultimately serving as a validated surrogate endpoint in a pivotal clinical trial, meaning available evidence being strong enough, showing a proven prognostic relationship between end-of-induction response and the clinical outcome of its survival; allowing to support regulatory benefit, a benefit-risk decision making as outlined on my previous slide; and appreciating here the necessary regulatory validation steps in this case requires a high level of convincing evidence and

scrutiny prior to agreement of its use. And again, I'm sure will hear about more on that point later today.

To conclude my short presentation and reflections, end-of-induction response in patients with high-risk neuroblastoma may have potential utility in forming and accelerating drug development efforts, so I'm very much looking forward to today's discussion and would like to thank the FDA again for the opportunity to participate. But for such an endeavor to be successful and to eventually benefiting patients, international multistakeholder collaboration and early interactions with the regulators is key, I believe, to ensuring that all available evidence can be independently reviewed to supporting a proposed intended use of such an endpoint within regulatory submissions.

Having said that, and with the focus on EMA procedures, I would like to take this opportunity to inviting the academic community to come forward, considering EMA qualification advice in that regard

as a necessary step in moving this discussion forward from a European perspective. We would be more than happy to guiding and supporting you in the necessary preparatory activities with the objective to, of course, continuing the discussion and continuous close collaboration with the FDA, ensuring that it benefits patients on both sides of the Atlantic.

That concludes my part of the presentation, and I would like to thank you very much again.

Clarifying Questions

DR. PAPPO: Thank you very much, Dr. Bradford and Dr. Karres.

We have about 20 minutes for questions, so we will now take clarifying questions for Dr. Bradford and Karres. Please use the raise-hand icon to indicate that you have a question, and remember to clear the icon after you have asked a question. When acknowledged, please remember to state your name for the record before you speak and direct your question to a specific presenter, if you can.

If you wish for a specific slide to be displayed, please let us know the slide number if possible. Finally, it would be helpful to acknowledge the end of your question with a thank you and end of your follow-up question with, "That is all for my questions," so we can move on to the next panel member.

We are now open for questions. I see Dr. McMillan.

DR. McMILLAN: Yes. Thank you. This is Dr. Gigi McMillan from Los Angeles.

Dr. Karres, you mentioned that you thought independent verification of early endpoints would be needed for there to be enough evidence to use them for making decisions about trial progression. Can you elaborate a little bit on that?

DR. KARRES: This is Dominik Karres. Thanks a lot for the question.

Indeed, what I refer to here is with a perspective on European regulatory requirements with regard to pediatric investigations plans as an example, ensuring that any decisions with regard to

novel agent introduction into frontline treatment, and then considerations with regard to which product to continue moving forward into full development in that indication, would certainly require some discussions in terms of understanding, from our side; and considerations, what would be considered acceptable threshold levels in terms of responses seen for individual products; if that answers your question. Thank you.

DR. McMILLAN: Yes. Thank you very much.

DR. PAPPO: Any additional questions?

I had a question, and maybe this is a little bit for later because I know that we're going to have some talks by statisticians. But are there any thoughts to how this will affect protocol design if you develop a new endpoint, and how you're going to interpret data with overall survival, and PFS, and EFS, and if you are going to take patients off protocol or give them an alternative regimen if they don't have a CR or PR at the end of induction, or should we leave that for later, for statistical discussion?

DR. BRADFORD: This is Diana Bradford. I think it may be helpful to have a little bit more discussion from the statisticians before we delve into that question. If others from FDA feel differently, we could do it I guess now.

DR. PAPPO: Okay. We'll just wait for the statistical presentations later in the day. Thank you.

Anybody else have any questions that they would like to ask Dr. Bradford or Dr. Karres before we move to the next guest presentations?

Dr. Seibel?

DR. SEIBEL: Yes. Perhaps Dr. Bradford could provide a little bit more detail about the VERITAS trial that patients in Europe will go to if they have an inadequate response to induction.

DR. BRADFORD: Yes. This is Diana Bradford. I may turn to Dr. Karres for his input, as I only have a very high level understanding of this trial, and my understanding it's open to patients with poor response to induction chemotherapy, such as on the HR-NBL2 trial, and they would then proceed to a

randomized trial of MIBG therapy plus autologous transplant compared to arm that's tandem transplantation.

Dr. Karres, do you have any further comments?

DR. KARRES: This is Dominik Karres. No, nothing in addition to add to that. You have summarized the concept of that study. Thank you.

DR. McMILLAN: Could you define what -- I know this may be difficult -- is considered poor response?

DR. BRADFORD: This is Diana Bradford. I'm sorry. I don't have that information on hand --

DR. McMILLAN: Sure.

DR. BRADFORD: -- but I can certainly find out.

DR. McMILLAN: Okay. Thank you.

DR. PAPPO: At the time of enrollment of this trial, patients would have not seen or received any kind of GD2 antibody, correct?

DR. BRADFORD: This is Diana Bradford again. That is my understanding.

DR. PAPPO: Thank you very much.

Any other questions? If you have had your question answered, please be sure to put your hand down.

We have a question from Dr. Glade Bender.

DR. GLADE BENDER: Hi. Julia Glade Bender from Memorial Sloan Kettering. You mentioned that the anti-GD2 naxitamab map was approved on the basis of overall response rate, but this was in a relapsed population.

I was just wondering if either representatives from the FDA or the EMA could comment on the context of an upfront study versus a relapsed trial and how endpoints might be viewed differently, depending on where the patient is in their disease trajectory.

DR. BRADFORD: Yes. This is Diana Bradford. I think the ability to support a regulatory approval on the basis of overall response rate is a very complex question. It depends on many factors. Oftentimes this is considered in the relapsed or refractory setting, and the preference for upfront

therapy would be to demonstrate an improvement on a survival-based endpoint, traditionally.

There are situations where in rare diseases, with strong biological rationale based on the mechanism of action of the drug, the molecular defined subset, for example, frontline indications have been granted for some of the targeted agents, based on an overall response rate endpoint supported, of course, by very important information on the duration of responses.

I'll turn to Dr. Karres for any of his perspective on this as well.

DR. KARRES: Thank you very much. This is Dominik Karres. Indeed, the situation in Europe is similar, that depending obviously on the intended target population -- in the context of alternative treatments, unmet medical needs, et cetera -- a single-arm study based on a primary endpoint of overall response rate, supported through duration of response, has been accepted in the past; that the main issue always relates to the ability to attribute any treatment effect seen to the

individual compound under investigation. But indeed, in a frontline setting, an event-driven endpoint such as event-free survival in a randomized fashion would certainly be preferred from a regulatory perspective. Thank you.

DR. GLADE BENDER: Thank you. Those comments were very helpful.

DR. PAPPO: Does that answer your question, Julia?

DR. GLADE BENDER: Yes. I think those were very helpful comments, and my question was answered.

DR. PAPPO: Any additional questions?

(No response.)

DR. PAPPO: Also, Nita, I believe that your hand is still up, so if you want to put it down.

(No response.)

DR. PAPPO: If there are no additional questions for Dr. Bradford and Karres, we will now proceed with guest speaker presentations, starting with Ms. Leona Knox, and this will be followed by Dr. Navin Pinto and Dr. Maja Beck Popovic.

Guest Speaker Presentation – Leona Knox

MS. KNOX: Thank you, Dr. Pappo, and thank you for the invitation to join this important meeting today. My name is Leona Knox. I'm an advocate and head of research at Solving Kids' Cancer in the UK, and I want to talk about accelerating cure for high-risk neuroblastoma from the perspective of the family.

The reason I am here as an advocate is my beautiful little boy, Oscar, which is a story that is very familiar to many of you. A few short weeks after this photo was taken, Oscar was diagnosed as having high-risk neuroblastoma at 3 years old. The disease was already present in his major organs, and it spread through his bones from his skull to his ankles; so much so that I asked our oncologist, "Is there any point?" and he outlined the treatment path. I did not want to put Oscar through all of that for him to die anyway. What he said was, "It's a challenge, but there's a chance," and from that moment onwards we put everything we had into getting him whatever treatment was needed to save

his life.

We enrolled Oscar on the SIOPEN HR-NBL1 trial, but the disease response after induction was insufficient for him to proceed on the protocol. He had multiple lines of therapy, he experienced severe toxicity, and ultimately we were unable to save him. He died in 2014 at age just 5 and a half.

So looking at frontline treatment for high-risk neuroblastoma, it's a dire picture. Before I speak to these statistics, I want to acknowledge that these are only the reality in more affluent countries, and that children and their families in low- and middle-income countries face an even worse fate. But here, despite intensive multimodal therapy, survival rates remain in and around 50 percent, so half of all children diagnosed with this disease will die, and of course those are the children that we track. The reality is that even fewer grow into old age.

Neuroblastoma accounts for 10 to 12 percent of deaths from malignancy in childhood, which is a

[indiscernible] representation given the incidence rate. Many children experience early disease progression despite the intense induction chemotherapy used to obtain maximal reduction in tumor burden ahead of the consolidation and post-consolidation phases of treatment, which highlights the importance of identifying the most effective initial treatment strategy, and sadly, 1 in 5 children do not achieve even a partial response to the current standard-of-care treatment despite the toll it has on their young bodies.

So all in all, our children suffer too much and too often without success. More effective treatments are desperately needed from the point of diagnosis.

I think it's important to keep in mind, and I know you all do, that during all these conversations about evaluating response, that behind every data point there is untold suffering. Neuroblastoma is an embryonal cancer diagnosed at around 18 months old, and at this age, our children are only starting to venture from our arms. We're

learning about their personalities and their social skills are really starting form. So to take a child at this point and have them undergo frequent and prolonged periods of hospitalization, separated from siblings, grandparents, cousins and peers, it has an enormous effect on their development, which we see play out in many different ways.

As parents we have to watch our children suffer the most horrendous side effects of treatment. We're watching for every hint that something is not ok, and the fear of not seeing the results when it comes to scans and other tests is all-consuming. We live every single day with the worry that our child will die, and that has a profound effect on us as parents, too.

The suffering doesn't end when treatment is completed. The children who do survive, they're facing a long list of potentially severe and life-changing complications, mentally and physically, brought on by the treatment itself. And although relapses aren't common, they do occur, and they devastate families all over again.

In taking a step back just for a minute, we know incredible things have been achieved, and many lives have been saved because of the diligent work and sheer determination of those before. But for the most difficult-to-treat cancers, we seem to have reached a plateau, not because of the lack of scientific progress and advances in knowledge, but because of the limitations in bringing those potential breakthroughs to the clinic and into standards of care.

More precise risk classification, post-optimization, and intensification of readily available therapies has allowed for incremental improvements in recent years, but to really change things for the children who are not well served by these approaches, we need brave, bold thinking and more agile approaches, which are still designed to protect them.

As a parent, the question raised in the Dubois paper really hits home. We do not want our children participating in clinical research that has little relevance, or as soon as it is completed

simply because the trials take too long to design, open, accrue, and, evaluate. This is a major challenge in the context of rare diseases. In high-risk neuroblastoma, where the phase 3 trials take an incredibly long time to produce a readout, it seems we are constantly doing [ph] the opportunity for further improvements.

Looking at the progress that is being made in terms of clinical research for high-risk neuroblastoma, it's a sobering picture. This is a very primitive search on clinicaltrials.gov and PubMed, but it speaks to the thousands of publications and hundreds of clinical trials which have resulted in just one class of targeted agents being incorporated into frontline therapy for neuroblastoma since the 1980s. I am not disparaging of the work that is being done or the commitment of the people involved at all, but when we look at what the impact is for children diagnosed with this disease, something seems quite wrong.

I speak from firsthand experience when I

tell you what a shock it is to find out at diagnosis that treatment for this disease is still being defined. It is incredibly unnerving at an already difficult time. We must work together to find ways to do better, and we must do it faster.

Dr. Bradford has already covered the Unituxin path to approval, so I won't spend too long on this, but we know this class of agents that was incorporated into frontline care has taken an incredibly long time to get there. This timeline from Bird, et al. clearly demonstrates the challenges involved in developing and evaluating costly new drugs in rare pediatric cancers. We just can't let this be the norm. We need to find ways to move the needle more and leave it much faster.

So how can we do this? It's evident that there is a clear need to find ways to assess the efficacy of new drugs more rapidly, but still robustly. The question is no longer if or why; it is how. I think having these conversations openly and having a much more coordinated approach by

cooperative groups, industry, regulators, payers, and with patient advocates is a strong start.

ACCELERATE has been championing this approach for many years, identifying the problems and actively working to find solutions. Early interaction between all stakeholders is vital, and I think it's highly relevant to point out the willingness of the regulators to participate in early conversations, including Dr. Bradford and Dr. Karres' talks today, and I've heard this discussed many times before.

We all want to see effective drugs reach as many children as possible, and even though we are all coming from different angles, we are pulling in the same direction, so let's work on making that even stronger. Of course, it would be a mess to ignore the fact that we need major investment to streamline clinical research in pediatric oncology and to achieve efficiencies that would be crucial in making breakthroughs and making them more quickly. Philanthropic funding is not enough. The scale of what is required is too much for us to be

able to deliver what children need.

Of course, none of this is easy. If it were, then we would not be here having this conversation and the solutions would already be implemented. The need for scientific robustness and ability to generate data to support regulatory filings which satisfy regulators and payers isn't arguable, but the major challenges of conducting clinical trials in a rare disease such as neuroblastoma across a network of 150-plus institutions must be recognized.

Everyone needs to work together to enable researchers to work in an environment that is conducive to progress, ensuring they have all the tools they need to maximize the impact of every study. It is another challenge. Are we gathering all the right data, and for long enough? Are we sharing and comparing it in a cooperative manner, and are we making sense of all of it, not ignoring the importance of quality of life and other considerations even long after clinical trials have ended?

I'm focusing more on the specifics of today's topic. How do we evaluate response in the modern era? How do we ensure the validity of historical controls when we have improved imaging techniques and the possibility of liquid biopsies? These things need to be embraced and figured out for the benefit of today's children and future generations, and what is stable disease, and what does it mean in the context of one spot versus 15 spots? Of course, there are others, so let's identify them and bring them into the conversation and maintain the ethos that this is difficult, but it's not impossible, and our children need us to figure this out.

So what's next? We need to define robust methods of using earlier endpoints. We need RAPID assessment, promising new therapeutic strategies, and we need to work closely with the FDA and the EMA who can help equip the right people with the right tools to gather, analyze, and rapidly report the scientific evidence to move the phase forward as quickly as possible with children who

desperately need it, and I'll end there. Thank you very much for your attention.

