CENTERS FOR MEDICARE AND MEDICAID SERVICES

Medicare Evidence Development & Coverage

Advisory Committee

August 22, 2018

Centers for Medicare and Medicaid Services

7500 Security Boulevard

Baltimore, Maryland

Panelists

Acting Committee Chair
Joseph Ross, MD, MHS

Committee Vice-Chair
Aloysius B. Cuyjet, MD, MPH

MEDCAC Members
Joseph S. Cheng, MD, MS, FACS, FAANS
Diane Civic, PhD, MPH, MSW
Naftali Z. Frankel, MS
Melissa M. Garrido, PhD, BS
Thomas F. Goss, PharmD
Thomas James III, MD, FACP, FAAP
Joel Lamon, MD, FACP
Carla Perissinotto, MD, MPH

Industry Representative
Shamiram Feinglass, MD, MBA

Guest Panel Members

Stephen Gottschalk, MD
Doug Olson, MD
James C. Yang, MD

Invited Guest Speakers
Elissa Bantug, MHS
Ethan Basch, MD
Ilia Ferrusi, PhD
William Go, MD, PhD
Paul Kluetz, MD
Claire Snyder, PhD

CMS Liaison
Tamara Syrek Jensen, JD

Executive Secretary
Maria Ellis

TABLE OF CONTENTS

Page

PANEL PROCEEDINGS

(The meeting was called to order at 8:10 a.m., Wednesday, August 22, 2018.)

MS. ELLIS: Good morning and welcome, acting committee chairperson, vice chairperson, members and guests. I am Maria Ellis, the executive secretary for the Medicare Evidence Development and Coverage Advisory Committee, MEDCAC. The committee is here today to focus on the state of evidence on CAR T therapies, CAR T cell therapies that are approved by the Food and Drug Administration. We are seeking the MEDCAC's recommendations regarding collection of patient-reported outcomes (PROs) in cancer clinical studies. The MEDCAC will specifically focus on appraisal of evidence-based PRO assessments to provide information that impacts patients, their providers and caregivers after a CAR T cell

therapy intervention for patient's cancer.

The following announcement addresses conflicts of interest issues associated with this meeting and is made part of the record. The conflict of interest statutes prohibit special government employees from participating

in matters that could affect their or their employer's financial interests. Each member will be asked to disclose any financial conflicts of interest during their introduction. We ask in the interest of fairness that all persons making statements or presentations disclose if you or any member of your immediate family owns stock or have another formal financial interest in any company, including an Internet or e-Commerce organizations, that develops, manufactures, distributes and/or markets consulting, evidence reviews or analyses, or other services related to PRO assessments or CAR T-cell products. This includes direct financial investments, consulting fees, and significant institutional support. If you have not already received a

disclosure statement, they are available on the table outside of the room.

We ask that all presenters please adhere to their time limits. We have numerous presenters to hear from today and a very tight agenda, and therefore, cannot allow extra time. There is a timer at the podium that you should follow. The light will begin flashing when

there are two minutes remaining and then turn red when your time is up. Please note that there is a chair for the next speaker, and please proceed to that chair when it is your turn. We ask that all speakers addressing the panel please speak directly into the mic and state your name.

For the record, voting members present for today's meeting are Dr. Aloysius Cuyjet, Dr. Joseph Cheng, Dr. Diane Civic, Mr. Naftali Frankel, Dr. Melissa Garrido, Dr. Thomas Goss, Dr. Thomas James III, Dr. Joel Lamon, Dr. Carla Perissinotto. A quorum is present and no one has been recused because of conflicts of interest. The entire panel, including

nonvoting members, will participate in the voting. The voting results will be made available on our website following the meeting.

I ask that all panel members please speak directly into the mic.

This meeting is being webcast via CMS in addition to the transcriptionist. By your attendance, you are giving consent to the use and distribution of your name, likeliness and voice during the meeting. You are also giving

consent to the use and distribution of any personally identifiable information that you or others may disclose about you during today's meeting. Please do not disclose personal health information.

In the spirit of the Federal Advisory Committee Act and the Government in the Sunshine Act, we ask that the advisory committee members take heed that their conversations about the topic at hand take place in the open forum of the meeting. We are aware that members of the audience, including the media, are anxious to speak with the panel

about these proceedings. However, CMS and the committee 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 topics during breaks or at lunch.

Please remember to discard your trash in the trash cans located outside of this room. Guests are prohibited from taking photographs on the CMS campus. And lastly, all CMS guests attending today's MEDCAC meeting are only permitted in the following areas of CMS single

site; the main lobby, the auditorium, the lower level lobby, and the cafeteria. Any persons found in any area other than those mentioned will be asked to leave the conference and will not be allowed back on CMS property again.

And now, I would like to turn the meeting over to Ms. Tamara Syrek Jensen.

MS. JENSEN: Thank you. Good morning. One, I wanted to thank everyone for coming today, this is an important meeting for us, and thank you to the panel for surviving all of

their travel hardships last night in attending.

First and foremost, we are trying to get the temperature lowered in here and the humidity lowered in here, so hopefully that will happen. Like you, I am also very warm, so hopefully that will happen in the next half an hour to an hour.

I know we're running a little bit behind time so I am just going to cede my time. Do we go to the chairperson, or do we -- oh yeah, sorry, this is why she's here. Just a reminder.

If you have not signed up -- we have the invited public speakers. The folks that

would like to speak today that are not on a list, there's a list out back, you need to sign up before ten a.m. this morning. So please, if you would like to speak, or you think you want to speak in the next hour or two, please get your name on a list out back, and then there is also a disclosure form that you also need to sign that Maria just talked about. And I will continue to remind you of that deadline,

because it is a hard deadline, so that we can then incorporate that into the agenda.

It is a full agenda, there are 23 voting questions, so we are going to make sure that we move through this because we really do need to get to those voting questions and we want to hear what the panel has to say about those voting questions, so that is why the time is so important today.

So -- did I miss anything else? All right. With that, I'm going to cede it to the chairperson, Dr. Ross.

DR. ROSS: Hi. My name's Joe Ross, I'm a general internist on the faculty of the School of Medicine and Public Health at Yale. Thank you to everyone for joining us today,

this is going to be a very exciting discussion about this new therapy that's revolutionizing clinical medicine in some respects.

I have been given a lot of advice about how to coordinate this meeting. All I can say is I will be very strict on time, it is not personal, so if I cut you off, please,

please, please respect that, and it's because I'm trying to give everybody an opportunity to speak, there are many people who want to have their say today, and we are very much looking forward to that.

Maria, is now the time where we're supposed to introduces ourselves and disclose our conflicts?

MS. ELLIS: Yes.

DR. ROSS: Okay. So as I said, I'm Joe Ross, on the faculty at Yale in the School of Medicine, and I just, I have no personal conflicts. I do want to note that my research group at Yale receives funding from Johnson & Johnson as part of a clinical trial data sharing efforts, and we also receive funding from CMS and FDA for research work, it's all through Yale, and we formally receive funding

from Medtronic and the Blue Cross Blue Shield Association.

DR. CUYJET: Hi, I'm Aloysius Cuyjet, I have no disclosures to make.

DR. CHENG: Joe Cheng, chair of

neurosurgery at University of Cincinnati, no disclosures.

DR. CIVIC: Hi, Diane Civic, no disclosures, though I do work at Anthem.

MR. FRANKEL: Naftali Frankel, no disclosures.

DR. GARRIDO: Melissa Garrido, no disclosures.

DR. GOSS: Tom Goss, Boston Healthcare Associates, so I'm a paid consultant but I have not done any work in the CAR T area.

DR. JAMES: Tom James, senior medical director at Highmark Blue Cross in Pittsburgh, and I have no disclosures.

DR. LAMON: I'm Joel Lamon, no disclosures.

DR. PERISSINOTTO: Carla Perissinotto, no disclosures.

DR. FEINGLASS: Shami Feinglass, I work for industry. We are not involved in

CAR T.

DR. GOTTSCHALK: I'm Steve Gottschalk, a member of St. Jude Children's Research

Hospital. I have research support from a company called Teva Therapeutics and I have patent applications in the CAR T cell therapy field.

DR. OLSON: Doug Olson, patient advocate, no disclosures.

DR. YANG: James Yang, surgery branch of the National Cancer Institute. The National Cancer Institute has a cooperative research agreement with Kite Gilead.

DR. ROSS: Great. And thank you for every member of the panel for being here today and to the CMS staff for helping to organize this meeting.

I'm going to turn it over now for the first presentation from CMS from Katherine Szarama.

DR. SZARAMA: Good morning. Many thanks to the panel in advance for your consideration and helpful discussions today around this important topic. Currently CMS is reviewing the evidence on chimeric antigen

receptor (CAR) T-cell therapy in response to a

formal complete request for a national coverage determination.

It is critical to identify the information needed for beneficiaries to make informed treatment decisions with their providers. Therefore, today we will be discussing the study, collection and dissemination of health-related quality of life with patient-reported outcome (PRO) assessments. Specifically, the purpose of this meeting is to obtain MEDCAC recommendations regarding how existing PRO assessment tools should be incorporated into future clinical studies, including future clinical studies on CAR T-cell therapy, and clinical study design characteristics, study duration, and suitable study controls. To this end, CMS has provided background materials to support your assessment of the strength of evidence on five voting questions.