(Pause.)

DR. PAPPO: Dr Pinto, we cannot hear you.

DR. PINTO: I'm here.

Guest Presentation – Navin Pinto

DR. PINTO: Good morning, everyone. My name is Navin Pinto. I'm a pediatric oncologist at Seattle Children's Hospital. As mentioned in the beginning of the presentation, I have an uncompensated role as a member of the Scientific Advisory Board of Y-Mabs Therapeutics.

I won't belabor this, as it's been, discussed extensively, but high-risk neuroblastoma, the topic of our discussion today, is an ultra rare disease that affects less than 500 children per year in the United States, which represents 12 to 25 cases per million individuals. Aggressive multimodal therapy is necessary to achieve cure, and relapsed high-risk neuroblastoma is generally fatal.

We've talked about the two FDA approved

therapies for patients with high-risk neuroblastoma and, again, just a different representation of a topic that has been previously described is that the goal of induction therapy is maximal reduction in tumor burden.

So with the combination of multimodal chemotherapy and surgical resection of tumor, the goal is achieving as little disease as possible before moving on to the subsequent phases of the therapy. We've mentioned the painfully slow pace of drug development in ultra-rare diseases, and now this slide has been shown three times to illustrate how slow progress has been in neuroblastoma.

As Dr. Bradford mentioned, we are already using surrogate endpoints to make regulatory decisions in high-risk neuroblastoma. Obviously, overall survival is the ultimate measure of effect of a given drug, but given the time to read out for such an endpoint, a surrogate endpoint of event-free survival has been used in the past to make regulatory decisions for high-risk neuroblastoma, most notably, the event-free

survival benefit of dinutuximab as post-consolidation maintenance in patients with newly diagnosed high-risk neuroblastoma. But I think the crux of our problem is that oftentimes it's very hard to predict which patients will be failed by our standard or novel therapies, and earlier readouts of benefit are desperately needed.

There's been a long-standing recognition that early responses to therapy are often predictive of event-free and overall survival, and multiple investigators have shown that early responses to therapy can predict event-free and overall survival. The most notable publication was done by Greg Yanik and colleagues at the University of Michigan, with collaborators from the Children's Oncology Group, that showed that patients who had an end-induction Curie score, or MIBG score, less than 2 fared much better than patients who had a score greater than 2.

This analysis was done in an era where many patients were not receiving tandem myeloablative treatments, and many of these patients did not

receive post-consolidation dinutuximab. So in an effort to re-evaluate this in the modern era, I was involved in a retrospective analysis that I'd like to spend some more time describing.

We looked at four consecutive trials performed in the Children's Oncology Group for patients with high-risk neuroblastoma conducted in the 2000's, and patients on those trials with at least one response assessment during induction were eligible for this analysis. Importantly, the response criteria were uniform during this period of analysis.

The 1993 version of the International Neuroblastoma Response Criteria were used to evaluate response to therapy. The primary outcome of this analysis was to evaluate the partial response rate or better at end induction, and the secondary outcomes were complete response at end induction and progressive disease at end induction, and their impacts on outcome. We evaluated baseline clinical variables like age and stage, as well as available biologic variables for their

impact on outcome.

So in total, 1315 patients were potentially available for this study; 1280 of those patients had at least one response assessment, so formed the analytic cohort. You can see the breakdown by trial for each of these groups, with the majority of patients coming from the two randomized phase 3 studies, A3973 and ANBL 0532.

These are the results of that analysis, and similar to Dr. Yanik's presentation looking at Curie score, we saw that patients that had at least a partial response to induction fared much better both in event-free and overall survival compared to patients that had less than a partial response to induction. That effect was also demonstrated in a statistically significant way for patients that had a complete response to end induction versus those patients that had less than a complete response.

The conclusion from this section, international neuroblastoma response criteria have been built via an international consensus, but are complex. The INRC requires evaluation of anatomic

imaging, functional imaging like MIBG scans, and histologic response elements such as evaluation of the bone marrow compartment for metastatic disease. Using central review of these multiple data points is cumbersome and complex.

I've hopefully demonstrated that patients that have a partial response or better to induction chemotherapy tend to have more favorable outcomes, and one can conclude that interventions that improve the end-induction partial response rate will likely also lead to improvements in event-free and overall survival.

I'd like to just highlight that we have opportunities to test this hypothesis prospectively. The Children's Oncology Group is currently performing a randomized phase 3 study with the major question of the study occurring during induction, which is a randomized study of patients to receive standard treatment with or without the introduction of MIBG therapy during induction therapy.

Also as mentioned by Leona, the pace of

development in neuroblastoma proceeds, and oftentimes new data emerges while ongoing phase 3 studies are happening. One of the most exciting advances recently in high-risk neuroblastoma is the realization that combining anti-GD2 immunotherapy with chemotherapy leads to remarkable responses, first demonstrated in the relapsed setting, and recently published in the upfront setting by our colleagues at St. Jude Children's Research Hospital.

Sara Federico and Wayne Furman published a report showing that by incorporating a humanized anti-GD2 antibody into the induction schema for patients with newly diagnosed high-risk neuroblastoma, a remarkable end-induction response rate was seen with the vast majority of patients having either a complete or partial response to therapy and very few patients having less than a partial response.

This is the end-induction results for those patients, and even at an early time point, two cycles into treatment, you can see that many

patients had a dramatic reduction in their burden of disease. This translated to a remarkably high event-free and overall survival for patients treated on that study.

Obviously, this is a single institution study that needs some additional validation, and the Children's Oncology Group is planning our next phase 3 study, which will be a randomized study, again, asking an induction question. Again, the current planned protocol is to randomize patients to receive either standard induction chemotherapy with cytotoxic chemo and surgery alone versus the incorporation of dinutuximab into the induction chemotherapy regimens, and to evaluate the impact of induction dinutuximab therapy on event-free and overall survival.

We will have a readout of end-induction response on this study, so this, again, provides an additional opportunity to evaluate the impact of end-induction response to a novel therapy on overall and event-free survival.

In conclusion, we're currently evaluating

MIBG therapy incorporated into induction therapy on ANBL 1531. Chemoimmunotherapy with dinutuximab during induction will be studied in our subsequent phase 3 study, ANBL 2131. This allows for a prospective evaluation of novel induction regimens and their impact on event-free and overall survival.

I think that if we find that in both studies these interventions lead to better end-induction responses, and those better end-induction responses translate to better event-free and overall survival, we should have the information we need to suggest that this early-response time point can be used as a surrogate biomarker for regulatory decisions, and this hopefully will accelerate the path to approval for novel agents. Thank you for your time and attention.

DR. DONOGHUE: Thank you so much, Dr. Pinto. This is Martha Donoghue.

Guest Speaker Presentation – Mara Beck Popovic

DR. BECK POPOVIC: Hello. Good afternoon. My name is Maja Beck Popovic. I'm a pediatric

oncologist in Lausanne, Switzerland at University Hospital. It's my pleasure to continue the discussion on end-of-induction response as evaluation in high-risk neuroblastoma patients.

Many things have already been said. I will start just here showing my disclosures, and build on the fact that neuroblastoma is a very complex disease. This has been largely slow. Especially in high-risk patients, the needs for therapeutic improvement concerns many parts of the treatments. You have, as my colleagues have shown beforehand, different blocks of treatments. We are today discussing induction treatment mainly, but of course, patients will need also at other time points in their treatment, also in a relapsed setting, improvement in therapeutic approach.

Also, to respond to one of the questions that was asked earlier this afternoon, or this morning, is that patients who are treated within a high-risk regimen receive high-risk induction treatment. It has been shown the randomization is currently ongoing in the SIOPEN regimen. Patients

who have adequate metastatic response will go on to consolidation, radiotherapy, maintenance, and those who have either a refractory disease, which is defined as insufficient metastatic response based mainly on MIBG, are considered refractory patients and will go in the European setting to the VERITAS protocol that has been described earlier.

As I said, we are discussing induction today, but of course we are very much aware that this will not respond to questions further in the treatment and also in the relapsed setting of many high-risk patients.

So the question is, whether end-of-induction evaluation is a surrogate endpoint to event-free survival in patients with high-risk neuroblastoma? It certainly is an important time point, but we will, anyway, need event-free and overall survival as a complement to the end-of-induction question.

I would like to show or to add to the complexity of neuroblastoma as a disease also the complexity and effort that has been made over many years, over now almost 15-20 years, of an

international collaboration in order to develop and define a common language, because this is what we need in this rare disease, is to have a common effort and to talk about the same things when we define disease risk groups and how we evaluate response to treatment. These are criteria that we can, once developed, use in common collaborative studies.

I think that most of you who are present here today are aware of the international task force that has been developed, starting at the early years of early 2000, and collecting in a common database clinical and biological data on patients from U.S., from Europe, and from Japan. Currently, there is information of almost 25,000 patients in this quite unique and common database.

This has allowed progress in the INRG staging system to develop maybe a simplified staging system which is based on pretreatment, imaging-defined risk factors, which allows quite quickly and rapidly to define whether patients can be operated up front, or have a more extended local

disease, or are metastatic; and to incorporate the common established criteria for an internationally accepted pretreatment risk group classification; and to incorporate a consensus statement on molecular and radiographic techniques; and also a consensus statement on assessment of minimal residual disease, which allows us today to have a common risk group assignment that takes into account the staging system, age of the patient, histology, biological factors. And when we discuss very low, intermediate, high-risk, and very high-risk patients, to know about what kind of patient category we are talking and evaluating.

A parallel effort that has also been mentioned in a former presentation is the international initiative to define response criteria. These response criteria -- the criteria for diagnosis and also for the termination of response to treatment -- started in the '80s and have been reviewed in '93 -- Navin Pinto has shown this in the presentation -- and have in a further effort been modified in 2017 by incorporating

modern imaging techniques and incorporating new methods for quantifying bone marrow disease. This has been done over several years by many experts from different countries, in very regular conference calls, to end up in this work that allows, also when evaluating response to treatment, the use of common criteria and to have the same language.

This system of evaluating response is complex. I wish to show maybe for people less familiar with neuroblastoma how complex it is because in order to be able to have acceleration in the development, and in incorporating new drugs, we must think of what criteria we are going to use in order to have a simple and efficient tool to evaluate.

Now, neuroblastoma is a complex disease, and as it has already been shown at the very beginning from Leona, it can spread over all the body, and many aspects have to be evaluated. So the INRC system in the latest version defines assessment of primary tumor, of soft-tissue metastases, bone

metastases, and a new more refined and precise way, bone marrow infiltration in aspirates and trephine biopsies.

All this evaluation then serves to evaluate and to define overall response, and to define how complete response should be defined, partial response, minor response, stable disease, and progressive disease. This gives us the opportunity to have a uniform assessment of disease response, to improve interpretability in our common effort, and to facilitate collaborative trial design.

Now, what tools are used to evaluate primary and metastatic soft-tissue disease and response? It has already been mentioned by Navin, anatomic imaging, but also, then, functional imaging. We have for the evaluation of metastatic bone disease also MIBG imaging, but then also the type of imaging used to evaluate osseous lesion that has not an involvement of soft tissue, and then a precise description of how to evaluate metastatic bone marrow disease.

Now, these tables I show you are not

intended to read them all through, but just to show and to illustrate the complexity, and also to show that for the definition of primary and soft-tumor response to treatment, we use always anatomic variation and MIBG or FDG-PET imaging, and to evaluate tumor response at metastatic soft tissue and bone site, the same. When you see the description, this reflects -- I'm sorry. I'm using my pointer from the computer and not the good one. If you look at the details of the evaluation, this illustrates very well the complexity that the disease imposes by itself.

Here is the evaluation and definition of minimal marrow disease. We all know that bone marrow infiltration is one of the very big challenges in how to evaluate in what uniform way and how also to organize review if you wish to implement a central review. The combination of all these individual components gives us the tools to define complete remission, partial remission, minor response, stable disease, and progressive disease.

We have a stratification elaborated that

allows us to define homogenous treatment groups. We have our various risk-group patients and EFS mainly used to modulate treatment. We know that when patients have very good event-free survival, we can reduce treatment intensity a lot. If they have low event-free survival, under 50 percent, we need to intensify treatment. Having these tools that have been developed over many years allows us to have a comparison of risk-based clinical trials conducted in different regions in the world and helps us to develop international collaborative studies.

Early-phase trials need also a definition to help how to construct them and how to define them. I would like to take the opportunity here to cite Julie Park's and collaborators' work that has been recently presented and submitted in how to develop criteria for early-phase trials, which are supposed and which are intended to help us in the development and in the acceleration of the development of new treatments in neuroblastoma high-risk patients.

The aim is to establish a consensus approach to conduct clinical trials, which needs a precise and better definition of progressive refractory disease to establish a clear definition of eligibility criteria for early-phase trials, the comprehensive extent of disease evaluation at certain time points, and definition of response evaluation, bone marrow being one among the major challenges.

My comments and thoughts to today's discussion is that we have, through our international collaboration, developed common tools for risk-group assignment. We have developed common tools for uniform response evaluation. We have developed an international common database with data that can be used for project evaluation and developing research questions, and we have a consensus on a harmonized way of how to conduct early trials.

We need to accelerate development of new drugs in patients with neuroblastoma to improve the patient's pathway, and this has been very well

explained and shown by Leona at the very beginning. We need to accelerate introduction into frontline treatment, and then standard of care. This needs the close collaboration we have already mentioned, which includes also patient advocates; and in pivotal studies, end of induction is certainly an acceptable endpoint.

Not to forget, we will still need to have developments in the relapsed and refractory setting, where safety, pharmacokinetics, and preliminary activity data are still needed. We have throughout all these common efforts set an international collaboration that has also developed tools we can use today to evaluate disease based on a common language.

I would like to conclude with saying that end of induction as an endpoint, yes, can be used, and there is certainly a need for acceleration in the upfront setting of high-risk neuroblastoma patients and as an intermediate endpoint which, however, needs to be complemented with event-free survival. We have tools that can be used to

evaluate end-of-induction response, and it might be a solution to have simplified INRC maybe using the metastatic response by MIBG score because this has been well documented and published.

How shall we do this? By working together, SIOPEN and COG, hand-in-hand with FDA, EMA, and with patient advocates to agree on the end-of-induction response criteria. What has been done in the past certainly will be helpful for now and for the future. I thank you very much for your attention.

DR. PAPPO: Thank you very much for these excellent presentations.

Before we move to the clarifying questions section, Dr. Popovic, we were unable to capture fully all of your disclosures in the transcript. Would you mind just reading them again? Sorry for the bother.

DR. BECK POPOVIC: Oh, no problem. I'm currently the SIOPEN president, and SIOPEN is still receiving royalties for the trials of dinutuximab beta that has been agreed on many years ago. I

have been involved twice in discussions with Y-Mabs with any compensation and with any decision.

Clarifying Questions

DR. PAPPO: Thank you very much, and sorry for the bother.

We will now take clarifying questions for Ms. Knox and Drs. Pinto and Popovic. Please use the raise-hand icon to indicate that you have a question, and remember to clear the icon after you have asked your question. When acknowledged, please remember to state your name for the record before you speak and direct your questions to a specific presenter, if you can. If you wish for a specific slide to be displayed, please let us know the slide number, if possible.

Finally, it would be helpful to acknowledge the end of your question with a thank you and the end of your follow-up question with, "That is all for my questions," so we can move on to the next panel member.

We can get started with some questions. I had a quick question for Dr. Pinto.

In your analysis of end-of-induction response, did you have any follow-up data on the patients that did not achieve a PR and how they were treated, their ultimate outcome, and if there were any hints or any signal that some subset of patients could be actually retrieved; or there's not enough genomic data or anything to make any conclusions about that?

DR. PINTO: Thanks for that question. Yes, I think it highlights one of the biggest challenges in neuroblastoma. I think because there's widespread recognition that less than a partial response is a predictor of poor outcome, many providers are oftentimes seeking additional salvage therapies to try and drive patients into a better end-induction remission, so many patients would come off protocol therapy to receive additional therapies.

Now, there is a subset of patients, obviously, that continue on therapy. All of the studies I highlighted would allow the patients even with stable disease to continue on to subsequent

therapy, so those patients obviously had less than a partial response. We did capture some of those patients, but many of the patients came off protocol therapy, and then we do not have robust data about what therapies those patients received and how they responded to that therapy. We just have vital information about alive or dead.

DR. PAPPO: Thank you very much. That answers my question.

Ro Bagatell is next.

DR. BAGATELL: Hi. This is Ro Bagatell from the Children's Hospital of Philadelphia. I'd like to thank the speakers for really excellent presentations that have highlighted so many of the complexities of high-risk neuroblastoma therapy.

My questions are for Drs. Pinto, Beck Popovic, and possibly Bradford. You've all talked about the very lengthy treatment that is administered to patients with high-risk neuroblastoma and the heterogeneity within the neuroblastoma patient population. I am curious as to your thoughts about how we apply the data from

Dr. Yanik's analysis and Dr. Pinto's analysis, end-induction response in the evolving setting with much more lengthy therapy over time; the addition of tandem transplant; the addition of post-consolidation therapy; and even in the era of ALK-directed therapy, a targeted agent throughout the entirety of treatment plus a continuation phase.

I'm just interested in how you think about the heterogeneity of the treatments, the length of the treatments, and the heterogeneity of the patient populations as we try to interpret the end-of-induction response data that you've presented.

DR. PINTO: I'll take a stab at that. This is Navin Pinto, and Maja, I would appreciate your comments and thoughts as well.

Yes. Ro, I think you've highlighted a really big problem. The average course of therapy for a patient with newly diagnosed high-risk neuroblastoma is nearly 18 months of intensive therapy, and then obviously we're waiting for

biomarkers, or readouts, like 3-year event-free and overall survival. So I think the highlight of the path to approval for dinutuximab highlights the process from IND to standard readout of event-free or overall survival and how long that process takes.

I think this is where the opportunity for something like end-of-induction response may really serve this unmet need and help accelerate approvals. We've now shown, time and time again, that patients that do worse at early points in therapy fare worse, eventually. So I think if we can move the needle earlier in therapy with induction strategies, that can potentially accelerate approval.

Now, unfortunately, I think one thing that we're all interested in is not just induction therapy. We're interested in ways to modify our current consolidation regimen to make them more effective and hopefully less toxic, and we're interested in both post-consolidation regimens that can sop up minimal residual disease, and then

remission maintenance strategies that can prevent relapses.

So again, I think this is not a one-size-fits-all problem. Obviously, we want to encourage innovation in the other phases of high-risk neuroblastoma care, but again, I think we have an opportunity for those interventions that make sense in the induction regimen to have a potentially quicker path to regulatory approval, where there's clearly an impact of those interventions.

DR. BECK POPOVIC: Thank you, Navin. Maja Beck Popovic is speaking.

Thank you, Ro, for this question. The patient heterogeneity is really a problem, in fact, we are not sure when we evaluate at the end of induction, at the static response, whether, really, we have all the same patients. We say it's a complete remission or partial remission. Is it really the same for all these patients?

I think that the development and the implementation of biomarkers as an additional tool

to evaluate disease response will be helpful, helpful at the end of induction and probably in other treatment blocks that are as important for further treatment in neuroblastoma patients.

I see it personally this way; that this will help us then to identify subgroups or see whether when we evaluate end of induction, with what we have as tools now, if this is really all the same patient population we are evaluating or not. Then of course, if we have biomarker-guided treatments, this might then shorten, or prolong, or modify at various time points for patients their treatment.

I don't know if this response answered your question.

DR. BAGATELL: Very helpful. Thank you.

DR. PAPPO: Does that answer your question, Ro?

DR. BAGATELL: Yes. Thank you.

DR. PAPPO: Okay.

Dr. Ted Laetsch, you're next.

DR. LAETSCH: Thank you. This is Ted Laetsch. I just want to thank the speakers for

their excellent presentations. The evidence, clearly in my opinion, demonstrates that the end-of-induction response is able to predict EFS.

I did have a question for Dr. Pinto. As I consider the next planned COG trial, I just wonder how you think about moving a known active agent, immunotherapy that's part of standard upfront therapy now, from maintenance to induction so that now immunotherapy will be given before the surrogate endpoint rather than after it, and wonder if you think there's a potential for that to improve end-of-induction response by providing more active therapy before that time point, but potentially not impact EFS or OS.

DR. PINTO: Again, I think that's where this opportunity exists to truly see if these types of end-induction responses do translate to improved event-free and overall survival. Again, I think if there's enough data in the aggregate to suggest that we can make these decisions before the completion of 2131, that would be fantastic, but 2131 will hopefully provide an opportunity.

I think that, Ted, you've highlighted that this is an issue where we're using the same drug, both in the induction phase of care and in the post-consolidation maintenance. I don't pretend to understand why the addition of dinutuximab to chemoimmunotherapy has had such a big impact in the relapsed setting, but it clearly has. I think, as Dr. Federico and Furman's recent publication highlights, that seems to also have a very robust signal in the newly diagnosed setting.

So I think we're excited about the incorporation of dinutuximab, at least hypothesized, that in the setting of concomitant chemotherapy, the mechanism of action of tumor control may be different. Those are my thoughts about that plan and happy to hear anybody else's thoughts.

DR. PAPPO: Ted, does that answer your question?

DR. LAETSCH: Yes, it does. Thank you.

DR. PAPPO: Ira Dunkel, you're next.

DR. DUNKEL: Thank you, Dr. Pappo. Ira

Dunkel, Memorial Sloan Kettering. I think my question is primarily for Dr. Pinto, but perhaps others might wish to comment, too.

I guess my question is, when you have an intervention that you deem very promising, like the St. Jude earlier use of dinutuximab, I wonder if you could discuss when it's most appropriate and necessary to study it in a phase 3 trial versus in a multicenter phase 2 trial, which obviously has disadvantages but also would increase efficiency using less patients and shortening the trial duration. Thank you.

DR. PINTO: Thanks, Dr. Dunkel, for your question. I just wanted to highlight I apologize if I misspoke during my presentation, but the antibody used in the St. Jude trial is not dinutuximab; it is a humanized antibody with an additional mutation to prevent complement fixation, which is the main mediator of pain with anti-GD2 antibodies. So it was a novel antibody similar to dinutuximab, but distinct. That is probably the major reason for a confirmatory study with a

different, more widely available GD2 antibody.

I think the other issue that we need to confirm is I'm a person who was born in Peoria, Illinois, and there's a famous phrase of "Will it play in Peoria?" meaning something that's done at a very esteemed, very well resourced center like St. Jude translate to smaller centers that are still providing care to patients with high-risk neuroblastoma?

So I think that the thought of the COG leadership was that the most sound way to do that was in a randomized phase 3 setting.

DR. DUNKEL: Thank you very much. I don't think that you misspoke. I think that I misspoke, but thank you for correcting me there.

DR. BECK POPOVIC: If I can just add, I can only confirm these needs. We have similar reflections and thoughts also from the SIOPEN view, that we need some -- the combination of induction treatment, that is a little different, the COG induction regimen. We need some safety information, but the aim is then to have it quite

quickly, in a randomized way, implemented in induction, so I think I go the same way.

DR. PAPPO: If I may add for a couple of minor comments from Dr. Pinto for Dr. Dunkel, a couple of the other differences of this antibody, the St. Jude antibody also was 98 percent humanized, so it's really not chimeric.

The other issue is that the level of glycosylation is significantly less than with other antibodies, and we believe that that is important to increase ADCC, which is one of the main mechanisms for this antibody to work, so there are some minor differences there.

Your question was answered, Ira.

I'm going to go to Dr. Kraus now.

DR. KRAUS: Yes, and thank you very much for all these presentations. They're extremely helpful. I was impressed by the data presented particularly by Dr. Pinto and the relation of greater than PR, CR, event-free, and overall survival. On the research and development level, the Kaplan-Meier curves are showing phenomenal

differences, and we rarely see Kaps that wide, so this is very valuable and important.

The interesting thing I saw, it looks like a wider gap with the greater than PR than a CR, which may or may not be something you would predict. Maybe you would; I don't know. But I wanted you to comment on it and ask if you'd dug into duration of response, and if that's at play here. But all in all, I think this is very informative, important data on a large patient group. Thank you.

DR. PINTO: Great, and thank you for that question.

DR. KRAUS: Yes.

DR. PINTO: This is Navin Pinto. I think I'll try and tackle the duration of response question first because I think, again, it highlights another very difficult question in neuroblastoma.

So oftentimes, patients with residual disease or persistent disease will not settle for that, and oftentimes will cycle between multiple salvage and experimental therapies in order to try

and achieve a remission. So really, the duration of a complete response is probably well known, but in patients that have partial responses, this is a very difficult question to answer.

I think with the emergence of remission maintenance strategies like a GD2 vaccine that's being developed by, first, Memorial Sloan Kettering, and now Y-Mabs Therapeutics, and a remission maintenance drug, DFMO, being investigated by the Beat Childhood Cancer Consortium in the United States, that makes the challenge even harder, even in patients with a complete remission, so duration and response is a really difficult question to tackle.

I think the point that you highlighted, it would be easy to say -- I don't think we would be even having this meeting if the CR curves at end induction were flat at hundred percent or higher than they are. So it's clear that even at end induction, patients that have a remarkable response, with the combination of chemotherapy and surgery mostly, some of those patients do,

unfortunately, go on to relapse and die of their disease. So it's not a perfect biomarker and does highlight the need for additional strategies in consolidation and post-consolidation maintenance. But again, I agree with you that it is a relatively powerful biomarker of overall response.

DR. KRAUS: Thank you. That's very helpful. Appreciate it.

DR. PAPPO: Dr. Donoghue, you have a comment?

DR. DONOGHUE: Thank you, Dr. Pappo. I actually have a couple of questions if that's ok. I want to thank, first, the presenters for their really informative and helpful presentations, and my first question is for Dr. Pinto.

Thank you so much for presenting, at a high level, the analysis of data from several COG trials, looking at that correlation between end-of-induction response and EFS and overall survival. My question just relates to whether -- and I apologize if you touched upon this and I missed it -- those analyses at all controlled

for other factors that could be predictive of patient outcome such as N-Myc status, age, et cetera, and whether those analyses showed similar results looking at that association between end-of-induction response and EFS and OS.

DR. PINTO: Thank you. That's an excellent question. I did not highlight that during this talk, but it is highlighted in the manuscript for others' reference. But briefly, we did look at clinical and biologic factors that were known for this group of patients, clinical factors like age at diagnosis and clinical stage using the previous staging system; as well as, for many of the patients, biologic information like amplification of the MYCN proto-oncogene, and for us, a smaller subset of patients, segmental chromosomal aberrations, which have been demonstrated by the SIOPEN group to be predictive of outcome.

In summary, the only biologic factor that survived a multivariable analysis as predicting end-induction response was the presence of an 11q segmental chromosomal aberration. 11q loss in

neuroblastoma has been a long-standing biologic factor associated with poor prognosis both in non-high-risk neuroblastoma and in high-risk neuroblastoma, so this was the only biologic factor that survived that multivariable analysis.

DR. DONOGHUE: Thank you, Dr. Pinto. That addressed my question.

I have one additional question, and this is for both Dr. Pinto and Dr. Beck Popovic, related to the SIOPEN trial and the planned ANBL 2131 trial. I just wanted to make sure my understanding is correct that with the ANBL 2131 trial, the plan is for patients who have progressive disease during induction to then switch to extended induction, so peel away a bit from the main trial, versus with the SIOPEN trial, that patients who have an inadequate response to induction therapy would be eligible for the VERITAS trial to go on to receive 131 MIBG.

I just wanted to check and see if that is correct, and whether there is a difference between those trials in terms of how response to induction

is being assessed and deemed inadequate or adequate.

DR. PINTO: Yes. I'll start with the COG 2131 plan. This trial is still in development, but the current proposal is that, as you mentioned, patients with progressive disease during induction have historically come off protocol therapy and not been well captured by current COG protocols, but in addition, patients with a poor end-induction response, which is in defined in the protocol, those are patients largely with persistent metastatic disease at end induction and will be eligible for an extended induction phase of chemoimmunotherapy with the hopes to capture as many of these poor end-induction responders as possible and to get a better understanding of some of the questions that other panelists have raised now.

That's the COG perspective, and Maja can provide insight on SIOPEN.

DR. BECK POPOVIC: Yes, thank you. From the SIOPEN perspective, end of induction, poor

metastatic response means more than 3 MIBG spots still active, and there are some additional factors for bone marrow evaluation.

These patients currently can go on to the VERITAS protocol, which starts by 3 courses of irinotecan and temozol as re-induction, and then a randomization to receive a double transplant. On one hand, it is MIBG with topotecan followed by BuMel, and on the other, high-dose thiotepa followed by BuMel. These patients, then, if they respond well to these treatments, will go on further to surgery, local radiotherapy, and maintenance. This is the current setting for high-risk patients that have insufficient metastatic response at end of induction.