The first question requests your vote on the confidence in the strength of evidence validating seven PRO assessments that we will describe here.

The patient-reported outcomes-common terminology criteria for adverse events, PRO-CTCAE, is a free assessment developed in 2008 to supply meaningful data and improve understanding of symptomatic adverse events from multiple disease states, based on the hypothesis that collecting information directly from patients improves the precision and reliability of symptomatic adverse event detection. Validity was based on comparison to established scales and follow-up based on outpatient clinical visits. Completion rates were over 90 percent. Results showed 98 percent of PRO-CTCAE items were significantly associated in the expected direction with established assessments.

The M.D. Anderson Symptom Inventory was developed in the year 2000 from established brief pain inventory and brief fatigue inventory to be specific to cancer patients. It contains 13 core symptom items and six interference items, with multiple formats available. It contains, in addition to validity testing, the MDASI and other quality of life assessments were collected for

comparison at five event-dependent time points in treatment for multiple myeloma and non-Hodgkin lymphoma. Completion rate was 82 percent. Results showed symptom severity and interference scores correlated with specific treatment events, and were not sensitive to demographics, diagnosis or laboratory measures.

The European Organization for Research and Treatment of Cancer Quality of Life Questionnaire, EORTC-QLQ-C30, was based on the following six characteristics: standardization, cancer specificity, psychometric strength, practical application to cancer clinical trials, appropriate self-administration, and cross-cultural applicability. The background materials provided include a sample of the validation work of this assessment, and has been recognized in meta-analyses by Kotronoulas et al. and colleagues as the standard measure on which more recent assessments are compared.

The University of Washington Quality of Life assessment was developed specific to head and neck cancer patients based on three

desired characteristics; that the assessment

take less than ten minutes to complete, it is simple to understand, and measures health-related quality of life longitudinally. Version four includes 12 domains, a single-item quality of life question, and a free text section. The background materials include three publications measuring the application and validation of this assessment. Results showed 42 percent of UW Quality of Life items were significantly associated in the expected direction with the EORTC-QLQ-C30 summary score. Completion rate was 79 percent.

The Patient-Reported Outcome Measurement Information System, PROMIS, uses item response theory and computerized adaptive testing to build on existing items in the patient-reported outcome and quality of life instruments database with focus on the following six desired characteristics: One, appropriate context, the instructions associated with answering the item; two, the appropriate stem, the part of the item that

makes it unique; three, consistent response options; four, minimal time spent answering; five, instrument of origin; and six,

domain-specific rather than disease-specific measurement. To this end, the modular structure can be customized and administered specific to each individual patient. The background materials provided summarize the clinical validity for nine PROMIS measures in five PROMIS domains in over 1,000 patients across six clinical conditions. The completion rate was 95 percent.

The Electronic Self-Report Assessment in Cancer is not a single assessment, but an electronic self-report method that contains questions regarding common cancer symptoms and quality of life measures, including the EORTC-QLQ-C30 and patient health questionnaire, PHQ-9. The background materials provided summarize the impact of such methods, where ERSA-C scores seem to communicate patient health-related quality of life, and reduce symptom distress with additional engagement

that did not increase time during an outpatient clinic visit.

Lastly, the Functional Living Index in Cancer was validated in a study by Schipper and colleagues as compared to Karnofsky, Beck

Depression, Spielberger State and Trait Anxiety, and Katz Activities of Daily Living, as well as the general health questionnaire and McGill's pain index. The assessment was designed to assess the overall functional state of the patient with the following desired characteristics: One, cancer specificity; two, functional orientation; three, patient self-administration; four, high compliance; five, reproducibility; six, sensitivity to a range of clinical practice and intensity of therapeutic intervention; seven, face, content, construct, and concurrent validity and reliability.

Question two asks for the consideration of evidence on the following desired characteristics of a PRO assessment: Breadth of measures in emotional, social and

physical well-being; quick throughput to apply to clinical study; transferable to community practice settings; measures that are not sensitive to differences in age; line of therapy; comorbidities; and that are generalizable to study of combinations of therapies; used in a net benefit analysis based

on symptom burden and well-being.

Additional sources of consideration include the Alliance For Clinical Trials in Oncology and other partners' recommendations for geriatric oncology research, recognizing that elderly patients are less willing to compromise on health-related quality of life and prefer to maintain function and independence during cancer treatment and management, according to these authors. We also ask for your discussion of two following questions regarding patient-reported outcome assessments and other desired characteristics.

Question three asks for your consideration of an appropriate measurement period for a valid PRO assessment.

Question four asks for your consideration of an appropriate measurement duration for a valid PRO assessment.

Question five asks for your consideration of an appropriate control population for a valid PRO measurement.

Thank you.

DR. ROSS: Thank you, Dr. Szarama. Our next speaker is Dr. William Go, the vice

president of clinical development at Kite.

DR. GO: Hi, my name's William Go. I'm the vice president of clinical development at Kite, a Gilead company. First of all I want to take this time to thank the committee for the invite to speak at this, as well as to introduce chimeric antigen receptors and the discussions in terms of patient-reported outcomes. My background is I'm a PhD in T-cells and I'm a hematology oncologist and BMT by training.

So Kite, a Gilead company, is committed to the research and committed to discovering and developing a novel T-cell

immunotherapy for patients with unmet needs, especially in the cancer space. And we are committed, and we applauded CMS and FDA on patient-reported outcomes, and we recognize the importance of incorporating PROs in our own drug development process and in the overall assessments. While Kite recognizes the importance of PROs in the measurement of clinical trials, the PRO CAR T science where these interests are most appropriate for CAR T remains still to be determined and still is

evolving, and still is quite early in the development.

Today I want to first introduce the CAR T therapy of YESCARTA as well as the transforming technology of CAR T, which has shown huge benefit for patients who have exhausted all therapies in large cell lymphoma, and have not had any other options in standard of care therapy.

So, I know this is a busy slide and I know we have limited time, and so what we'll do is first talk about what chimeric antigen

receptors are. Chimeric antigen receptors are where you have a, where you take a patient's own T-cell in an autologous setting and reprogram them to express a chimeric antigen receptor where it is an antibody specific to the target antigen, in this case CD-19, as well as then costimulatory domains and also signal one and two, in this case it's CD3-zeta, as well as CD28.

This was recently FDA approved in October of 2017, and these were in patients that we studied in refractory large B-cell lymphoma. These are patients that usually have

less than six months left to live and in cases where we normally, I would be having hospice discussions with these patients. This is intended to be a one-time infusion, and it's limited at certified healthcare facilities who have BMT cell therapy experience. So this turnaround process of a patient's own truly personalized medicine is approximately 17 days from door to door, and these patients really have very limited time in their disease state

in terms of their outcomes and their overall survival. So any delay in this process from evaluating the patient, collecting their apheresis, their T-cells, to engineering their T-cells, to then returning them, conditioning chemotherapy and infusion, and then ultimate recovery is very key, every day matters for these patients.

And we are also in the post-approval commitment to the FDA as well as others, working with others which we will talk about in terms of registries and future investigations of PROs.

For the sake of time, I won't go through this study design but just, again, this

has been published in the New England Journal of Medicine. High level, we only study patients with refractory, chemo-refractory large B-cell lymphoma. Again, in the SCHOLAR I retrospective global analysis, these patients only have approximately six months to live, with only a seven percent complete response rate, as well as a 26 overall response rate.

As you can see in the updated analysis to the right, that with a median follow-up time of 15.4 months, with everyone having an opportunity to have one year of long-term follow-up, the best overall response in this case after one dose of cellular immunotherapy of CAR T, of axicabtagene ciloleucel, was 82 percent with a complete remission, meaning no evidence of cancer throughout the trial at one point at 58 percent. At the time of the data cutoff, 42 percent of the patients still had ongoing durable remission, with 40 percent having complete remission.

This data is unprecedented, especially in this patient population. And what, again, I wanted to highlight is that in this kind of refractory patient population, that this

benefit-risk has been reported as positive, and this is why this is the basis of our full approval, regular approval with the FDA.

Of course in oncology as you know, with the box warning, we have two predominantly box warnings which is very consistent across

the CAR T space, as well as other T-cell targeted spaces, which is both cytokine release syndrome and neurologic events. Cytokine release syndrome is pretty much on target with what we expected because we are reengineering someone's own immune system to fight their cancer, and what we can see is that even though that their Grade 3 and higher events are clearly significant, that this is generally reversible and generally self-limiting.

The median time of hospitalization in our trial was approximately 14 days, and we already reported at scientific forums that predominant long-term side effects have really been very well managed, consistent with autologous stem cell transplant and really focusing more on other aspects such as infectious risk, which is consistent with the B-cell aplasia that you do see in CAR T.

Again, though, this is truly different compared to what you would see typically in end stage of the large cell lymphoma.

Now to the Medicare population. As

you all know, the median age of diagnosis of large cell lymphoma is 58, and so therefore you would expect to see a fair amount of patients over the age or equal to 65. In our pivotal study we had 25 percent of them at age 65 or older. What is remarkable, which is very different than auto and allogeneic stem cell transplant is we did not put age limits on our trial or any of our trials. In fact, we just really focus on the performance status of ECOG PS zero or one, but you can see that we actually treated on trial up to 76 years old, and currently we have been in the commercial setting, we had patients up into their 80s.

What you can see here is the observable overall response rates are very similar in patients greater than 65 in age compared to the overall ZUMA-1 population. Specifically, over objective response rate was 89 percent versus 82 percent in the greater than 65 years old, the complete response rate

was 70 percent versus 58 percent, and the most important part, the durable remission rate was

48 percent versus 42.

Interestingly, compared with the patients for safety, the patients greater or equal to 65 had a lower incidence of SAEs and a lower incidence of Grade 3 or higher infections. However, a higher incidence of Grade 3 or higher neurologic events, as you can see here. The higher incidence of neurologic events in subjects greater than or equal to 65 years of age was driven by events that would be expected to more frequently afflict the greater than 65-year-old patient population in general, such as delirium, agitation, and disturbance in attention.

So now we're pivoting to what we were really asked for in terms of the committee, and what I wanted to highlight is that in the post-market commitment setting to the FDA, Kite Gilead is committed to patient safety and following patients after treatment, and agreed to this extensive post-market registry. As a former transplanter myself, and our CIBMTR colleagues in the room, we felt that CIBMTR is

the right group and forum to do this registry study.

Just to take a step back for people who are not aware, in America we have the American Society of BMT. With the American Society of BMT, we have other organizations that support this as well as FACT accreditation for cellular therapy, as well as the Center for International Blood and Marrow Transplant Research, or CIBMTR.

The registry objectives are quite clear. You can see that we're really looking to make sure that we have long-term safeties, especially around these cellular therapy products, and this registry may include data from other countries as well and is modular, and there's a steering committee to oversee the registry, and the steering committee is made up of representatives of key certified sites, opinion leaders and experts in the field, as well as us, Kite, as the manufacturer.

DR. ROSS: Dr. Go, you have five minutes more.

DR. GO: Yeah, two slides left, or three slides left.

I wanted to go to the data elements. Because of the five minutes left, I'm not going to go through the data elements here, but I just wanted to show you that this is robust, we just wanted to show you that this is clearly robust and that we thought about everything, and this has been in discussions with the FDA as well as other PRO experts, as well as other people in the field. Okay.

So, where are we with the registry? The registry has been enrolling patients as early as March of 2018. All 61 certified sites for YESCARTA are also registered as CIBMTR centers. As of July 6, 2018, you can see that 220 cell orders shipped since the registry enrollment, and 82 YESCARTA recipients were reported in the registry. You can probably see what's that delay, this is pretty standard as a former transplanter, that because of the data resources and the man or woman power resources to enter the data, that we usually do this in either a monthly or quarterly fashion, so this is actually expected based on that period of time. So we are clearly looking to make sure that we don't have any substantial missing

data, and to ensure that the enrollment is going to be clear, and expected to be at 90 percent based on previous registry experience.

We are using PROs in our ZUMA trials right now. Again, PROs are highly variable, and they may be specific to disease state or disease-specific AEs. We are doing this in ZUMA-1 and in single arms of cohorts three and four, but more importantly, we are doing this in ZUMA-7, our randomized controlled trial. We feel at Kite Gilead that PROs are best served and best interpreted in a prospective randomized Phase III setting, especially globally, and this is the first to my knowledge, the largest randomized global Phase III study in terms of CAR T, and we're looking at axi-cel in terms of in second line as compared to the standard of care in second line. Right now the FDA approval is after two systemic lines of therapy; this is after one systemic line of therapy, and we're building in PROs prospectively.

So as my final slide, the Kite

experience with PROs to date, Kite generally uses PROs. However, again, we feel very

strongly that it's best served in the randomized Phase III setting. As several CAR T randomized cell trials are underway with different PROs selected for further data collection and randomized controlled trials is warranted to consider the best instrument selection or modification and appropriate timing of assessments, which are key as PRO instruments, and their applicability including scientific rigor are still evolving. The context of CAR T therapies, we feel PROs are not quite ready for real world coverage decisions at this time.

Therapies like YESCARTA are transformative and considered one-time, and would have huge clinical benefit for patients with limited options. These patients have advanced disease and dismal survival times, and without access to CAR T therapy would be detrimental for patients.

Kite is dedicated to continuing

follow-up of YESCARTA patients, and therefore have an extensive multiyear post-market registry as well as a strategy of examining PROs prospectively. PROs are a valuable tool

and Kite is committed to their study and their value in cancer care. However, we do not believe PRO data collection should be linked to coverage for these transformative therapies at this time. However, we welcome the opportunity to partner with CMS as well as other experts in the field and key stakeholders to explore these topics further.

Thank you very much for your time.

DR. ROSS: Thank you, Dr. Go. Our next speaker is Dr. Ilia Ferrusi, from US Oncology, Novartis.

DR. FERRUSI: Thank you, Dr. Ross. Okay, I have my slides. It is my distinct pleasure to be here today to speak to this MEDCAC committee on behalf of Novartis, in particular in the context of these groundbreaking innovative CAR T therapies that are changing the course of treatment for

patients who have exhausted all of their options. I'm very grateful to be here to continue this discussion and to ensure patient access.

Before getting to the topic at hand, I would like to share that, a bit of my

background with the MEDCAC committee. I'm a paid employee of Novartis and I do hold an interest in the company. In terms of my training, I did receive PhD training at McMaster University in health research methodology with a specialization in health technology assessment, and I've worked in the pharmaceutical industry five years. My research experience spans a total of seven years in both the academic and industry settings. During my time in industry, I've supported the inclusion of the patient's voice in the drug development process through the development, validation, and interpretation of patient-reported outcomes in ophthalmology, dermatology, aesthetics, women's health, and oncology.

For the benefit of today's panel, I would like to share Novartis' experience with patient-reported outcomes in clinical studies of Kymriah, the first approved CAR T therapy. This may benefit CMS and the panel as it considers whether patient-reported outcomes should be included in future clinical studies of CAR T clinical research.

A patient-reported outcome consists of any measurement based on a report coming directly from the patient regarding their health or their condition without interpretation, any interpretation of that report by a clinician or anyone else. Novartis collected such PRO data during its registration studies of Kymriah in leukemia and lymphoma to specifically measure changes in patient-reported disease burden and its impact on quality of life. This information was in turn provided to regulators to inform the risk-benefit analysis of Kymriah as an investigational medicine. However, with the approval of Kymriah for both leukemia and

lymphoma, we are concerned about the burden of additional PRO data collection on both patients and providers beyond the context of registration studies for these very sick patients.

Kymriah is the first approved CAR T therapy, having received two FDA approvals in the past 12 months. The first was for treatment of patients up to 25 years of age with B-cell precursor acute lymphoblastic

leukemia that is refractory or relapsed after second line therapy. So again, this is a patient population who have exhausted all standard of care treatment options.

Our second approval was for adult patients with second line or later relapsed or refractory large B-cell lymphoma including diffuse large B-cell lymphoma, high grade B-cell lymphoma, and diffuse large B-cell lymphoma resulting or arising from follicular lymphoma.

Kymriah is available for both approved indications through a restricted risk

evaluation and mitigation strategy program via a network of certified treatment centers, and it may be administered on an outpatient or inpatient basis. For both approved indications, these relapse refractory patients have a very poor prognosis with prior therapeutic options. They also experience substantial disease burden, which motivated us to incorporate patient-reported outcomes to measure changes in that disease burden and quality of life from the patient perspective in our registration studies. For the purposes of

today's MEDCAC I will focus on our PRO experience in that diffuse large B-cell lymphoma population, which bears the greatest relevance for Medicare patients.

Standard of care in second line diffuse large B-cell lymphoma consists of high dose chemotherapy with autologous stem cell transplant, but as many as three-quarters of those patients may not be eligible to undergo the transplant procedure for various reasons. While salvage chemotherapy could be an option

for these patients, response rates remain low and survival is quite short. Taken together, the prognosis is grim for these patients who are relapsed or refractory after first line treatment. Approximately 21 percent will survive to 18 months following their relapse.

We'll skip ahead to our DLBCL slides here. The safety and efficacy of Kymriah in DLBCL was established in the JULIET study, a single-arm multinational registration trial that enrolled patients from the U.S. and nine other countries in North America, Europe and Asia Pacific. Efficacy was measured as best overall response rate comprised of both

complete and partial response. With a median follow-up of 14 months, Kymriah demonstrated a best overall response rate of 52 percent.

The most common adverse events in JULIET included some of those that are well known and the previous speaker did address, cytokine release syndrome and neurologic events, but also prolonged cytopenia, infections, and febrile neutropenia. In our

clinical study, most adverse events were resolved within the eight weeks of treatment.

Now to patient-reported outcomes. With the objective to measure disease burden and quality of life, the functional assessment of cancer therapy for lymphoma, or FACT-Lym, was administered in the JULIET study. We selected this specifically because it's a disease-specific PRO measure that was developed and validated specifically for use in adult lymphoma patients. This is of particular importance to Novartis as we sought to collect information from patients using measures that have established validity and reliability for the specific disease context being studied, and required a measure that had been translated and

validated in several different languages, given the multinational nature of our study.

The FACT-Lym includes four cancer subscales from the FACT-G measure, which is a general cancer measure, and that includes physical, social, emotional and functional well-being over the past seven days, and it

also includes a lymphoma-specific subscale containing items that are particularly relevant for lymphoma patients and the symptoms that they experience. We administered the FACT-Lym at screening and at follow-up months three, six, 12, 18 and 24. This administration schedule was designed specifically with the intent to measure changes in disease burden, not exactly the burden of treatment, while minimizing the patient's questionnaire burden.