DR. DONOGHUE: Thank you so much. That answers my questions. I appreciate it.

DR. BECK POPOVIC: Thank you.

DR. PAPPO: Dr. Reaman has a question.

DR. REAMAN: Thanks, Alberto.

I guess this is primarily for Dr. Beck Popovic. You mentioned, I think, the difficulty

with respect to assessment of response as it relates to bone marrow disease.

Can you just provide a little bit more detail with respect to how central review for response assessment for marrow disease might be accomplished, should be accomplished, could be accomplished, within the context of a multisite, even multicenter, study?

DR. BECK POPOVIC: Yes. Thank you very much.

(Crosstalk.0

DR. REAMAN: And whether you think that's necessary? Sorry.

DR. BECK POPOVIC: Yes. Thank you very much for your question. I think that this is feasible. It can be organized when it's planned, prospectively. We have been suffering in the past from the fact that the bone marrow evaluation is done in laboratories that are acknowledged as experts in the field, but it was not planned for regulatory issues at the end, and this needs another organization.

I mentioned just the difficulties because it has also needed quite a lot of work to agree on how bone marrow shall be evaluated exactly and the response to it. In a prospective setting, this is feasible because in a prospective setting, if the criteria that have been developed are implemented, then it's a question of putting the labs together and organize a central review.

In the past, in our European studies, not having planned beforehand against the regulatory aspects, this review would have been a problem, whereas MIBG response has very early been implemented as one of the main factors and with a uniform scoring system, which has proven quite efficient also in the evaluation. So my wish was not to say that it is not possible, but it might be more complex. But when it is planned in advance, I'm sure that this can be done.

DR. REAMAN: Thank you. And I didn't mean to imply that you said that it was not possible or feasible. I was just wondering how important, and what are the plans to accomplish this.

Currently, to sort of follow up on your response, are you planning that marrow assessment include just histologic examination or using specific immunohistochemical techniques, or molecular --

DR. BECK POPOVIC: Yes.

DR. REAMAN: -- techniques, and all of the above?

DR. BECK POPOVIC: So there is histology, but there is immunocytology, there is also immunohistochemistry, and there is also RT-PCR technique, which is not yet validated as such. So it is not histology only. For histology, however, in the trephine biopsies, the limit of 5 percent has been set, which means that minor presence of cells is something that is acceptable and can be considered as negative.

DR. REAMAN: Thank you. Then just for clarification, is this something that may come a part of the INRC requirements for evaluations --

DR. BECK POPOVIC: Yes, it is part of the INRC requirements. It's in one of the tables I've

shown, yes.

DR. REAMAN: Okay. Thank you very much.

Then just one other quick question for both Drs. Pinto and Beck Popovic; the absence of a complete response covers a broad group of patients. Are there any indicators that lack of response in one area -- be it primary tumor site, visceral metastases, bone metastases, bone marrow -- that there may be prognostic significance to a specific area or specific disease site where there is lack of response, complete response?

DR. BECK POPOVIC: I think there are good indications that really a lack of response in bone is one of the major bad prognostic factors.

DR. REAMAN: Okay. Thank you.

Was this something that was evaluated, Dr. Pinto, in your analysis of patients who did not have a complete response at end-of-induction therapy?

DR. PINTO: Unfortunately, we didn't have as detailed of information. We had an overall assessment of response using the INRC criteria, but

the individual elements of response were not available for this analysis, and I think again highlights that as we've built a complex response, criteria, deconvoluting that to assess its impact can be difficult.

DR. REAMAN: Okay. Thank you. That answers my questions. Thank you.

DR. PAPPO: Dr. Seibel, DO you have a question?

DR. SEIBEL: Yes. Nita Seibel from NCI, and this is for Dr. Popovic.

For the patients who would go on VERITAS who are MIBG non-avid, how will they be treated?

DR. BECK POPOVIC: Thank you for your question. Patients who are MIBG non-avid are evaluated by FDG-PET, so then the criteria will be used the same, but it's not MIBG if they are not avid.

DR. SEIBEL: And then they will be non-randomly assigned to the arm that doesn't include MIBG for the VERITAS trial?

DR. BECK POPOVIC: This is a very good

question. I suppose, yes, but I have to look up in the protocol, but this makes, of course, sense that they could not be then treated by MIBG, yes.

DR. SEIBEL: Okay. Thank you. That answers my question.

DR. PAPPO: We still have a few minutes for additional questions before we break for lunch, so I will give you a minute or so to raise your hand.

(No response.)

DR. PAPPO: I don't see any additional, hands. I want to thank all the presenters for their outstanding presentations and the panel for being so interactive.

Since there are no additional questions, we will now break for lunch. We will reconvene at 1:00 p.m. Eastern Standard Time. Panel members, please remember that there should be no chatting or discussion of the meeting topic with anyone during the break. Additionally, you should plan to rejoin at around 12:50 p.m. to ensure you are connected before we reconvene at 1:00 p.m. Thank you very much, and enjoy your break or your lunch.

(Whereupon, at 11:58 a.m., a lunch recess was taken.)

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

(1:00 p.m.)

DR. PAPPO: Welcome back to an afternoon session. We will now proceed with a speaker and FDA presentation from Drs. Lisa McShane, followed by Dr. Anup Amatya.

Dr. McShane?

FDA Presentation – Lisa McShane

DR. McSHANE: Thanks very much.

Good afternoon, everyone. I am pleased to be here to share some thoughts from a statistician's perspective about how we might use early endpoints to support drug development in high-risk neuroblastoma. Early endpoints can serve in many roles, and I will discuss what evidence is required to support the various uses.

Here are my disclosures. I have none. I do want to emphasize that I am not employed by FDA; I work for NIH. There will be a speaker following me, Dr. Amatya, who will be providing a more in-depth discussion of regulatory considerations for use of early endpoints. My talk will be a

little more on the conceptual side.

A good starting point is to make sure that we're speaking the same language. For this I turn to the BEST resource. BEST stands for Biomarkers, Endpoints, and Other Tools. It's a resource that was developed by a working group charged by the FDA-NIH Joint Leadership Council to develop a glossary of harmonized terminology for biomarkers, endpoints, and other tools useful in medical product development or regulated product evaluation. It contains clear definitions of useful terminology and many explanatory examples, so I want to make you aware of this and encourage you to take a look at the website.

For purposes of today's talk, my focus is on end-of-induction response, which would fall into the category of response biomarker in the BEST glossary. This is the definition of response biomarker from the BEST glossary. We divide this category into two main groups, the first being pharmacodynamic biomarker, which indicates biologic activity of a medical product or environmental

agent without necessarily drawing conclusions about efficacy, or clinical outcome, or even linking the activity to an established mechanism of action.

What might be a more familiar term, or a popular term but a distinct entity, is "surrogate endpoint biomarker." It is an endpoint that is used in clinical trials as a substitute for a direct measure of how a patient feels, functions, or survives.

Too often, people prematurely take the leap from an early-response biomarker to surrogate without having evidence to establish that it can be used reliably as a substitute, and I underscore the word "substitute" for a definitive clinical endpoint.

We'll touch on evidence for surrogacy later in this talk, and Dr. Amatya will also talk more about it. But importantly, there are still things that an early endpoint can be useful for without it meeting the very rigorous requirements for surrogacy, so I'll start with some of those easier topics first.

We already use many response biomarkers in drug development programs. Within the class of pharmacodynamic biomarkers, we might use early endpoints to enrich for patients for whom a modified treatment strategy may be evaluated. For this role, we want a biomarker that is measured after some initial course of therapy and is prognostic for subsequent outcome; in other words, it's a correlative long-term clinical outcome, and I'll talk about the rationale for enrichment in a minute.

Early endpoints may also be used to drop drugs early in the development process. For example, many phase 2 trials for solid tumors use tumor response as the primary endpoint. If a drug can't produce tumor shrinkage, chances that it will improve survival are greatly diminished; therefore, poor performance on tumor response might lead researchers to not continue on to a phase 3 trial, or in a multiarm trial, certain treatment arms might be dropped early for poor performance on an early endpoint. Similarly, in an adaptive

phase 2/3 clinical trial, we might not proceed to the phase 3 stage.

Use of an early endpoint for enrichment can be a very efficient strategy for drug development. The regulatory definition of enrichment from the FDA guidance document on this topic is shown here. Enrichment refers to prospective use of any patient characteristic to select a study population in which detection of a drug effect, if one is in fact present, is more likely than it would be in an unselected population.

It might be used to reduce inter- or intra-patient heterogeneity or to enrich for patients in a certain prognostic category. This might be a poor prognosis subgroup for which we expect more events, thus increasing statistical power for detecting treatment effects, or it might be a good prognosis subgroup for which treatment de-escalation might be considered.

Finally, there is predictive enrichment in which we select patients based on some biological evidence that they are more likely to respond to

the investigational intervention under study; for example, and a classic example in oncology would be a somatic mutation targeted by a small-molecule inhibitor or antibody therapy. The type of enrichment I will focus on today is prognostic enrichment.

Shown here is demonstration of the prognostic ability of end-of-induction response for event-free survival and overall survival in a study including 1280 patients across four high-risk neuroblastoma trials. You heard about this trial already from Dr. Pinto.

As shown in panels A and B, responders have longer event-free survival than non-responders, whether we define responders as PR or better or as CR or better, according to the 1993 response criteria used here. The same holds for the overall survival endpoint. All of these associations were statistically significant. You also heard this morning that these held up even after adjustment for other clinical covariants.

Another study looked for similar

associations of end-of-induction response with event-free survival and overall survival using the 2017 response criteria. Here, response was defined as minor response or better, according to the definition that you see in the table at left.

The associations did not quite achieve statistical significance for event-free survival but did for overall survival. However, it's really important to recognize the extremely small sample sizes in this study. There were only a handful of non-responders. These findings are certainly interesting, and we hope they pan out, but they would need confirmation in a larger study.

If you buy into the idea of end-of-induction response being prognostic, how can we use that in an enrichment strategy? As this diagram shows, what one might typically do is use the early endpoint to separate patients into two groups. Those who respond by end of induction perhaps go off study to get usual care. Those who do not have response by end of induction are randomized often to standard of care versus an experimental therapy.

There are two reasons why we might want to randomize this particular group, namely the non-responder group. I mentioned before the event rate will be higher in this group, and that translates into higher statistical power for treatment comparisons. This group might also be seen as the one most urgently in need of better treatments.

For sake of completeness, we could also think about a predictive enrichment strategy if when looking at these two groups divided by end-of-induction response outcome we can identify biomarkers that distinguish these groups and are actionable in the sense of having match targeted therapies. Again, I won't have time to go into predictive biomarkers, but I just want to remind you that that is another option.

As I mentioned, this prognostic enrichment strategy has been a successful drug development strategy for many pediatric cancers, but I want to be clear that we should not take it for granted that intensifying therapy in the higher risk,

non-responder group will always lead to a better outcome.

The very high-risk cohort of the B-ALL trial, AALL1131, provides an example of how one might attempt to acquire evidence to support clinical benefit of intensifying therapy in a biomarker-defined subgroup that has worse outcome based on an intermediate endpoint.

On the right are the criteria for eligibility for randomization in the very high-risk cohort with eligibility criteria, including day 29 bone marrow MRD greater than or equal to 0.01 percent; or induction failure, meaning greater than 25 percent blasts in the bone marrow; or M3 on day 29.

The randomization compared a control therapy with two different levels of intensified therapy after a standard 4-drug induction regimen. The CONSORT diagram on the left shows the number accrued to the 3 arms, and patients were randomized 1 to 2 to 2 between February 27, 2012 and September 13, 2012.

On the left are details of the three treatment arms in the randomized very high-risk cohort. Arm 2 delivered the most intensive therapy, adding cyclophosphamide, etoposide, and clofarabine during the second half of consolidation and delayed intensification. Unfortunately, as the 2018 cancer paper by Saltzer reports, the experimental arm 2 was stopped early due to excess toxicity, particularly grade 4-5 infections and pancreatitis. The trial continued with only the control and experimental arm 1 using a 1 to 2 randomization.

Unfortunately, the news did not get much better over time. In February 2017, experimental arm 1 was closed for futility with a hazard ratio of 0.606 favoring the control arm. With additional follow-up in December of 2017, the evidence was even stronger that the experimental arm was not superior to control, with a difference of 4-year disease-free survival of 85.5 percent for control versus 72.3 percent for experimental arm 1.

Another thing I want to point out here is

that the 4-year disease-free survival of 85.5 percent reported for the control arm was quite a bit higher than the 70 percent originally predicted, based on data available for patients with very high-risk features treated in the preceding B-ALL studies, and even the experimental arm disease-free survival rate was numerically slightly higher. This serves as a reminder of the need for randomized-controlled trials.

Now a few words on the other uses of end-of-induction response that fall into the category of pharmacodynamic biomarker; there are several points to consider when using end-of-induction response as an early endpoint to assess drug activity, which, keep in mind, might not necessarily translate to efficacy.

The prime consideration here is that we want the early endpoint to be good at ruling out drugs that have minimal activity and little chance of improving more definitive clinical outcomes, while not prematurely discarding too many good drugs. The suitability of an endpoint for this purpose may

depend on the drug class. For example, in solid tumors, we don't necessarily expect a cytostatic drug to shrink tumors very much.

The other use I have listed here is a little bit more tricky. When we use an early endpoint for selection among drugs -- for example, in head-to-head comparisons of drugs in a phase 2 trial for decisions about moving drugs from phase 2 into phase 3 -- we're hoping that the endpoint can at least predict large differences in efficacy or has reasonable ability to rank drugs for efficacy with respect to longer term definitive clinical endpoint. There's a lot of trickiness here, as I mentioned, and I think that will become apparent as I talk a little bit more about surrogate endpoints.

I just mentioned that using early endpoints for this preliminary selection can be tricky, and it can be tempting to actually replace a more definitive long-term endpoint with an early endpoint, and not have to measure the definitive endpoint at all, which might take a lot longer time and require more resources. So there's a lot of

attractiveness to having this replacement endpoint.

This really brings us to the idea of a surrogate endpoint. People often will use the term "surrogate" without being very rigorous about what they mean or what conditions might be required to use something as an honest-to-goodness surrogate as a replacement endpoint, so I'll briefly introduce some of the key concepts relevant to surrogate endpoint in my last few minutes, but in the next talk, Dr. Amatya will give you a more detailed regulatory perspective.