Here are some of our results. As compared to baseline scores, we observed clinically meaningful improvements in patient-reported quality of life at months three and six of follow-up among those patients achieving a partial or complete response. These improvements met thresholds of clinical importance despite the initial toxicity

experienced after a one-time treatment. Importantly, this PRO data proved useful in corroborating the clinical data that we also observed, indicating that clinical improvements were associated with corresponding improvements

in quality of life for these patients.

It's also worth noting here that data missingness was highest among nonresponders, as these patients left the clinical trial for additional treatment elsewhere, and this is very much reflective of real world practice as well. This created challenges for interpreting PRO data and should be a key consideration for future clinical studies of CAR T and its alternatives.

Follow-up does continue for this JULIET study population, with data for months 12, 18 and 24 forthcoming.

We also used the Short Form 36 questionnaire to measure health-related quality of life. You may be more familiar with this, it is among the most widely used questionnaires owing to its more general nature, which facilitates comparison to other diseases. This was also requested by regulators outside of the

United States.

Again, here we observed clinically meaningful improvements on the Short Form 36

across general health, bodily pain, vitality, physical functioning, role-emotional, role-physical and social functioning subscales, with only the mental health subscale not showing a clinically meaningful improvement over baseline. Again, these findings did corroborate improvements in clinical endpoints as well as improvements noted in the FACT-Lym.

A bit about PROs and incorporating them in your research. To support the collection and interpretation of PRO data, there was significant effort and investment made to prepare for PRO administration, collection of PRO data, and reporting of such data. I won't go into every point on this slide, it is extensive, but I wanted to be able to share this with the audience for consideration.

The extensive training, infrastructure and analytic expertise needed to support the meaningful inclusion of PROs in clinical research bears important consideration for

future clinical studies that this MEDCAC

committee may be considering. Many of these needs, the infrastructure and whatnot, do exist in the context of clinical trials for registration purposes, but outside of this setting may require additional investment to ensure appropriate instrument selection, administration and interpretation. In the context of CAR T studies that does that would necessarily impact both the sites where CAR T is administered, as well as any subsequent location that the patient may go to for their follow-up care and for the comparators, which may not be at one of these FACT-accredited sites.

So therefore, we do urge caution to this MEDCAC panel in its mission to identify a PRO measure for future CAR T studies. Although CAR T therapies are currently approved for relapsed refractory diffuse large B-cell lymphoma, applications may include other relapsed refractory populations in the near future, including adult acute lymphoblastic leukemia, chronic lymphocytic leukemia, multiple myeloma and follicular lymphoma in the

future. Symptoms, adverse events and burden may vary across these diseases, which vary also in their acute versus chronic nature, and that may necessitate the use of different measures depending on research objectives.

This is just one of the important contextual factors that makes it challenging to advise the panel on a single PRO measure for all future studies of CAR T in advanced cancers. Moreover, it is important to clearly define research objectives when considering whether to collect patient-reported outcomes data, particularly given the burden associated with advanced disease experienced by the patients that are likely to receive this therapy. Despite the significant support of our registration study and the infrastructure therein, we experienced challenges collecting PRO data, which resulted in missingness and risks to data quality and interpretation. This could very well be amplified in larger trials or in real world practice.

In closing, I would like to thank the committee for the opportunity to share our experiences, and I look forward to addressing

questions later in this meeting.

DR. ROSS: Thank you, Dr. Ferrusi. Our next speaking is Dr. Paul Kluetz, the associate director of patient outcomes at the Oncology Center of Excellence at the FDA.

DR. KLUETZ: So, good morning, my name is Paul Kluetz and as was mentioned, I'm a medical oncologist currently serving in the U.S. Food and Drug Administration, serving as an associate director in the newly formed Oncology Center of Excellence, and I've been leading efforts to formulate a patient-focused drug development effort for oncology that will cross the centers with the advent of the 21st Century Cures Act. I have no financial relationships to disclose. I will be using several examples of PRO instruments within my slides, mainly to illustrate important characteristics that we think about for the use of PRO data for regulatory review. These are examples and not blanket endorsements of one instrument over another, and we definitely encourage sponsors to come and talk with us about which type of tool they want to put in

their trial.

So, I think it's important, and we'll probably hear about this more today, that I'm only one stakeholder, or I represent only one stakeholder. As you can see, data from controlled clinical trials are used for many different stakeholders, and especially when there's only one clinical trial usually at the time of approval in cancer, and so part of the challenge in designing a PRO strategy for industry is that they need to meet the needs of all these different stakeholders who may have different needs, and this is why the FDA oncology group and the work that I've led has included domestic and international payers, as well as international regulators, industry and importantly, patients, in our efforts to try to standardize the field.

So, the outline for my presentation is as follows. I'll introduce some definitions. Because this is a unique research space, it has its own language, and I want to make sure we're all on the same page. I will talk about

clinical outcomes. I'll talk about an effort we've led to provide some core concepts to measure as an expectation to create some

standardization. I'll go on and talk about measurement tools, how do we measure these concepts and what do we use for characteristics for patient-reported outcome tools. And then I'll talk about importantly endpoints, what questions are we asking, because if we don't know what questions we're asking, we can't formulate the correct endpoint to answer that hypothesis.

So in this slide I'll start to use some common terminology. So, a concept is any aspect of an individual's clinical, biologic, physical or functional state or experience, it's the broadest of things that you start with in your trial. A concept can be clinical, like pain, function or survival, or it can be nonclinical, like a biologic pharmacodynamic biomarker or PSA. When your concept is something that describes how someone feels or functions or survives, it's called a clinical

outcome, and this has important regulatory implications because we provide regular approval based on clinical outcomes that are viable in how a patient feels or functions. And if we're not looking at survival, we're

looking at fields and functions that we use a clinical assessment, and one of these types of assessments is the patient-reported outcome which we will talk about today.

Finally, a test or tool or instrument is actually the assay that you use to measure your clinical outcome, and in the case of patient-reported outcomes, this is the questionnaire and its scoring manual. And an endpoint is the most precisely defined variable that you're really talking about, intended to reflect an outcome of interest that's statistically analyzed if you want to make a claim of treatment benefit to address a particular research question.

All these definitions are available at this link that I've provided, which was an NIH and FDA collaboration to create some standard

terminology called the best philosophy.

So just to walk down this terminology example, you may be interested in pain in your clinical trial, the concept of interest is pain. Pain is a measurement of how someone feels, so it is a clinical outcome assessment that measures pain is considered to be a

patient-reported outcome because patients are usually in the best position to quantify a non-observable symptom. The PRO instrument you might want to use could be the brief pain inventory, it's got 15 questions, but you're not going to use the whole thing, you might only use one question as your endpoint, and so your endpoint could be a two-point decrease in the third question, worst severity in 24 hours, of this questionnaire, that's seen and confirmed by a second result two weeks later with no increase in analgesic use. So I only put this up to show you how specific the endpoint can be in order to really create a well defined and reliable assessment.

So as we've seen, a clinical outcome

is a measurement of how someone feels, functions or survives. I've said overall survival is, or I should say that overall survival is a common endpoint we use, and if you're not measuring that, you're measuring symptoms or function. These are the four types of categories of clinical outcome assessments that we can use and they really have everything to do with the source of the data, where is the

data coming from.

A clinician-reported outcome comes from the clinician after a history and physical; for instance, safety is reported in clinical trials as a clinician-reported outcome typically. Observation-reported outcomes are, say, a parent reporting on an infant on an observable sign. Performance outcomes is like a six-minute walk test where you're asking a patient to perform something. And then patient-reported outcomes are as was just described, a questionnaire, and it's information coming directly from the patient without interpretation or change by a clinician

or anyone else. So when I ask you what's your pain and you say it's eight out of ten, that's the answer, it's an eight, and that's how we use it for our endpoint.

Increasingly and interestingly, there's a fifth type of data stream that we're starting to see more often, and this is mobile technology tools like sensors and wearables, and I think this holds promise to complement patient-reported outcomes especially with respect to physical function. It's certainly

an area evolving in regulatory and scientific interest, and we're working in several collaborations to try to understand this better.

My personal view on wearables is that for concepts like physical function, I think it can complement patient-reported outcomes data, and I think both used together could provide internal validation and be quite powerful.

So, patient-reported outcomes have been in trials for a long time, we have instruments that are 30 years old. Why now,

why is there so much enthusiasm for patient-reported outcomes or clinical outcomes in general? Well, number one, there's recent legislation that, FDA has been charged to attack patient-focused drug development, to take a look at patients and see what's important to them throughout drug development, and part of what's important to them is measuring how they feel and function, so we need to try to do a better job with that, and that's what we've been doing.

I think everyone realizes we're in an era of lots more options now so we can inform

our therapeutic choice better if we have more data on how patients feel and function. Technology is definitely improving our capabilities, not only to measure PRO through electronic capture that creates structured data that's easier to aggregate, easier to utilize, but also, again, with wearable devices and biosensors creating passive data streams, and this allows you to look at patients outside of the clinic, and so you can identify probably a

real stream of real world data that we're starting to see.