The discussion often starts with the famous Prentice criteria, which states, "idealized conditions for a surrogate." The treatment has an effect on the true or definitive endpoint, for example, survival; the treatment has an effect on the surrogate; the surrogate is associated with or prognostic for the true or definitive clinical outcome; and the surrogate must fully capture the net effect of treatment on the true clinical outcome.

The problem is that this rarely holds,

especially the last point. Even for one treatment, and much less for multiple treatments, one might wish to compare. So while this is a conceptually appealing set of requirements, these conditions are generally impractical to meet.

The visual representation of the Prentice criteria seen in this diagram, basically all of the action of the treatment is thought to happen on a direct pathway through the surrogate endpoint to the definitive endpoint. An early endpoint with properties displayed here would be both a good prognostic indicator and a good surrogate endpoint, at least for the particular treatment. Most difficulties arise when we're trying to compare two or more treatments because we don't know if the mechanisms of all drugs lie on exactly this same pathway.

At this point, we need to get a little more granular. There's some confusing terminology out there that we need to straighten out. We need to distinguish between the notions of individual-level surrogacy and trial-level surrogacy. An

individual-level surrogate is an endpoint or variable that is a correlate or prognostic for the true clinical endpoint within the context of specified treatments and patient population. This may be demonstrated, in fact, even in the context of a single cohort or clinical trial, but as Fleming and DeMets caution in their landmark paper, "A correlate does not a surrogate make."

So what else do we need to be able to use it truly as a replacement endpoint? What we need is to establish that it's what I would refer to as a trial-level surrogate, and that's an endpoint or a variable that can replace the true clinical endpoint. That's a very big hurdle to clear; actually replace the definitive endpoints.

What does it take? Well, demonstration of trial-level surrogacy generally requires a meta-analysis of clinical trials to show that a conclusion about treatment effect, based on the surrogate, reliably predicts the conclusion obtained using the true endpoint, and this has to hold across trials because, remember, the whole

point of having a surrogate is that you'd like to conduct a new trial and not have to measure the definitive endpoint. If you could just measure the surrogate, that could save us lots of time and resources.

Here are some examples that illustrate why prognostic ability of an early endpoint does not guarantee trial-level surrogacy. I know that may seem very counterintuitive. You think, well, if you have an early endpoint that is highly prognostic for the endpoint, and you can improve the outcome on that candidate surrogate marker, how could it possibly be that that would not correctly predict the result on the definitive endpoint?

Well, here's an attempt in explaining that concept, and if you get nothing else from this talk, I hope you will understand the examples that I have on this slide.

So let's suppose that with some baseline therapy, 20 percent of patients will meet the early endpoint and 80 percent will not, and at meeting that endpoint is a favorable prognostic indicator

for the definitive endpoint of event-free survival. So here we might be tempted to say that treatment A is superior to treatment B because it results in more patients achieving the favorable endpoint. So if you compare the green bars there for treatment A versus treatment B, you see that treatment A results in 60 percent, meaning this favorable endpoint, and treatment B achieves only 40 percent.

But there is a potential flaw in the logic as illustrated by the rows of this table. That logic assumes that the responders under treatment A will behave the same as the responders under treatment B with respect to event-free survival. If that is the case, then treatment A will show superiority like scenario 1 in the table. But there's no reason that this has to be the case. The two treatments might have effects other than through the early endpoint. Those effects could be different and translate to different impact on event-free survival.

So if you look at scenarios 2 through 4, you can see how the early endpoint could remain

prognostic across the board, but the effect on event-free survival, comparing treatment A to B, as shown in the last column highlighted in orange, could go either direction. I could conclude that either treatment A is better or B is better, just depending on how that relationship between the intermediate endpoint and the definitive endpoint plays out for the two different treatments.

Dr. Amatya will talk more about the points on this slide, but I do want to remind everyone that you need to think carefully about how you conduct a trial-level meta-analysis to empirically validate an endpoint as a surrogate. You need to specify, first of all, what is the clinical benefit measure of interest? What is the endpoint that you want to replace, and how are you quantifying the treatment effect on that endpoint?

I would remind people that when you're using time-to-event endpoints, generally a randomized trial will be required unless you're dealing with a very, very good risk group.

The method of measuring the surrogate is

important, what we've heard today about two different versions of the end-of-induction response criteria that have changed from '93 to 2017. The class of drug could be very important. As I mentioned, the mechanism of drug may be very different, and can really lead you astray in you're thinking. The patient population may be very important, and whether there are biologically defined tumor subtypes for which the drug might behave differently. Remember that extrapolation to a new class of drugs or patient population not covered by the meta-analysis can be risky, so think carefully about how you set it up.

Here's an example of a trial-level meta-analysis conducted for another pediatric cancer. This study looked at end-of-induction minimal residual disease as a candidate trial-level surrogate for event-free survival in B-ALL. The analysis included 4830 patients from two large randomized phase 3 trials that were asking a question about different corticosteroids, specifically dexamethasone versus prednisone,

during induction therapy.

MRD was assessed as a 3-level variable: negative, low-positive, or positive, as you can see defined at left. In the COG trial, patients were also randomized to receive either Capizzi or high-dose methotrexate regimens, in addition to the corticosteroid randomization, in a 2x2 factorial design.

The figure to the right shows event-free survival curves by treatment within minimal residual disease categories for the following patient groups: A is for the overall group; B in the European trial; C in the high-dose methotrexate COG group; and D in the Capizzi group. So there's a very strong prognostic effect of end-of-induction response that's clearly evident from these plots.

But in order to conduct a trial-level meta-analysis, centers within each trial were grouped according to geographic region to define many trial units. Ideally in a meta-analysis you would like many different trials, and people often recommend at least 10, but there just weren't that

many in this particular disease area.

The groupings also accounted for chemotherapy regimen. Within each trial unit an odds ratio for treatment effect on MRD and a hazard ratio for treatment effect on event-free survival were calculated, and these were plotted as you see on the right. Each point corresponded to a trial unit, number labels on points refer to sample size, and shading indicates the chemotherapy regimen.

The association between treatment effect on MRD, which is the X-axis, and the treatment effect on event-free survival, which is on the Y-axis, was poor, yielding an R squared of 0.09. So MRD was not validated as a trial-level surrogate for event-free survival in this example.

There are many reasons why an early endpoint might fail to validate as a trial-level surrogate. First, the early endpoint maybe is not capturing the relevant biology, and there was a comment made this morning about looking in the different disease compartments. There are many components to the end-of-induction response criteria. So maybe it's

important to be looking in the bone marrow, or in other metastatic sites, and not just other more local measures of disease. I often felt that the reason many of these kinds of response criteria don't work is many of them are looking at the primary tumor site, and the tumor cells that tend to be the bad actors and harm the patient are the ones that metastasize.

So not measuring early endpoint in the best way or the right time is another possibility. In general, the closer the measurement of the surrogate is, or the candidate surrogate is, to the definitive endpoint, the greater the chance it's going to be a good surrogate because lots of things can happen in between; for example, effects of therapies delivered after measurement of the early endpoint, and especially if those therapies are chosen based on observation of the early endpoint.

I already mentioned the value of early endpoint could depend on biological subtypes of tumors. We need to potentially restrict to a particular class of therapeutic inventions. Maybe

it works differently for targeted versus non-targeted, et cetera. And we could simply just not have enough trials, or the trials are too small, or we don't have enough range of treatment effect; lots and lots of reasons that I'm sure Dr. Amatya will expand upon.

In conclusion, I think we have to really sit down and carefully define the intended role for the early endpoint. People too often start talking about surrogates. There are many other ways earlier endpoints can be used in a productive way in drug development program, so make sure you know where you're aiming before you shoot.

Plan ahead to collect the right evidence to support that intended role, and this may require harmonizing measurements of the early endpoints; identifying sufficient number of trials in the relevant patient populations with the right drugs, et cetera. Surrogacy analyses typically need randomized trials with early endpoints measured after delivery of treatments of interest. There have been many other efforts in adult cancers where

people have tried to do meta-analyses looking for surrogacy, and they didn't even include randomized trials in their set of trials.

It's important to appreciate that premature adoption of a reasonably likely surrogate may also thwart efforts to complete ongoing phase 3 trials designed to assess a true definitive endpoint, so don't jump too quickly. Thank you very much.

FDA Presentation – Anup Amatya

DR. AMATYA: Good afternoon. I'm Anup Amatya, currently the acting lead mathematical statistician at FDA, Division of Biometrics V. Dr. McShane discussed many of the key considerations in the validation of early endpoint to support drug development.

From a regulatory perspective, the fundamental question when using early endpoint is whether the decisions based on such endpoint would be the same had we waited for a trial to meet the definitive clinical endpoint. By definitive endpoints, I mean those endpoints that directly measure how a patient feels, functions, or

survives, but there is overall survival or long-term clinical benefit endpoints, so there's event-free survival in the setting of neuroblastoma.

Additionally, to rely on an early endpoint for regulatory decision making, we also need to have adequate data to understand the relationship between the magnitude of observed treatment effect on an early endpoint and a meaningful improvement in definitive endpoints.

The degree of uncertainty that is acceptable in the answers to these questions depends on the context in which the early endpoint is being considered for regulatory use, including the regulatory pathway being pursued for marketing approval.

There are two ways in which surrogate or intermediate endpoints may be used to support marketing approval of a drug or a biologic. If a surrogate is validated and shown to be a reliable predictor of clinical benefit, that endpoint may be used for regular approval.

The second pathway is accelerated approval, which allows the use of intermediate endpoints that are reasonably likely to predict clinical benefit. I note that intermediate endpoint is the term that we use interchangeably with early endpoint.

As this pathway leaves some room for uncertainty, confirmatory evidence of clinical benefit is typically required in order to support regular approval after accelerated approval is granted. But before using an early endpoint to support regulatory decision making, the strength of evidence supporting the ability of the endpoint to predict clinical benefit needs to be evaluated.

As Dr. McShane pointed out, it is not enough for an endpoint to be prognostic for that endpoint to be a reliable predictor of clinical treatment effect on the definitive endpoint. The use of Prentice criteria to establish reliability, while theoretically appealing, is likely to be impractical. The current approach to early endpoint validation relies on meta-analytical methods to work around this difficulty with

Prentice criteria.

Commonly discussed meta-analysis methods are listed here. The last two methods listed have been used by FDA for evaluating early clinical endpoints of interest. As we have noted, all these approaches require data from multiple randomized trials, and some of them also require individual patient-level data. This is one of the major challenges in the validation of early endpoints.

As the validation of an early endpoint can be complex, FDA has two mechanisms through which the agency can provide feedback on the development and use of novel endpoints to support approval, firstly, through FDA's formal drug development tool, the qualification program. The formalized process and steps to follow are in the FDA guidance, which is provided at the website listed at the bottom of this slide.

The second mechanism is to discuss them with a specific regulatory review division. An example of an early endpoint that uses the second regulatory mechanism is completed at 30 months in

follicular lymphoma. I use this example to illustrate how the validation is accomplished and to highlight two important caveats when using early endpoints validated through the meta-analysis approach.

Progression-free survival, or PFS, is a generally used endpoint for assessing the efficacy of a new drug in first-line follicular lymphoma, but the expected median PFS in the first-line setting for follicular lymphoma is long. It's over six years, and it is continuously improving with new therapies. A potential early endpoint was considered to facilitate drug development in this disease.

In 2015, the Follicular Lymphoma Analysis of Surrogate Hypothesis group, which is also called FLASH, explored a utility of 30-complete response, which is also indicated as CR30 on this slide, as a surrogate for PFS in first-line follicular lymphoma trial.

Thirteen studies were selected. Eight of these studies were induction trials, shown with

triangles on this figure, and five were trials that included maintenance treatment, shown in circles. Nine trials incorporated rituximab in at least one of them, shown in gold color shapes, and four trials did not include rituximab, shown in blue color shapes.

There was a total of 3,837 evaluable patients across these trials. The relative sizes of these trials are represented by the sizes of the shapes from this figure. The analysis was conducted using two meta-analysis approaches that used individual patient data.

The meta-analysis demonstrated consistent results with both methods for a trial with measurement of surrogacy. The point estimate of the R square, which is the indirect measure of correlation, was 0.88 with corresponding 95 percent confidence interval 0.77 to 0.96 using weighted least square approach, and 0.86 with confidence interval of 0.75 to 1.0, based on the second approach that also took account of patient-level correlation between the two endpoints. Sensitivity

analyses were also conducted, and adults were mostly consistent with the primary analysis results.

Overall, as can be seen on this figure, the results appear to support the use of 30 months complete response as an early endpoint in patients with previously untreated follicular lymphoma. However, some caveats must be kept in mind when considering the use of an early endpoint validated in this manner.

First, in this specific example, the trials through meta-analysis evaluated the use of cytotoxic agents and rituximab, and therapeutics that were via a different mechanism of action may impact responses differently than the traditional cytotoxic agents, which may in turn influence the correlation between response rate and progression-free survival. As such, 30 months complete response rate should only be used in trials of therapeutics that have a similar mechanism of action and can be expected to have similar response patterns.

Second, the majority of patients included in this meta-analysis had a 3 or 4 in intermediate or high-risk disease. Among patients with early-stage disease and those with low or intermediate risk scores, correlations were weak. As such, the use of 30-month complete response rate as an early endpoint may only be appropriate in future trials that enroll a similar patient population to those in meta-analysis. Their data also suggested that 30-month complete response may not be an adequate surrogate for patients with low to intermediate risk scores or stage 1 to 2 disease.

Now, circling back to neuroblastoma, what is needed to development an early endpoint? First and foremost, the early endpoint for consideration must be biologically plausible. Some data seem to exists to support biological plausibility of end-of-induction response. Regarding standardization, a consensus on response criteria seems to have been agreed, and that will be helpful in this process, at least going forward.

One of the challenges, perhaps, is the

limited number of randomized clinical trials available in this setting. To validate a trial of correlation, as Dr. McShane mentioned in her presentation, there should be an adequate number of randomized clinical trials that have captured end-of-induction response and a definitive endpoint, such as EFS. Appropriate statistical analysis of course should be performed to establish a reasonable really strong correlation between improvement in early endpoint and the improvement in how patients feel, function, or survive.

Additional consideration should be given to potential confounding of surrogacy; if the intensity or treatment regimen is in consolidation or maintenance depends on response to induction therapy; and even if the validation process is successful, we need to acknowledge the limitation that these methods are highly context dependent and contingent on disease, stage, patient population, and therapy that are used in the trials included in the meta-analysis.

Early interaction with FDA is going to be

important to ensure that a drug development program meets the general statistical considerations for validation of candidate early endpoint; discussion and agreement on definitions of endpoint; details of the trials to be included in meta-analysis; and a detailed analysis plan will be important in this process.

In summary, an accelerated approval program may be used to expedite approval for serious life-threatening disease based on early endpoints that are reasonably likely to predict clinical benefit. The candidate early endpoint such as end-of-induction response should be validated to show that it is reasonably likely to predict clinical benefit, which generally requires multitrial approach.

The collaboration and cooperation between all stakeholders and early planning of future trials, including the ones conducted by academic investigators, will be crucial to fully utilize a limited number of trials that are feasible in pediatric diseases such as high-risk neuroblastoma.

Additional research in methodology and perhaps exploration into alternative data sources may also be needed to overcome the limitations posed by current reliance on multitrial approach. Thank you for attention.

Clarifying Questions

DR. PAPPO: Thank you, Dr. McShane and Dr. Amatya, for your excellent presentations.

We will now take clarifying questions for Drs. McShane and Amatya. Please use the raise-hand icon to indicate that you have a question, and remember to clear the icon after you have asked your question. When acknowledged, please remember to state your name for the record before you speak and direct your question to a specific presenter, if you can. If you wish for a specific slide to be displayed, please let us know the slide number, if possible.

Finally, it will be helpful to acknowledge the end of your question with a thank you, and end of your follow-up question with, "That is all for my questions," so we can move on to the next panel

member.

We have a question from Dr. Steve DuBois.

DR. DuBOIS: Thank you, Dr. Pappo. Steve DuBois from Dana-Farber. I have a question for Dr. McShane and a question for Dr. Amatya.

For Dr. McShane, I really enjoyed your presentation, and I just want to highlight and dig a little bit more deeply into this issue of insufficient range of treatment effect. We understand from Dr. Pinto that about 80 percent of patients will have an end-induction partial response or better, and I wonder what your thoughts are on the challenges if we were to start using end-induction response as a surrogate endpoint.

Does that actually make it more difficult for us to run our trials as compared to an EFS endpoint, which currently is at about 50 percent?

DR. McSHANE: Thanks for that question. Yes, in fact it does make it more difficult. If you have a very limited range of treatment effect on either your surrogate, your candidate surrogate, or your definitive endpoint, you don't really have

much room to play. To establish a correlation, you need to show that things on the very low end on one endpoint correspond to low on the other endpoint, and vice versa on the high.

This has been one of the criticisms of some of the meta-analyses that have been done in adult solid tumors; that you really need to have that variation in order to detect a correlation. So it could be that a little bit later early endpoint might actually work better than end of induction. Furthermore, between end-of-induction and a more long-term endpoint, there will be a lot of things that happen in there. Treatment may have been altered based on the patient's response or non-response, or other supportive therapies. Lots of things can happen.

Many of the success stories -- and it's interesting. Even with the lymphoma example that Dr. Amatya gave, he was using a 30-month endpoint, which is going to be a whole lot closer to that long-term endpoint. That actually gives you a big advantage in terms of being able to establish

surrogacy. In adjuvant colorectal cancer, I think they've shown 3-year disease-free. I could be a little wrong on this, but a 3-year endpoint is a pretty good surrogate of overall survival, so that's definitely an important consideration.

DR. DuBOIS: Yes. Well, it almost makes me -- I had come into this thinking that favorable response is being a partial response or better, but I almost wonder if the bar needs to be -- if we opt to move in this direction, complete response or better, where we know that only 20 percent of patients have an end-induction complete response. So it's interesting.

DR. McSHANE: Yes. And I'll just add one more thing. I do think it will be important to look at the various components of the response criteria for the reasons I mentioned during the talk. It could be that what's really giving a bad outcome for the patient is the disseminated tumor cells.

DR. DuBOIS: Yes.

DR. McSHANE: I would definitely, if I were

doing this one, look at those separate compartments -- bone marrow and metastatic sites -- and not just take the overall.

DR. DuBOIS: Yes. Thank you for that.

For Dr. Amatya, I think the methodology that you presented seemed really enviable, and I think maybe the challenge for us in neuroblastoma is that most of our recent high-risk neuroblastoma trials have actually not focused on induction -- or I should say completed trials have not focused on induction randomization.

A lot of the questions have been focused on the consolidation phase or post-consolidation phase. It's only our two ongoing studies, one in the COG and one in SIOPEN, that are looking at induction question. I worry that the very nice methodology that you showed, we won't actually be able to do that type of a meta-analysis for a long time.

So it's a long way to get to my question, which is, at what point would we be able to say that we've repeated an analysis showing a strong

correlation between end-induction response and subsequent outcome? When can those correlations, or enough of them, be accepted as a surrogate without the data available to do the type of meta-analysis that appears to be the preferred approach?

DR. AMATYA: Thank you for that question. That's one of the challenges, especially in this disease area. I think that's the challenge in some of these areas, too, with the different kinds of surrogate endpoints, the potential surrogate endpoint.

The question is when do we know that we have enough information? And that I think is at a point where we can discuss it in the context of the application and the information that we have at that point, and not only the potential surrogate endpoint, but also other information from other trials at that point.

I don't have a straight answer to that, basically, because if we want to follow the statistical threefold, then we to have some of

these methodologies, and the straight alternative path isn't available at this point. It will have to be discussed with the interested party, these endpoints, to see if there's any path to move forward.

DR. DuBOIS: Yes. Thank you for that.

Nothing further for me, Dr. Pappo.

DR. PAPPO: Question, Mark Conaway?

DR. CONAWAY: Yes. Mark Conaway, University of Virginia.

Thank you, Dr. McShane and Dr. Amatya, for those very interesting presentations. I had two technical and one much more general. The technical question is, in those analyses of complete or partial response, or early endpoint versus event-free survival, did they take the time element into account?

For example, I presume -- though I didn't really see the definitions -- you need to be event free long enough to be declared a responder. Was that taken into account at all by like a landmark analysis or anything like that?

DR. McSHANE: Yes. If I could answer that one. I think you're referring to the examples I presented, probably. It's an excellent question, a very astute observation by a statistician.

One of the analyses did use the time, I think, of randomization or start of induction as the time point. The other one did use more of a landmark approach. And as you know well, it's a little bit difficult to know exactly what to do there, because as you pointed out, when you have a patient who didn't make it to the landmark time point -- in this case, around 30 days or so for end of induction -- well, no, I guess it's longer than that in neuroblastoma. But you have a patient who didn't make it to that point, then you can't even really measure that outcome, or if they drop out and you don't know what happened to them, you're sort of predicting -- you're using something that happened in the future to divide the patients into subgroups.

I think it was in the trial that used the baseline, there actually was not a lot of drop out

until that early endpoint time, and I don't recall the other one. The other one did use more of landmark analysis. But when you use a landmark analysis, what Dr. Conaway is referring to, is you basically drop all the patients who didn't make it to the end of induction -- in this case -- and then you say, starting from that point, what is their residual outcome?

There's not really a perfect way to deal with that statistically unless you do something fancier like use the time-dependent covariate analysis. I think that the results in those cases were probably strong enough that it wouldn't have made much of a difference in the net conclusion that there that was prognostic ability, but it's definitely a statistical fine point that needs to be looked at, and should be part of any analysis plan that you develop for a meta-analysis.

DR. CONAWAY: Yes, thank you. Yes, that was mostly clarifying as to whether that was done. And I agree completely; these analyses are really challenging, and there is no perfect way to deal

with it.

DR. McSHANE: Right.

DR. CONAWAY: The second question is much more general. I really appreciated your point that most of the work on surrogates treats the surrogate as a direct substitute for the definitive endpoint.

Are you aware of work that's been done on maybe multiple endpoints? Because usually the definitive endpoint is being collected anyway. So has there been work done, or not, on a direct substitution but using the intermediate endpoint as primary, supplemented by perhaps a lower bar of information for treatment differences on the definitive endpoint; or perhaps not a one-to-one substitute of surrogate for definitive, but maybe multiple surrogate endpoints would be more predictive of a definitive endpoint?

Are you aware of any work done in that area?

DR. McSHANE: Yes. I don't know if that's directed at me or Dr. Amatya. I can take a first crack at it.

DR. CONAWAY: Either.

DR. McSHANE: I've certainly heard of people working on analyses where they're even using machine learning approaches to take all kinds of things into account when trying to predict that longer term endpoint. So, in theory, it seems like something you could do.

I guess there may be regulatory issues with regard to how such an approach would be reviewed and how you decide to combine things. It's something you probably would want to specify up front, or you'd want to have a real clean separation between the development of such an endpoint and the eventual validation. You don't want to be doing the combining on the fly because you're likely to just come up with spurious things.

But I don't know if Dr. Amatya has anything else to that.

DR. AMATYA: I also have not seen a proposal like that, looking at multiple potential surrogates for multiple endpoints. But we do have an accelerated approval pathway, and that seems to be an approach where you've had an effect on that

reasonably likely surrogate, and then pursue then or do a different second trial, and then try the earlier signal maybe a few years later. But that's one pathway I can see if you use not a replacement endpoint, but a reasonably likely early endpoint, and that could be discussed in the regulatory setting.

DR. McSHANE: My impression, Dr. Amatya, if you could comment on this, is that in many kinds of decisions, there are secondary endpoints that are given consideration as to whether they are supportive or not of a result on a primary endpoint, so I think informally that happens. Whether anybody has formally proposed a combined sort of endpoint and tried to validate it as a surrogate, that I'm not so sure.

DR. AMATYA: [Indiscernible].

DR. PAPPO: I had a question for Dr. McShane, and it goes back to one of her slides. I think it was slide number 9. I don't know if they can put up slide number 9, but it's when you designed a clinical trial for end-of-response

assessment, and you had those that had a good event of response and those that did not have a good end-of-induction response.

I had two questions. The first one is, why would you take those patients off study that had a good end-of-induction response? Wouldn't that group provide important information regarding validating your prognostic factors for those who respond or do not respond, and would serve as a good control to be sure that your therapy is as effective as you thought it was going to be initially when you compare it to the experimental, to the standard amount of patients that did not respond?

The second question was, when you have the end-of-induction response, non-responders when they are randomized, you're going to run into relatively small numbers of patients, especially if our therapy gets better and better. For example, when you introduce chemoimmunotherapy up front, would you consider doing some modifications in the way you're going to analyze the data; for example,

relaxing the alpha even though it's not going to be as strong, but just to provide some sense of this new experimental therapy is working or not?

Those were my two questions.

DR. McSHANE: Okay. Thanks.

Yes. So the off-study part, I guess it's a question of whether you're actually continuing to look at those patients and answering an investigational question. With off study, the thought is that you're probably giving them whatever you would normally give, and they're already showing signs of responding favorably to the usual strategy for treatment. You're correct that you could certainly choose to follow those patients to see if their outcome remains as good as you think it should be. It's just a matter of what you're really defining as your clinical trial question.

We do have a number of studies where we use these enrichment strategies, and for lack of resources, often we can't afford to follow all the patients who didn't make it into the enriched

subgroup. But you make an excellent point, that there may still be value in some situations and following those patients.

Your other question was about the fact that as our treatments get better and better, or at least better at producing responses, that this group could become very small. Yes, that is a real challenge, and in fact a challenge -- this is kind of what personalized medicine is all about, right? I mean, the better we get at tuning therapies, the smaller the group will be that needs to have something better than the existing therapies.

We have trials even with biomarker-guided enrichment for targeted therapies. For example, we have adjuvant 1 trials right now where the rate of patients who make it into the enriched group and get randomized might be as low as 10 percent or 5 percent, and it becomes extremely difficult, especially in a rare disease, to conduct trials that way.

So then you kind of get into the question of, well, when you're in that rare disease setting,

or at least a rare subset within a disease, should we relax some of our usual statistical criteria, like relax the alpha level thing? We're going to be happy if we can use an alpha 0.10 or 0.15. That's really a question that the whole community has to deal with. What is our tolerance level for making a mistake? And sometimes we do have to settle for a lesser confidence in the outcome.

DR. PAPPO: Thank you. That answers my question.

We have a question from Dr. Mishra-Kalyani.

DR. MISHRA-KALYANI: Hello. This is Pallavi Mishra-Kalyani from FDA statistics. Actually, no; I was trying to add responses to some of the prior questions. I guess I can just give a quick note now, if that's ok.

To the previous question regarding ethical endpoints being validated together or used together for the definitive endpoint, I just wanted to point out that in the context of the regulatory review process, we do somewhat follow this method of looking at several endpoints regardless of the type

of endpoint that is the primary endpoint in the trial.

So regardless of whether we're using an early clinical endpoint or a definitive clinical endpoint for the trial, we do look at various endpoints as supportive evidence, and we'd like to see that there is benefit demonstrated on various endpoints to ensure that what we're seeing is truly robust, or the treatment benefit that we're observing is truly robust and documented by several different mechanisms or different markers.

I was just trying to add that to the previous response. Thank you.

DR. PAPPO: Thank you very much.

We have one last comment from Dr. Donoghue, and then we'll go to the OPH session.

DR. DONOGHUE: Thank you, Dr. Pappo. I hope I lowered my hand.

I just wanted to make sure Dr. Mishra-Kalyani had a chance to chime in, and she has, so thank you.

Open Public Hearing

DR. PAPPO: Thank you so much.

We will now begin the open public hearing session. Both the FDA and the public believe in a transparent process for information gathering and decision making. To ensure such transparency at the open public hearing session of the advisory committee meeting, the FDA believes that it is important to understand the context of an individual's presentation.

For this reason, the FDA encourages you, the open public hearing speaker, at the beginning of your written or oral statement to advise the committee of any financial relationship that you may have with a sponsor, its product, and if known, its direct competitors. For example, this financial information may include a sponsor's payment of your travel, lodging, or other expenses in connection with your participation in this meeting.

Likewise, the FDA encourages you, at the beginning of your statement, to advise the committee if you do not have any such financial

relationships. If you choose not to address this issue of financial relationships at the beginning of your statement, it will not preclude you from speaking.

The FDA and this committee place great importance in the open public hearing process. The insights and comments provided can help the agency and this committee in their consideration of the issues before them.

That said, in many instances and for many topics, there will be a variety of opinions. One of our goals for today is for this open public hearing to be conducted in a fair and open way, where every participant is listened to carefully and treated with dignity, courtesy, and respect. Therefore, please speak only when recognized by the chairperson, and thank you for your cooperation.

Speaker number 1, your audio is connected now. Will speaker number 1 begin and introduce yourself? Please state your name and any organization you are representing for the record. Thank you.