We're going to hear from Dr. Ethan Basch today on his work using patient-reported outcomes to monitor symptoms. It has shown some benefit to patients in the clinical setting. So I think clinical care is starting to utilize patient-reported outcomes more for communication and again, that's probably going to become more common and provide a stream of real world data.

And finally, we've talked about PRO-CTCAE as a library. These item libraries are very interesting, they're allowing us to be more flexible with our trial design, so we can

select only those symptoms that are likely to occur, and not be asking patients about symptoms that are not going to occur only because they're on the legacy instrument that was developed in the era of cytotoxic chemotherapy, for instance.

So we talked a little bit about clinical outcome assessments. I'd like to talk

a little bit about core concepts and what we've been working on to try to organize a standard approach to our work. So, many things can be measured with PRO, many many things. You can look at symptoms, you can look at side effects, you can look at symptoms of disease, work life, sexual function and overall health quality of life. You can ask many things, distress.

What should be the core concepts that we focus on for our work in evaluating a therapy? Because the heterogeneity in PRO concepts that are being measured right now and the heterogeneity in the tools that are being used to measure them, and the heterogeneity in the assessment frequencies has been very problematic for the FDA and I think for the scientific community in general, so I think a

concise set of PRO concepts can create consistency, and the focus for our work in evaluating a therapy should be on isolating the therapeutic effect. We want to know what the drug is doing to the patient and their disease rather than other nondrug influences.

So to illustrate this issue of isolating the therapeutic effect, we can use this figure which is a conceptual framework of health-related quality of life or quality of life more broadly, adapted from Wilson and colleagues in the '90s. You can see the investigational therapy that's going to be given, it's the blue circle, and it's going to affect biologic and physiologic variables. It's going to have on target and off target effects. It's going to create symptomatic side effects and it's going to hopefully alleviate the symptoms of disease, and that net benefit or drawback is going to be associated with functional changes, which then feeds into general health perceptions, which then feeds into quality of life.

And as you can see, as you walk down from the proximal symptoms to function to

health to quality of life, there's a lot of inputs that begin to come into play that are nondrug influences, things like motivation, psychologic factors, support value preferences,

spirituality, et cetera, and so we have at the FDA focused on proximal symptom and functional measures because we would like to isolate the effect of the investigational therapy to the extent possible. It's never going to be perfect.

And so in 2016 when we were asked to review the patient-reported outcome landscape, we proposed to focus our current work on looking at symptoms of the disease, disease symptoms, symptomatic adverse events and side effect impact, and physical function, the ability of patients to work and perform their activities. These are all key components of health-related quality of life, but not all there is to health-related quality of life. Importantly we know, as well as other stakeholders, would like to know about health-related quality of life as well as some of these other distal domains, and we know from international payers, et cetera, that they need

this. So anything that we do, we want to make sure that it's going to get the data for

everyone that needs it, and so it's going to require a different way of thinking about PRO measurements.

So what about the measurement tools, what can we do to measure these concepts? So, a fit for purpose patient-reported outcome instrument for FDA purposes is appropriate for its intended use. That means that the study design, the patient population and the therapy under study need to be appropriate. What's the baseline function of the patients, what kind of side effect profiles are we going to be seeing, is this a TKI versus immunotherapy, is this a cytotoxic versus a TKI? You know that there's going to be very different toxicities for each. The instrument has to validly and reliably measure the concept, and it needs to be both clinically relevant and important to patients, and both of them need to occur for them to be a good measure in a clinical trial, and they don't always align.

So, an example is sexual function in prostate cancer. If you have a local therapy

and the patient has an intact prostate and intact sexual function, it's important to patients, and it's going to change potentially with the therapy, which is prostatectomy or radiation. On the other hand in the metastatic setting where patients have had their prostatectomy, are on lifelong androgen deprivation therapy and they are very unlikely to regain their sexual function, while it's important to patients, it's a terrible trial measure. I could have a wildly effective drug in that setting that showed no benefit in sexual function, and you may undervalue in that setting the effect of the drug.

Finally, and I think importantly, we have to be able to communicate this data on the FDA label in a way that is non-misleading and accurate, and so it has to be well defined what you're measuring, and I want to really get into well defined a little bit because I think that's important for instrument selection for FDA purposes. Well defined means that the questions within the PRO scale ask about the concept that they're measuring. For instance, the concept of physical function can be defined

as a patient's assessment of their ability to carry out activities that require physical effort, and here I give two examples of well-defined physical function domains.

So for PROMIS, in a PRO measurements information system by the NIH, there's four questions, all asking about varying levels of activity, and that will be rounded up into a score or a scale and the physical function rate will be better or worse and we can label that, knowing that all those questions were related to physical function.

Similarly with the EORTC, their physical function domain within the QLQC30 has five questions ranging from lowest to highest function and they're all asking about physical function. So when we label physical function as improved, it's not misleading, this is what was being asked.

Another example of a well-defined scale could be the concept of disease symptoms. You may have several symptoms that you wrap up into a single score, and that symptom scale or score should be well defined, these should all be symptoms of the disease that you're

studying. This is the myelofibrosis symptom score that has six cardinal symptoms of myelofibrosis, and this was labeled in the FDA.

So I showed you what is well defined. What's not well defined? In this example I'll give a fictitious example of a fatigue score, and if you take a look at these six questions which are intended to measure the concept of fatigue, is this a well-defined scale? Now I've made this up so this is not a real scale. So in the first three questions, these are asking about symptoms directly related to fatigue, how tired are you, how much weakness do you have, and what's your energy level, but the last three questions are problematic from an FDA standpoint. So what level of pain do you have and how much numbness and tingling do you have may contribute to your fatigue, and what is your health-related quality of life may be the impact of fatigue, obviously high fatigue will cause low quality of life, but it's not fatigue itself. So now to wrap that up in a fatigue score, if you say that fatigue

was improved, it's kind of a little bit misleading, because what if it's all driven by

quality of life, or all driven by neuropathy? So then we have to break it down and try to look at the item level, and that becomes problematic for FDA.

So as I mentioned before, we're only one of like many stakeholders, and so whatever the strategy is for core concepts that have well-defined physical function and disease symptoms, treatment side effects, we need to be able to create a strategy where everyone gets the data that they need, so we need a thoughtful combination of existing static tools like Short Form well-designed physical function, for instance, coupled with item libraries so you can select the right symptoms for your context, and it's going to require a modular approach, and what do I mean by that?

Conceptually if you look at a modular approach, it would mean that all these functional domains in health-related quality of life can be measured separately, scored

22 separately, and are well defined and you could
23 actually create an endpoint out of any one of
24 these boxes, cognitive function, physical
25 function, health-related quality of life, and

1 within that you can have a little more fidelity
2 and flexibility on these core proposed FDA
3 concepts that I mentioned before using item
4 libraries, or standalone disease symptom scores
5 like myelofibrosis, or now there's
6 non-small-cell lung cancer SAQ that was just
7 qualified.
8 Examples of modular PRO instruments I
9 wanted to give you, the EORTC QLQC30 is one
10 example of a modular design. Now it is a
11 30-question health-related quality of life
12 form, but it's not all added up and provided
13 one score and that's it. You can see this
14 taken directly from the scoring manual, that
15 each of the functional scales are, they are
16 scored on their own, they have symptom scales
17 and items that are also scored on their own,
18 and a two-question global health-related
19 quality of life score or health score that's

also scored on its own. So it's very modular, where we wanted to look at physical function specifically and if the company wanted to add that as an endpoint, a specific endpoint, we could do so and it would be considered well defined.

Another example of a modular system that was mentioned earlier is the PROMIS system that was designed specifically to be modular, to be able for you to choose specific scales and domains, and add them to your trial as you see fit.

And then the third example of a tool that you could apply to a modular type of framework would be an item library where you can select symptoms in a flexible way to specifically address trial questions, and this is the PROCTCAE, which is an item library we'll probably hear more about that's specifically for symptomatic side effects, so it really looks at safety and tolerability as a clinical trial objective.

So, I would like to end by looking at

the big picture. We can select core concepts, we can agree on well-defined instruments, but how are you going to define your question and what is the endpoint going to be that's going to answer that question? Again as I mentioned, the definition of an endpoint ends in, to address a particular research question. We have to come up with some standard questions

that we feel PRO can help us answer.

So what are some common research objectives? Well, I mean for my purposes, we really look at safety and efficacy, so are you looking to explore safety and tolerability? You can certainly complement safety by looking at longitudinal symptomatic adverse events. You can also look at overall side effect impact, which in reviewing some slides, I think we may hear about a little later.

And then with efficacy as I described, you can get through symptom scores, improvement in symptoms, you could do improvement in physical function. If you had an esophogeal trial, esophogeal cancer, you may want to do

swallowing as a function of living, but you have to know what your research question is.

So I proposed, you know, a background of what different kinds of outcomes are. I just wanted to show you what I think the important data elements could be in future cancer trials. Currently this is really what we focus on and it's served us very well, as has been described by the CAR T results. We look at overall survival, we look at

progression-free survival, tumor measures, overall response rate, serum biomarkers, and clinician-reported outcomes, like CTCAE safety data and dose modifications, and that gives us a really good sense of the drug's safety and efficacy.

But we're in the era of patient-focused drug development, we can do better. Some of the things we're looking at include other clinical events that are occurring that are going to be important to patients, hospitalizations, ED visits, morbid procedures, supportive care use. We have some

of this in our trials base and we've begun asking companies to provide us with some of this data for us to further explore some of these concepts.