DR. ZELDES: Good afternoon. Thank you for the opportunity to speak today on behalf of the National Center for Health Research. I am Dr. Nina Zeldes, a senior fellow at the center. We analyze scientific data to provide objective health information to patients, health professionals, and policymakers. We do not accept funding from drug or medical device companies, so I have no conflicts of interest.

Our statement today is based on our organization's experience of working with thousands of patients and caregivers. We understand that patients and their parents urgently want new treatments for these terrible cancers and think they are willing to take almost any risks if a new treatment might possibly be effective, but they feel very differently when treatments do more harm than good. Of course, nothing is worse for a parent in making a medical decision that harms a child without providing meaningful benefits.

We agree with FDA's statement in its memo that randomized trials, quote, "have a continued

role in generating the evidence needed to improve treatment paradigms for patients with high-risk neuroblastoma despite the challenges associated with enrolling sufficient numbers of patients in a timely fashion and the length of time needed to conduct trials," unquote.

We note that the FDA points out that endpoints traditionally used to evaluate effectiveness of drugs for first-line treatment of patients with high-risk neuroblastoma are evident-free survival, typically defined as the time from randomization to the first recurrence of relapse progressive disease; secondary malignancy or death; and overall survival.

These are the most appropriate endpoints for two reasons. First, cancer treatments always have the potential of resulting in serious risk to quality of life. Parents of these children need as much information as possible when they decide what treatments to accept for their children. Any kind of surrogate endpoint that does not involve improved survival or improved quality of life has

the potential to have risks that outweigh the benefits.

Second, once a pediatric cancer treatment is approved, it can be very difficult, if not impossible, to conduct well-controlled postmarket studies to confirm whether the benefits outweigh the risks. Randomized trials with placebo are often impossible.

We acknowledge that researchers have emphasized the importance of identifying the most effective treatments to use during the induction phase of treatment in order to improve patient outcomes. It may be that end-of-induction response may predict event-free survival or overall survival, but that is not yet clear. In addition, the relationship between end-of-induction response and adverse events is also unknown. These are important to study, but meanwhile, end-of-induction response is not an adequate endpoint for these very important treatments.

Clinical benefits should remain the key endpoints for approval decisions of these

treatments. Surrogate endpoints that predict clinical benefits are not yet established, and until they are, we are concerned about their use as secondary endpoints unless the primary endpoint is also met. Thank you for your time.

Questions to the Subcommittee and Discussion

DR. PAPPO: Thank you very much, Dr. Zeldes.

The open public hearing portion of this meeting has now concluded and we will no longer take comments from the audience. The committee will now turn its attention to address the task at hand, the careful consideration of the data before the committee, as well as public comments.

We will proceed with the questions of the committee and panel discussions. I would like to remind public observers that while this meeting is open for public observation, public attendees may not participate except at the specific request of the panel. We will start with question number 1 from the FDA.

Would you please read the question?

DR. DONOGHUE: Question number 1. Please

discuss the potential benefits and limitations to using an intermediate clinical endpoint in the evaluation of new drug under development for the first-line treatment of patients with high-risk neuroblastoma.

DR. PAPPO: While everybody's getting their hands up, I could potentially start.

Based on our discussions, the potential strength of using this would be that perhaps you could treat all this population in a clinical trial and help better identify the factors that are associated with a poor response to induction therapy if you collect all the genomic or the clinical data in an organized fashion.

I think that this could potentially help expedite the testing of new drugs or drug combinations for this population of patients, which is really the ones that you want to target, the ones that are not going to respond, and develop resistance.

I think that, also, this could allow you to identify those drugs or drug pairs that could

potentially be inactive. The only issue is whether -- as has been raised before -- this PR interim introduction therapy is the right endpoint. Would a CR be better, or would an earlier response, for example, after two cycles, be a better predictor of outcomes, and could potentially help, quote-unquote, "salvage" those poor responders?

Those were my considerations, and now we have Steve.

DR. DuBOIS: Steve DuBois from Dana-Farber, Boston Children's. I think one potential benefit that I don't think has come up yet is that in the course of designing our clinical trials, we often are flying a bit blindly when we are designing our successor trial because we don't have, really, any idea about how the current trial is looking. And if we had an earlier readout that didn't require three years of follow-up, then that may allow us to be a bit more nimble in designing our successor trials.

DR. PAPPO: Thank you very much, Steve.

Ro Bagatell?

DR. BAGATELL: Hi. This is Ro Bagatell from the Children's Hospital of Philadelphia. While I agree with Steve about the need to be nimble, I would add to that comment by saying, when we recently mapped out our timeline for answering important questions in neuroblastoma, it became apparent to us that some of the pressing questions in high-risk neuroblastoma therapy will not be answered until the late 2030s if we go at our current pace.

So I completely agree with Steve about our need to be nimble, but I do think that there are limitations we just need to be aware of. And one of those is that the data that we have are primarily based on studies of cytotoxic agents in induction, and it's just hard to know how those data apply and what we can say about the layering on of additional agents of other classes on top of cytotoxic induction, and the impact that they would have on a marker like end-induction response as a proxy for more distant endpoints.

DR. PAPPO: Thank you.

Dr. Conaway?

DR. CONAWAY: Yes. Mark Conaway, University of Virginia. Yes, I agree, the benefit is a quicker assessment and quicker evaluation of therapies.

One question I had -- and I certainly have no answers on this -- is the slides that showed event-free survival and end-of-induction therapy response, it looked like the survival curve separated out really early, and dramatically early.

It seems like if you saw an early signal in terms of end-of-induction therapy, you would also be seeing, to some degree, a difference in event-free survival. Maybe the data wouldn't be mature enough. Maybe there wouldn't be statistical significance, but it does seem to me that you would be seeing signals on both of those endpoints.

So it isn't clear to me, because of the length of the induction therapy, that the intermediate endpoint is really going to save all that much time. It will save some time in the assessment of therapies, but it isn't clear to me

how much.

DR. PAPPO: Any additional comments? Is there consensus of the group that one of the potential benefits could also be to potentially identify early new promising drug pairs or single pairs that could potentially be incorporated into this high-risk population of the diagnosis?

(No response.)

DR. PAPPO: Any comments on that?

(No response.)

DR. PAPPO: Okay. If not, I'm going to go to Ted Laetsch.

DR. LAETSCH: Thank you. Ted Laetsch.

Alberto, I agree with your thoughts that that is certainly a potential benefit, and agree with the other members around this. I think recognizing some of the comments by the statisticians and the community representative, I think it is important to think through the importance of continuing to gather the longer term survival endpoints, as well on these trials, and wonder if the FDA can use some of its regulatory

discretion through things like accelerated approvals and requirements for subsequent confirmatory data to alleviate some of those concerns, while still allowing more rapid drug development for this patient population that's clearly in need.

DR. PAPPO: Thank you, Ted.

Dr. Kraus?

(No response.)

DR. PAPPO: Dr. Kraus, did you have a comment?

DR. KRAUS: Sorry. I was double-muted.

Can you hear me now?

DR. PAPPO: Yes.

DR. KRAUS: Yes. Sorry.

I was going to comment on your comment, Dr. Pappo. I think having therapies getting to patients earlier, particularly when there's very severe prognosis, including, almost uniformly, early mortality, is a big advantage for patients. Obviously, you don't want to get therapies there wrongly without an adequate benefit-risk, but that

time can be especially precious.

So I would support the concept that while the survival curves did separate reasonably early, those curves were over many years and decades with 5, 10, 15-year periods on those Kaplan-Meiers in the index, so it's still years. I think that's very important for patients.

The second part really refers to, I think, the comment somebody may about regulatory process, and I look back to, for instance, how a number of approvals and confirmations of approvals were made on an accelerated approval basis, and then a confirmation of that basis in chronic myelogenous leukemia, and I was involved in several programs there.

Earlier data yielding accelerated approval to allow the drug to get out there, and then longer term follow-up with the same single-arm trial with more information from various endpoints, but predominantly more surety around original endpoints and durability of such served to confirm, could be something that could be considered, particularly

when it's really hard to mount randomized trials, et cetera, et cetera. That's just my comments for consideration. Thank you.

DR. PAPPO: Any other additional questions and comments before I summarize our discussion for question number 1?

Julia Glade Bender?

DR. GLADE BENDER: Hi. Thank you, Dr. Pappo. Julia Glade Bender from Memorial Sloan Kettering. While I appreciate the potential statistical limitations, I was very struck by Leona Knox's plea on behalf of the patients, which is that in the absence of a trial, I think there is real disillusionment amongst the patient to population.

I think the statistical potential limitations in many ways are not the factors that they are thinking about. What it takes in terms of phase 1 and phase 2 research to actually even be considered to move up front in a phase 3 pediatric clinical trial is substantial. So I think the limitation is that we would come up with the wrong

conclusion that we'd ultimately find out when we looked at event-free survival. But then the question was, was that time lost or wasted because we had something better we could be doing?

I think when we look in sum, the benefits probably outweigh the limitations. There is a chance we would get it wrong, but I think there's also, based on all of the preponderance of evidence, more of a chance that we might get it right, and sooner.

DR. PAPPO: Thank you very much, Julia.

If I can summarize this, the panel believes that there are certainly some potential benefits to use this intermediate clinical endpoint for evaluation of new drugs in neuroblastoma. Some of the potential benefits include that you could potentially identify new drugs or new drug pairs that would target the patients that have the biggest risk for relapse, and this could be incorporated earlier into the treatment of patients with high-risk neuroblastoma.

Another benefit would be that an early

readout could potentially help plan successor trials earlier. The endpoint of end-of-induction therapy response may have some limitations based on the data that we currently have, which is mostly based on cytotoxics. That is unclear if with newer therapies this could potentially change.

It also is important to consider that we need to continue to look at long-term survival as an endpoint and that this could also provide means for better analyzing the data in clinical trials that are using this endpoint. As Julia said, one of the main drawbacks of this is that perhaps in the end, we got it all wrong, and really this does not correlate with outcome, and we could have potentially wasted, quote-unquote, "some time" doing this clinical trial without potential benefit for the patients. But based on the cumulative data that we have, it appears that the benefit might outweigh the limitations of this potential drawback.

Did I get it sort of right? Did I miss anything?

(No response.)

DR. PAPPO: I think it was great.

Okay. Let's go to question number 2.

DR. DONOGHUE: Thanks, Dr. Pappo. This is Martha Donoghue. I think some of this discussion may have already occurred under question 1, but I'll go ahead and read it.

Please discuss the strength of the evidence for using end-of-induction response as a prognostic factor and to assess antitumor activity of investigational treatments during the induction phase of treatment.

DR. PAPPO: Yes. I think some of this was addressed in question number 1, but I look forward to additional comments.

I don't, Julia, if you just forgot to put your hand down or if you have a comment.

DR. GLADE BENDER: I've taken my hand down.

DR. PAPPO: Thank you.

Anyone else would like to add to question number 2?

(No response.)

DR. PAPPO: Okay. So it appears that most of the answers to this question have been adequately addressed in question number 1, and we will move to question number 3.

DR. DONOGHUE: Please discuss how end-of-induction response is used in clinical decision making and the implications of its use in the design and conduct of clinical trials investigating new treatments for patients with high-risk neuroblastoma.

DR. PAPPO: If there are no questions or comments concerning the wording of the question, we will now open the question for discussion.

We have Steve.

DR. DuBOIS: Steve DuBois, Dana-Farber. I think I'll just highlight one of the points that Dr. Pinto made in his description of the ANBL 2131 proposed clinical trial, which would allow patients with an inadequate end-of-induction response to actually stay on study and be systematically followed, which is a bit of a departure from how we've done things in prior trials, where if a

patient elects to come off protocol therapy due to inadequate response, if they go on to one of our COG relapsed or refractory trials, we can track their outcomes and understand more fully what treatments they received, but otherwise, it's a very heterogeneous approach historically to those patients. So I think it will be really valuable to have those patients treated uniformly and tracked uniformly in the context of the same trial.

DR. PAPPO: That is a great point.

Anyone else that would like to comment on question number 3?

Dr. Kim?

DR. KIM: Hi. This is AeRang Kim from Children's National, and I agree with Dr. DuBois. And just to add, I think if we're using this end-of-induction response, currently if patients are coming off therapy because they've progressed, if we're using this as an early endpoint in the design of future trials and there can be a way forward for those patients in the future trials, and those that have progressed are now going to get

treatment A, B, or C because we know that there are going to be poor responders, that will be an opportunity to use that and to have those patients remain on clinical trials, and opportunity, as discussed before, to have new therapies and input at that time, which I think is a benefit. Thank you.

DR. PAPPO: Thank you very much.

Dr. McMillan?

DR. McMILLAN: Gigi McMillan, Loyola Marymount University. I'm a bioethicist and a patient advocate, and I just wanted to comment that with regards to the design of clinical trials using end-of-induction response, this is an example of a creative response or a creative strategy for a group of patients for which there has little been done in a timely manner, and it's not from science not trying its best.

But we have new ways of thinking about data and new ways of correlating what's happening in trials that are already in existence. These examples of a creative new design, this is an

ethical response for these patients and these families who don't really have much hope or much at their disposal at this time?

DR. PAPPO: Thank you very much.

Dr. Bagatell? Dr. Bagatell, go ahead.

DR. BAGATELL: Thank you, Dr. Pappo.

This is Ro Bagatell from the Children's Hospital of Philadelphia. I like the way that this question separated the discussion about how end-of-induction response is used in clinical decision making, then the second half of the question is about its implications in the design and conduct of clinical trials. The reason I like that is because I think the first part speaks to, a little bit, the history of how we've made clinical decisions in this disease and reminds us to be a little bit humble about things.

Years ago, those of us who took care of neuroblastoma patients followed the practice guidance that came from the Yanik data about end-of-induction Curie score, and when we had a patient with a even slightly high Curie score, we

sat people down and laid crepe, and told them continuing on therapy is essentially futile and we really have to rethink the goals of care and those kinds of things.

Then new therapy came along, and specific chemoimmunotherapy, and it really was a game-changer. We don't have that conversation with patients anymore. We have a different set of conversations; not that it's super happy, and based on the Pinto data, we know that often those patients don't fare as well, but it's not as clear-cut.

So I think we should just remember that in addition to trial design, we have to think about clinical decision making and the implications of putting a stamp of approval on end-of-induction response as the be-all and end-all when the world does change and evolve, and we just have to stay humble about that.

DR. PAPPO: Thank you very much for that comment.

I'm going to go a little bit off script. I

know that Dr. Popovic is not part of the panel, but she's an expert in this area, and she would like to make a comment, so I am going to allow that.