Today, germane to this advisory committee, I also think that patient-reported outcome and other clinical outcome data, and potentially biomarkers and sensors, can provide this patient-centered data. In my mind, if we could just focus on, start with these core concepts as the basis and make that consistent across trials, we could add things on top of

that as well, but look at disease symptoms, symptomatic adverse events, overall side effect impact, and physical function and ability to perform your activities.

By being clear on these core PRO concepts or clinical concepts that we can measure, we can begin to do what we've done with our other trial measures. Let's take a look at the standardization based on certain data elements, survival, highly standardized in how we assess, analyze and present these

endpoints. Tumor progression, very highly standardized, as well as tumor shrinkage, using RECIST criteria. Adverse event reporting, highly standardized using the CTCAE lexicon.

Now what we need to do better at is to take a look at these core PRO or clinical outcome assessments and begin to standardize those. What are the RECIST criteria going to be for patient-reported outcome core concepts, and that's really where we're spending a lot of energy right now. So again, at FDA oncology, we've talked about the concepts, we've talked about the instruments, and I think we have a relatively reasonable plan and path forward.

We're moving now onto standard objectives, endpoints, analytics, so that we can get data standards, start to create some SAS codes to really be able to really begin to scale this, because I don't think I can overestimate the amount of resources it takes to get through some patient-reported outcome data sets that may not be super standardized, and there's no clear specific endpoint in mind, they can be

very challenging. So if we start with standard objectives and concepts and create endpoints that we can all agree upon, and a couple of standard ones, I think we will go a long way.

So, a couple of examples to end, from going from concept to instrument to endpoint, if you're looking at disease symptoms as one of your core concept measures, you could use the myelofibrosis symptom assessment form as a well-defined instrument to assess those symptoms, and the endpoint could be 50 percent or greater reduction in symptom score by week 24, and guess what? That was exactly what was used as a key secondary endpoint, statistically tested, supporting the regular approval of Ruxolitinib in the treatment of myelofibrosis.

It can be done.

From concept to instrument to endpoint using physical function, we haven't had many physical function applications where they used it as a key secondary endpoint, none actually as a key secondary endpoint, but if the concept of interest is physical function, the

instrument could be PROMIS, could be EORTC, could be any well-defined functional domain, and one possible endpoint, just throwing it out there, could be what is the physical function change from baseline at 12 months, giving you enough time to experience the toxicity, to understand what some relatively subacute toxicity chronicity will be, and whatever benefit you're getting from the drug may start to appear.

Now this is, there's lots of problems with cross-sectional analysis when you're just picking one endpoint, and I want to illustrate that quickly because I think it's relevant for today. Let's take a look at a fictitious physical function over time diagram. On the Y axis you have the functional, the physical function outcome result, you have

administration of the drug at time zero, 30, 60 and 90 days, and this might be what a patient's physical function does over the time. The context here, which may sound familiar, is a therapy that's given once, not chronically,

that has significant toxicity early, that then begins to recover, and if the patients have significant functional decline at baseline, the significance of your disease may actually improve function later on as patients recover completely.

Now if you choose your cross-sectional endpoint at 30 days, you're going to see a decrease in physical function in this example. If you choose your physical function at 90 days, you're going to see an improvement of physical function from baseline, so you could have completely different results if you select an endpoint. So in endpoint A we're really looking at tolerability, and in endpoint B we're really actually looking at efficacy, and that's why you need to understand your disease, your context, your endpoint definition, and really be thoughtful about your strategy.

I want to also make one note, that

when we look at patient-reported outcomes data, it's within the context of everything about the drug, the patient, the disease, the tumor

measures, the survival, everything, and it's very different contexts in certain scenarios. So an adjuvant curative therapy that's given for a specific period of time, it's going to have significant toxicity, but it's a defined treatment duration and there's a potential for cure. That's one type of context to put your risk-benefit decision into.

And then if you have a palliative noncurative therapy, which is the vast number of therapies actually in solid tumor oncology, you treat them to progression, you have cumulative toxicity as you go on with treatment, and the benefit is typically limited to the time on therapy; when you progress, the therapy is no longer effective, and that's a different context to think about what your results show.

So in conclusion, clinical outcomes in my mind can complement and not replace survival, tumor and standard safety measures, and patient-reported outcomes are one type of

clinical outcome, there are others.

Patient-reported outcomes functional scales and symptom scales that create a score out of a number of questions should be well defined so that we can allow for clear communication in FDA labeling. Several patient-reported outcome measurement systems exist, and we don't need to reinvent the wheel. Some concepts are relatively agnostic to the scenario; that's why I think physical function is a nice concept to measure, it's important in CHF, it's important in COPD, it's important in metastatic cancer, it's important across a lot of different diseases. In fact, that's sort of what the PROMIS idea was mentioning.

I think finally, it's really critical that you come to the FDA with your specific trial. I think we can help, and we would like to start providing written guidance to try to standardize some of these things, which I think will help the community a little bit in a sense. And I think it's also critical, as I mentioned, to understand the treatment context and carefully consider taking a research objective and creating an endpoint that

supports that, so thank you for your time.

DR. ROSS: Thank you, Dr. Kluetz. You should know that I very nearly interrupted you halfway through before I was reminded that you were given 30 minutes to speak, not 15. So I will note now that coming to the panel, Dr. Claire Snyder, professor of medicine, oncology and health policy and management at Johns Hopkins School of Medicine and Public Health, along with Ms. Elissa Bantug, the manager of communications, education and survivorship at the Johns Hopkins Breast Cancer Program, both at Johns Hopkins, and they are given 40 minutes to speak.

DR. SNYDER: Good morning. It's my pleasure to be here this morning. I'm Claire Snyder, and I'm pleased to be here with our patient advocate partner Elissa Bantug to talk about some research we've done with funding from the Patient-Centered Outcomes Research Institute, looking at how to display patient-reported outcomes data so that patients and clinicians can understand what the scores mean and use them in practice.

I currently and have previously

received research funding from Genentech, and I receive royalties on a section that I write in UptoDate on cancer survivorship. Ms. Bantug has nothing to disclose.

To be clear about the purpose of our presentation today, CMS asked us to talk about our project in terms of how it involved engaging with patients and other stakeholders to develop improved presentation strategies for patient-reported outcomes, so our emphasis is on the stakeholder engagement methods that we used, as well as providing you some insights on our study results. I want to emphasize that when we looked at the different data display formats, it was agnostic to PRO questionnaires, so there's nothing that I'm going to present today that would impact one questionnaire versus another.

There are three different contexts or applications of PRO data that our study addressed. The first is individual patient monitoring. This is when a patient completes a PRO questionnaire about how he or she is doing, the data is fed to the clinical team, and is

used to inform his or her personal care. This

has been shown to improve communication, help monitor progress and inform management. This is an example data display format for an individual patient. It shows the current score and previous scores with possibly concerning scores highlighted in yellow. Because this application is less relevant to the discussions today, I just wanted to introduce it, but we will not be focusing on it.

More relevant to this group is how you display research study results, and this is just an example study from the New England Journal of Medicine looking at quality of life among prostate cancer survivors, and the results of this were translated into data appropriate for display to patients so that they can see that either external beam radiation or brachytherapy are less toxic to bladder control compared to surgery. Those same data were displayed in the actual journal publication in a much different way.

The problem with all these data

displays is that it can be very challenging to interpret because of the variations in PRO instruments, so I think we're all familiar with

the multitude of PRO instruments out there. The last time I checked, there were over 800 listed in the PROQOLID database, and across these instruments there's no standardization in how they are scored, scaled, or in how the data are presented. So in some PRO questionnaires, such as the SF-36 which we've heard about, higher scores are always better. On other PRO questionnaires, higher scores represent more of what is being measured. An example of this is the QLQ-C30, which we've also heard about. That means higher scores are better for function domains but worse for symptoms. And then on some questionnaires, lower scores are always better, this is frequently the case with symptom questionnaires or something like the sickness impact profile.

There's also variation in how PRO instruments are scaled, some are linearly transformed, zero to 100, with the best and

worst at the extremes, whereas others such as PROMIS are normed to a general population average of 50. Because of this, a score of 50 can mean entirely different things and it can be difficult for patients and clinicians to

keep track of that. And finally, there's variation in how the data are presented. We've seen examples of that this morning in terms of whether you show average scores over time or the proportion meeting a responder definition.

Based on all of this variation, clinicians have told us that they value PRO data, they want to use it in practice, but they have difficulty understanding it. So we undertook a three-part mixed method study, again with funding from PCORI, and in the first part of this study we looked at current approaches for displaying PRO data to find out what patients and clinicians found helpful and what they found confusing. In the second part of this study we took the results from part one and partnered with patient and clinician stakeholders to develop improved presentation

approaches. And in part three, we tested those approaches.

As I mentioned, we were brought here today primarily to talk about how patients and other stakeholders can be engaged in research, so it's now my pleasure to introduce Ms. Bantug, who will talk about our stakeholder

engagement approaches and her role as our patient partner.

MS. BANTUG: Thank you, Claire, it's a pleasure to be here today. I'm going to talk a little bit, as Claire mentioned, about our stakeholder engagement. So in a research study with 30 oncologists, almost all felt that PRO data had value. However, less than half felt comfortable interpreting these results. And as Claire mentioned, there was a lack of standardization when it comes to how these measures are scored, scaled and displayed. Research also indicates that some methods used to display this data are more easily understood than others.

When we designed this study, we wanted

to come up with a way to better display PRO data that impacts clinical practice. Our stakeholders were around and along throughout the continuum of care. We wanted to make sure that they were available to aid in better understanding and use. The stakeholders were part of our investigative team, as well as an advisory board that we put together. In addition, we wanted to make sure that they were

involved in study conduct. There was broad inclusion with stakeholders, as I mentioned, on the investigative team as well as the advisory board. This included doctors, nurses, physicians, caregivers, as well as being part of study subjects; this helped with the intervention and development.

And finally with how the study was implemented and disseminated, advocates were really helpful in study implementation and they were critical in the dissemination process. So looking a little bit at how we put together our core team, this investigative team, this included patients and caregivers, PRO

researchers, as well as clinicians. And as you can see here on your screen, there's a picture of me in the top corner representing a patient as well as a caregiver, cancer patient and taking care of my mom who's a cancer patient. On the bottom of your screen you can see a great picture of Claire here as a PRO researcher, as well as Michael Brundage, who's a radiation oncologist in Canada and a PRO researcher who's very well versed in a lot of these PRO metrics.

So we had this core team, we met together weekly, the Hopkins team. We also had a qualitative researcher who was part of the team, we met weekly to get together with Michael on the phone, and then we worked with the advisory board that we put together to really make sure we had broad representation. We felt like the stakeholder advisory board definitely improved generalizability, it gave us direction and guided our research project, it helped with connection using various groups including advocacy groups, journal editors,

professional groups, et cetera. Again, this helped us with implementation and dissemination.

And here you can see a list of our stakeholders. Again, we had patient advocates, we had someone representing nursing perspectives, caregivers' perspectives, clinical perspectives with nurses and researchers, journal editors, as well as PRO researchers here.

So looking beyond what we did with our stakeholder advisory board and our investigative team, we wanted to go broader.

We looked at the Johns Hopkins Cancer Center which is where we started our recruitment efforts, but then wanted to go beyond that to the Johns Hopkins Clinical Research Network, and this is where we focused our part one and part two of our study looking at how we recruited patients and clinicians.

Here's a map that you can see of our Johns Hopkins Clinical Research Network. You can see Baltimore is towards the top. Just

north of Baltimore is the Greater Baltimore Medical Center, GBMC, which is just north of Baltimore in Towson. As you move a little bit further you have Anne Arundel Medical Center in Annapolis, and then as you go down the mid-Atlantic seaboard and the 95 corridor, we have hospital representation both in Montgomery County, Maryland, D.C., and into Virginia. The idea was to incorporate a big academic medical center at Johns Hopkins, in addition to community-based hospitals. This is where we recruited part one and part two of our study.

And going beyond that, we wanted to go a little bit broader, and so we knew we had a really unique perspective looking at medical

centers both in an academic setting as well as a community setting, but wanted to sort of take it beyond the Baltimore regional area. So then part three of our study looked at, tried to look at a national sample, and with this we used an Internet-based survey approach, and this again is where our stakeholders were really instrumental in making sure we got

information out to patient groups, professional groups, using social media networks, et cetera, to really get this Internet survey out, to cast a wide net to get a lot of respondents. So as you can see, on the bottom is where we did our part one and part two, and as we expanded to part three in our Internet survey.

Some of the key things that we took away from this was, we felt like it was critical to employ stakeholders from the very beginning of the process from study design all the way through implementation and dissemination, they were involved in every part of the way. We really felt like we didn't just want these stakeholders listed on an application, but we had them together, we brought them together periodically in person as

well as on the phone. They helped us triage problems, helped with study design, as well as moving forward. We cast a wide net so we had a very broad group of stakeholders that were diverse in education, gender, ethnicity, educational background as well as experience,

to make sure we had a wide representation. And we, as I mentioned, it helped us with study conduct and effective dissemination.

And finally, if you want to hear more, read more about how we did this, we did publish these findings in the Journal of Community and Supportive Oncology about how we implemented our stakeholders along the process of this program. And with that, I will turn it back over to Claire to talk a little more in detail about our brief part mixed method study.

DR. SNYDER: There was a pointer up here but I lost it, but I want to thank Elissa for her cooperation and collaboration throughout this project, and now I wanted to share a bit about our methods and results.

So as I mentioned, the first part of the study looked at current approaches for displaying data to find out what patients and

clinicians found helpful and what they found confusing, and so we showed different examples of data display formats, some of which we've seen this morning, so this is bar charts of

average changes at nine months, these are bar charts of the proportion meeting a responder definition, these are cumulative distribution functions that the FDA has suggested, and these are line graphs of scores over time. This particular version includes confidence limits. We also showed a version without the confidence limits, as well as a version normed to a general population average.

And I'm not going to go into details on part one or two but just give you a taste, and what we found was very wide variation in the accuracy of interpretation not just among patients but also among clinicians. And for questionnaires where higher scores were better for function and worse for symptoms, people got confused. They did have suggestions for us which we tried to follow, but at the end of the day, both patients and clinicians preferred the line graphs of scores over time, with the clinicians valuing additional statistical

details such as P values and confidence limits, whereas patients not only did not want that

information, but found it detrimental to their understanding.

So because of this, we actually broke out how we looked at the presentation of research study findings so that we had one stream of research focused on how do you present this information to patients in like educational materials or decision aids, and a subsequent stream then focused on how you present this data to clinicians such as in journal publications. And again, we also focused on individual level data, but that is not the focus here. If you want to find out more about part one, we summarized our results in this publication.

In part two, we used what we thought was a very innovative approach of partnering with clinicians and patient stakeholders to develop the improved presentation formats, and so we used an iterative approach that I wanted to walk through with you. So what we did is we took the results from part one and our research team met and said okay, these are the things

that emerge from the qualitative interviews that we need to try and improve on as we refine these data display formats. And then we had formed work groups based on volunteers from our part one interview participants who said yes, I'm willing to work with the research team on their project, and so we met with this patient and clinician stakeholder group and said these are the issues that emerged from part one, these are the things we're considering doing about it, what would you recommend to us?

Based on the feedback from the work group, we narrowed down some data display formats and we tested those in additional individual interviews, and then we took the results from those individual interviews to our stakeholder advisory board, and we went through this process separately for each of our three applications for our individual patient data, for presenting research results to patients, and for presenting research results to clinicians.

So based on the findings from part one, we had several key questions that we needed to address. Both patients and

clinicians endorsed line graphs for displaying the data, but we decided that the data display format does not drive the analytic strategy, the analytic strategy drives the data display format, so we wanted to address both how to show average scores over time as well as the proportion meeting a responder definition, and then we wanted to figure out how we could deal with the issue of directionality to make it clear whether higher scores were better or worse, how to deal with scaling norms versus non-norms, and for clinicians, how to identify statistical significance and clinical importance.

So the results of part two were the formats that were tested in part three, and for patients, we tested three different proportion formats, pie charts, bar charts and icon arrays. Now it's notable that we tested pie charts. They were not included in part one but they were recommended to us. Previous data in other contexts do not support using pie charts but we said we should go ahead and test it, and we did, so I'll share those results with you in a bit.

We also looked at three different approaches for showing line graphs of mean scores over time. This is an example of the more line graphs, this is where higher scores represent more of what's being measured -- I found the pointer. So here higher scores are better for physical and emotional functioning, but lower scores are better for fatigue and pain.

This is an example of the better line graphs. We did not change the scoring of the instrument but we flipped the axis so that zero was on top, and this way lines trending up were always better.

And then this is an example of normed line graphs where everything was centered on the average for U.S. adults.

To highlight some of the features that are consistent across all of these displays, we provided as clear separation as we could between the function domains and the symptom domains, so if they were scored in different directions that would help people interpret

24 the data. We colored the lines and the labels.
25 We used clear wording on the X axis, patients

1 did not understand what randomized meant. We
2 provided headers that specified how to
3 interpret what a line going up means. And we
4 included descriptive labels on the Y axis, both
5 to help understand what the score means, but
6 also implicitly tells you about directionality.
7 For clinicians, we tested various very
8 similar formats except for, we did not do pie
9 charts because no one could imagine seeing a
10 lot of pie charts in journals. Everything was
11 done in black and white, because still, a lot
12 of things are printed out in black and white,
13 so these are the exact same pie charts.
14 But the other thing we did for
15 clinicians is add in P values. I would note
16 that all the data that is on these figures that
17 I'm showing, the data is the same, only the
18 formats are changing. So these are our bar
19 charts of the proportion meeting a responder
20 definition, and then for line graphs we also
21 did the more format, the better format and the

normed format, these are exactly the same, just in black and white, but we added three additional variations on the lines. Some variations just showed completely plain lines,

others added an asterisk for clinical importance with a legend that explained what that meant, and then others added confidence limits on top of that.

So if you want to learn more about how we worked with stakeholders to develop these presentation formats, those results were published in Supportive Care in Cancer.