DR. BECK POPOVIC: Thank you very much.

Many things have been said, but it is important, I think, that if we have the end-of-induction response to use in clinical decision-making, that we have a plan; that these patients are not lost. In the past, in our former high-risk study, patients that had insufficient response in end of induction had additional chemotherapy, which was planned, before they would go, then, to surgery, high-dose chemotherapy, et cetera, and now we have this implemented in the new high-risk protocol through VERITAS.

So I think it is important that these patients might not be lost and that we have a strategy to use what happens at end-of-induction evaluation to decide on further treatment.

This is just to add to the former comments. Thank you very much for letting me make the comment.

DR. PAPPO: Thank you.

I think that the overall consensus on this question number 3, it all started with Steve's highlight that it will be very important for these patients to remain on protocol and to track their outcomes since they will provide significant important information for the future, and this was a recurrent thing.

Again, the other answer that would go to this question is given the poor outcome of these patients, that it's a good thing to try to become creative and come up with new methods to try to assess response and identify new therapeutics for these patients. What Dr. Ro Bagatell said also is this concept of end of induction and how we identify these patients at high risk for failure ultimately may evolve over time as new therapies come into play.

Did that summarize our discussion for question number 3?

(No response.)

DR. PAPPO: Did I miss anything?

(No response.)

DR. PAPPO: Okay. We will now proceed with question number 4.

DR. YU: Dr. Pappo, I'm sorry. It looks like Dr. Glade Bender still had a comment on this question.

DR. PAPPO: I apologize. Please go ahead.

DR. GLADE BENDER: Julia Glade Bender, Memorial Sloan Kettering. Actually, I should apologize to you, but as you were restating our thoughts, I had a new thought, which was that we've also had a lot of discussion about poor end-of-induction response.

I just think as we move forward collecting good data on end-of-induction response, it is important because as we discussed in our session yesterday, there is also a subset of patients potentially with excellent end-of-induction response who subsequently might benefit from a question of whether or not they need all of the downstream chemotherapy, including high-dose chemotherapy in the future. And unless we start to

rigorously collect this data and the outcomes in a uniform manner, we won't be able to ask questions like that either; not just those who have a poor end-of-induction response, but those who have an excellent one, and maybe there could be a therapy reduction question in the future.

DR. PAPPO: Excellent point. Thank you.

If there are no additional comments or suggestions for question number 3, we will move to question number 4, and we will have the FDA read this question.

DR. DONOGHUE: Given the current strength of evidence for using response at the end of induction to predict patient outcome and assess antitumor activity, consider the appropriate use of this endpoint in clinical trials.

I think we may have touched upon this a decent amount already, but hopefully there's some additional discussion to be had. Thanks.

DR. PAPPO: If there are no questions or comments concerning the wording of this question, we will now open this question for discussion.

(No response.)

DR. PAPPO: Any takers?

DR. DuBOIS: Dr. Pappo, I have my hand raised. I'm not sure if you can see that.

DR. PAPPO: Okay. Sorry. Go ahead, please.

DR. DuBOIS: Steve DuBois, Dana-Farber.

At the risk of being extraordinarily obvious, I think it's crystal clear that as we're designing trials, we need to obviously capture detailed end-of-induction response; and not just the overall response category, but site-specific response because it certainly may be that end-induction response as an overall measure may have its limitations, but that perhaps disease domain-specific responses may be more informative such as response by MIBG scan or clearing the bone marrow, for example. As we're developing these trials, we'll never be in a position to develop and validate surrogate markers without that level of detailed data.

DR. PAPPO: Thank you very much.

Dr. Mitchell?

MR. MITCHELL: Thank you, Doctor. It's Mr. Mitchell, but I'm always happy to be promoted on these calls.

Given that the FDA standard for surrogate endpoints, I believe, is that they'd be reasonably likely to predict a clinical benefit, especially that they are important when we're dealing with an unmet need, as long as we continue to study and to confirm whether end-of-induction response, a good end-of-induction reduction response, is in fact predictive of positive clinical outcome, I think that there's a reason, given the strength of the evidence, to use this endpoint in clinical trials. Thank you.

DR. PAPPO: Thank you very much.

Any other additional comments for question number 4?

DR. KRAUS: Yes. I have one. Albert Kraus, industry representative, Pfizer. I agree with Mr. Mitchell. My personal view is it may go beyond just in trials in terms of reasonably likely to predict and be considered in drug reviews and

approvals. But around the appropriate use in trials, I think we heard articulated that it's used and induction maintenance is an important practice, and I have familiarity with it in other settings.

Normally, the designs that I'm familiar with in maintenance are different than what were described previously in slides. Normally, anyone who doesn't progress would be put on randomized maintenance, and somebody who progressed would go to new therapy because obviously the therapy wasn't helping them. But anyone who had a PR, or CR, or stable disease would be randomized to some level of continued therapy, which sounded like that was a clinical practice, too, in what I heard from Dr. Pinto; that it was hard for PRs or CRs because of varied subsequent additional therapy, but in a way, that seems to be how maintenance evolves in certain other settings.

So I think it's important to think about those designs and to absolutely measure it, especially if substantial induction response can be achieved, as seems to be noted. So I just wanted

to comment in that way, on those two points. Thank you.

DR. PAPPO: Thank you very much.

I don't see any other hand that are raised. Oh, I think Dr. Bagatell, yes?

DR. BAGATELL: Thank you, Dr. Pappo. This is Ro Bagatell from Children's Hospital of Philadelphia. I'd like to just remind everyone of something Dr. Pinto said in his talk because he's from Peoria.

I think we always have to think about the international neuroblastoma response criteria in the multicenter context, and remember that if you have stable disease in one component, you have stable disease overall. This comes to the point of how does it play in Peoria in the sense that we have a wide range of institutions that take care of children with neuroblastoma, and some have surgeons who don't operate on very many patients with neuroblastoma and may leave a good amount of tumor behind. And even if you have a fantastic response in the rest of your compartments, as Dr. DuBois and

Dr. Beck Popovic have said, you're still coded as stable disease.

So I just want to emphasize the importance of keeping the limitations of the INRC in mind because you may have patients who are thought to have an inadequate response to induction who actually had rather a stupendous response to induction, but residual disease left at the primary site.

So I think if we're going to use end-induction response as an endpoint in our clinical trials, as Steve suggested, we really need to look at the components and probably get more data on response in metastatic sites in the background for this use, but then also require that people who are using end-induction response in trials really break it down and help us understand what's happening in metastatic sites versus the overall response that might end up penalizing new drugs, based on having some primary tumor left behind.

DR. PAPPO: That's an excellent point.

Thank you.

If I can briefly summarize the discussion for question number 4, it will be very important going forward in a clinical trial, that uses end-of-induction therapy as a surrogate endpoint, to capture data in detail, specifically at specific sites since that could potentially affect the overall assessment of response.

That may be a problem, especially when you have a multicenter trial with multiple places where they may not have a lot of experience with treating neuroblastoma. There was an overall consensus that this is probably a good endpoint to continue to assess. And finally, to continue to, I would say, think about optimal study designs for this group of patients, especially those that have achieved a PR, or CR, or stable disease.

Did I leave anything out?

(No response.)

DR. PAPPO: We will move to question number 5 now.

DR. DONOGHUE: If there is sufficient

evidence to support future efforts, please provide recommendations regarding interest, feasibility, and future steps to validation of end-of-induction response as a clinical endpoint in the first-line treatment of patients with high-risk neuroblastoma. I think we may have covered some of this already as well, but certainly welcome additional discussion.

DR. PAPPO: I think so, too, but I think, Dr. Bagatell, do you have your hand up, or you just forgot to put it down from the last comment?

DR. BAGATELL: Sorry. I forgot to put it down. I'm doing it now.

DR. PAPPO: Okay. I'm trying to scan. I see David Mitchell.

MR. MITCHELL: That's a leftover. I'm putting it down now. I apologize, Doctor.

DR. PAPPO: Sorry.

Nita? Dr. Seibel?

(No response.)

DR. PAPPO: Nita, did you have a comment?

DR. SEIBEL: Can you hear me now?

DR. PAPPO: Yes.

DR. SEIBEL: Okay. Nita Seibel from the NCI. This has already been mentioned, but I don't think we can emphasize or highlight this enough, how crucial it will be to have accurate follow up on the patients, based on their end-of-induction response, so we can really correlate this as a clinical endpoint. I think that's some of our concerns, particularly in the ANBL 2131 proposal, is to make sure or to be assured that this will happen.

DR. PAPPO: Perfect. Thank you very much.

Steve, do you have a comment?

DR. DuBOIS: Yes. Steve DuBois, Dana-Farber. Personally, I think I've certainly heard a fair bit of interest from the group and from this rather lively discussion about continuing to explore some component of end-induction response. I'd like to thank the FDA for putting together today's meeting, and I've certainly learned a great deal.

I guess my only other comment would be that I think understanding what data are available today

and our desire to not wait until completion of additional clinical trials to pursue this, I'd just say that this would require a fair bit of creativity because I don't think the data in-hand can follow the presented rubric for how surrogates are traditionally validated, so just to encourage creativity from all stakeholders.

DR. PAPPO: Thank you very much.

Any additional comments for question number 5?

Sorry. Go ahead.

DR. KRAUS: Albert Kraus, industry representative. From all the discussion and my view of it, I think there is sufficient evidence to study it further. And I do agree, I think, that we won't achieve full Prentice criteria for absolute determination of surrogate criteria, which as FDA kind of highlighted, almost never happens. Most of these surrogate examples, even full surrogate examples I'm aware of in the drug review process, kind of didn't quite go that way. It was a judgment, and it was a judgment sometimes in areas

with huge patient populations.

So in this case, with very sparse patient populations and sparse data, I think judgments would be needed, adequacy and determination of relationships needed, et cetera, et cetera. I thought I'd just want to mention that because, at least from a sponsor's standpoint thinking about pharmaceutical company medicines and doing studies, it's very challenging when we get these rare areas to do some of the studies we might optimally want to design in any kind of time frame, which all of you around the phone are involved with those, so you know that.

But I'm just kind of stating the obvious there; that, therefore, we have to find ways to do the right thing for the patient rather than just throw our hands in the air. So that's just my comment.

DR. PAPPO: Thank you very much.

Dr. Glade Bender?

DR. GLADE BENDER: I just want to validate all that has been said, especially by Dr. DuBois

and Dr. Bagatell. I think the people who are most involved in designing this research are well aware of the limitations, and one of the limitations has been the lack of a clinical trial that's really set up to collect these data in a rigorous way. So there's certainly interest in doing that, and it is certainly feasible to do that.

Then I would also urge creativity about maybe internal to the trial, an intermediate step of potential validation, and even an interim analysis of whether the surrogate is looking like the standard, and somewhere in the middle even. I don't know if such a thing exists, but I think creative minds could come up with a way, and that the time is now to do so.

DR. PAPPO: Thank you very much for your comment.

I don't see any other hands, so I'm going to try to summarize the discussion of question number 5. I think there's definitely a lot of interest in pursuing this endpoint. Some of the members believe that there's already sufficient

evidence to support future efforts exploring end-of-induction responses as a clinical endpoint. Other members feel that some of the data is still evolving, and we need to continue to explore this endpoint, but we also need to make decisions based on the data that we have today.

It will be crucial to be sure that we have accurate follow-up on these patients so that we can have better clinical correlates. And finally, we have to become creative in our study design to better evaluate these potential surrogate endpoints.

Did I leave anything out?

(No response.)

DR. PAPPO: If not, we will now proceed with the FDA closing remarks from Dr. Martha Donoghue.

Closing Remarks – Martha Donoghue

DR. DONOGHUE: Thank you so much, Dr. Pappo.

On behalf of FDA, I'd really like to thank everyone who gave presentations today and took part in the discussion. I really wish I could pass the podium over to Leona Knox right now because I think

she would probably do a better job of summing this meeting up than I can.

What resonated with me was the commitment in the community to continue to work together to figure out the best way to make clinical trial design and clinical development overall more efficient for patients with high-risk neuroblastoma. I'm thinking about Gigi McMillan's comment -- I think it was your comment, but correct me if I'm wrong -- that using end-of-induction response is the ethical thing to do for patients. I certainly think that I'll walk away keeping that in my head.

We also talked a great deal about the strength of evidence for use of end-of-induction response as a prognostic factor in predicting outcomes in patients with high-risk neuroblastoma, and talked a bit about the complexity of assessing that endpoint; and given the strength of the evidence that we have in hand, it may or may not be validated for use as a surrogate endpoint reasonably likely to predict clinical benefit, and

the traditional methodology that we can use, and have used in the past, to validate it across the spectrum of validation as well, and the limitations perhaps that we have, given the rarity of high-risk neuroblastoma.

I was really heartened to hear so many comments expressing commitment to using this endpoint and continuing to follow patients so that we better understand use of end-of-induction response and its various components, and the correlation between end-of-induction response and event-free survival and overall survival.

We also spoke -- and I know we at FDA and our colleagues at the EMA are fully committed to early discussions and continued discussions -- on use of this endpoint in our regulatory decision making, as best we can, to help do what we can to approve drugs that are safe and effective earlier for the treatment of patients with high-risk neuroblastoma.

I guess I will leave it at there. I think we have lots of reasons for optimism, and now it's

time for us to roll up our sleeves and continue to work together on this.

I will see if Dr. Reaman would like to add anything.

(No response.)

DR. PAPPO: Greg, do you want to add anything?

DR. REAMAN: Yes. Thank you, Alberto. Thanks, Martha.

No, I think you've adequately summed it up. Again, I would just like to thank the committee members and the presenters. I think this was enlightening. I think at this point now, we have work to do. I think we've seen some impressive results about the prognostic significance, but whether this really has the predictive ability to be used as a surrogate marker, I think it's going to require additional work, and hopefully work that we all continue to do together because this does require multistakeholder investment of time and effort.

I again would thank the input of our patient

advocates and patient representatives, which is really the only reason why we're here and why we're doing what we do, so thank you all again very much. This is just the beginning of the story, so much work to be done. Thank you.

Adjournment

DR. PAPPO: Thank you, Dr. Donoghue and Dr. Reaman. I also want to thank the FDA staff for making this meeting a success. There are a lot of people that worked very hard to make this happen. I just wanted to highlight Joanna Malsch, and then of course Joyce Yu that kept me on track. And I want to thank all of you for your support and your attention, and I hope that we can get together next year in person.

We will now adjourn the meeting. Thank you very much.

(Whereupon, at 2:52 p.m., the meeting was adjourned.)