So finally, I'm going to talk about the results from our part three study. As Elissa mentioned, we did both an Internet survey and supplemented that with one-on-one interviews with patients and clinicians from the Johns Hopkins Clinical Research Network. For the Internet survey it was self-reported cancer patients or survivors, self-reported cancer clinicians, or self-reported PRO researchers who did not have to focus on cancer, but our one-on-one interviews purposely

sampled based on education, cancer type, clinical site and clinician specialty, and we did that also in parts one and two.

So, we had about 2,200 respondents overall and they were randomized to one of 30 versions of the Internet survey, and I'll go

over what those versions were. For the data presented to patients, we collected data from patients, clinician and PRO researcher respondents, but for the data presentation to clinicians we only randomized clinicians and researchers to that stream.

I'm going to go over this survey design in more detail, but the main point I want to make is that in testing the different formats we kept the accuracy questions the same and the data in the formats the same, so only the way the data was displayed differed. So we asked questions of interpretation accuracy, we asked questions of clarity to rate the clarity, and we also asked them to select which format they thought would be most useful. For the one-on-one interviews, they basically completed

the Internet survey but they thought aloud as they did so.

Our analysis was descriptive summaries of the accuracy questions and clarity ratings. We used a chi-square Fisher's exact testing for the most useful. Format, we used multivariable GEE logistic regression to further explore interpretation accuracy and clarity, and we did

qualitative analysis not just for the one-on-one interviews but the very extensive open-ended comments we got from the online survey.

So first I'm going to show you about how to present the results to patients. Our survey here included 629 cancer patients or survivors, mean age 58, predominantly female and white, with 23 percent who had not graduated from college. We had 139 clinicians, mean age 44, mean years in practice 16, and 249 PRO researchers, mean age 45, with 46 percent with more than ten years experience. For our one-on-one interviews it was just ten patients and five clinicians, but it was well

distributed by our purpose of sampling.

So we had six different versions of the survey to look at data display for patients and the variation across the surveys was the order of presentation, so pies, bars and icons were each shown first, second or third, and one of the surveys with line graphs either seen after the proportions or before the proportions, but each respondent only saw one version of the line graphs, because we thought

it would totally confuse them if we started mixing around the line graphs.

And this is just an example of the kind of interpretation questions we asked, things like at nine months, on which treatment did more patients improve with regard to doing physical activity. Again, the questions asked were the same based on the order of the format seen so that we could isolate any differences, not to differences in data or to differences in the accuracy questions, but solely on differences in the way the data were displayed. The answer choices were treatment X,

treatment Y, or treatments are about the same. For line graphs we asked similar questions, such as at 12 months, on which treatments are patients better able to do physical activities.

For proportions across patients, clinicians and researchers, the bar charts seemed to perform less well, and in our multivariable analyses this emerged more clearly, so both the pies and the icons were more accurately interpreted than bars.

For the clarity, you see a very clear and consistent stepped rating with pie charts

being rated most clear, followed by bar charts and icon arrays. And again, this emerges in the multivariable models where pies are considered preferable to icons and preferable to bars.

In terms of selected quotes, some people said, you know, pie charts are easy, some people are against pie charts, and one of my favorite quotes is, a pie chart is appropriate at a bakers' convention only. Bar charts, some people liked the side-by-side

comparison, but others thought it was more challenging to interpret the data that way. And for the icon arrays, people liked the personalization of the data, but they were considered to be very busy. Icon arrays are generally used in the risk literature where it's a very small number of colored figures.

And pie charts were also the winner in terms of being rated as most useful, with patients, clinicians and researchers all putting it at the top and it was statistically significant for all but the researchers.

In terms of line graphs, you don't see a super clear rating in the descriptive

analysis but what emerges in the multivariate analysis is that the line graphs were, higher scores were always better, were more accurately interpreted than either the normed or the more versions. More performed better than normed.

And the clarity ratings tended to favor, tended to favor better line graphs only when compared to more for somewhat clearer, so the clarity ratings weren't super different.

In terms of the selected quotes, people like seeing the scores over time, but they get confused with the directionality changes. If you show them a more version, they say sometimes lines going up is better, sometimes it's worse, that's confusing. If you show them a flipped axis, they say I don't understand why zero is at the top of the Y axis, and then for normed data, patients can get confused about where these data come from.

So in summary, our pie charts came out as most accurately interpreted, most likely to be rated clear, and rated best for proportions. And the line graphs with higher indicating better outcomes were more accurately interpreted, and more likely to be rated clear

than the more line graphs. These findings are in press for the line graphs and under review for the proportions.

Finally, I just wanted to share our results related to presenting data to clinicians. For this our final sample was 233 clinicians and 248 PRO researchers,

supplemented by ten one-on-one interviews with clinicians. For this one we had 18 different versions of the questionnaire because we had the three different variations on the line graphs. So everyone only saw one kind of line graph, better, more or normed, but we varied the order in which they saw confidence limits, clinical significance and plain versions, and then we also showed the pies before bars in half the versions and bars before pies in the other half of the versions, so that's how we ended up with 18 versions.

So if you look at the accuracy of interpretation, it looks pretty pitiful, with very few clinicians and researchers getting both questions correct, and I just want to explain that finding. So one of the questions was, which treatment did patients improve more

related to doing physical activities? So if you look here, more improved on treatment Y but the P value is not statistically significant, so we considered the correct answer to be about the same, but because many people picked

treatment Y, we reanalyzed the data to look at what proportion got the absolutely wrong answer, so they put treatment X and the answer should have been treatment Y, so these results should be a little more reassuring.

And there was actually no difference between bar charts and pie charts for getting the answer correct. However, pie charts were less likely to be interpreted incorrectly, so pie charts performed better there.

And then for the clarity there were really no differences, and that is shown in the multivariable models.

In terms of selected quotes, some people liked pies again, but others felt uncomfortable with the idea of pies and then four bars. They found that bars were more easier to read or that it took them longer to compare. And I misspoke earlier. I said we didn't use pie charts for researchers and

clinicians; I meant to say icon arrays were not used for clinicians and researchers.

And then in terms of most useful, we

did not find statistically significant differences, although bar charts were a trend among researchers.

When you look at the line graphs in the multivariable models and when you look at the proportion of getting the answer incorrect, normed versions were inferior, and those are shown in the multivariable results, so norms were more likely to be incorrect compared to better and compared to more. The differences were less distinct for getting the answer correct.

And in terms of the clinical significance, we added questions like, for which domains are the average scores clinically significantly between treatments at six months? So the answers are pain and fatigue, and across the different versions of the survey, certainly pain and fatigue were answered more commonly in physical than emotional, but it's not that they were universally identified as being clinically important, even with the asterisk.

And then when we showed confidence

limits, we asked questions like for fatigue, at which time points are average scores statistically significantly different between treatments? And again, the highest ratings were for the ones where there was clear separation between the confidence limits, but even when there was very clear overlap in the confidence limits, some were selecting them as being statistically significant. So this is just a caution in terms of how reliably these data are being interpreted.

In terms of clarity, we again find that the normed are inferior with both the more and better versions, rated as being more clear. And then selected quotes are, the more are confusing because of directionality changes, but the better is confusing because if you think a line going up with fatigue would be more fatigue, not better fatigue, although if they figured out that we had flipped the axis, then they thought it was clear. For the normed it was hard to see the magnitude of the difference. Some clinicians and researchers like the simplicity of just the asterisk for

clinical significance, others appreciated the addition of the confidence limits. So when we asked them to select their favorite line format, clinicians and researchers picked some version with confidence limits or clinical significance.

So to summarize, both clinicians and researchers are unlikely to pick the incorrect treatment, but they were more likely to be incorrect with bar charts. There were no differences in clarity, and researchers tended toward picking bar charts. For the line graphs, the normed versions come out inferior, but the inclusion of clinical importance or statistical significance is appreciated, and all of this is in much more detail in this publication.

So finally, to conclude on our next steps, and returning to our theme of stakeholder engagement, at the end of our project our stakeholder advisory board said that the evidence we had generated is sufficient to inform recommendations for data display but did not define those recommendations on their own, and so they

recommended that we bring together a broader group of stakeholders to develop stakeholder-driven evidence-based standards. We were pleased to get additional funding from PCORI to conduct a Delphi process, taking the results from our study and from the few other studies that had looked at this question, and those results are currently under review.

So with that, I want to acknowledge our team, especially our patient co-investigator Elissa and our team of stakeholder advisors, and Michael Brundage, who was the co-principal investigator. Thank you very much.

DR. ROSS: Great, so thank you, Dr. Snyder and Ms. Bantug. So we are now, it's about ten o'clock, we're about ten minutes ahead of schedule, and we're up for a break?

MS. JENSEN: Yeah, we're up for a break, but I just wanted to remind anybody who wants to speak that is not already on the list to sign up out back, because we'll pick up that list after the break. So if you want to speak, make sure that you are signed in and you sign

the disclosure agreement.

DR. ROSS: So we'll give everyone 15 minutes, so please be back in the room at 10:15.

(Recess from 9:58 to 10:15 a.m.)

MS. ELLIS: If everyone could take their seats, we're about to get started. Dr. Basch, if you could make your way to the podium.

DR. ROSS: So, thank you as everyone is getting back to their seats. We're about to restart. Our next speaker is Ethan Basch, professor of health policy and management and the director of the Cancer Outcomes Research Program at UNC Chapel Hill.

DR. BASCH: Thanks very much, nice to be here, and thank you to CMS for the invitation to speak to you today. I'm a medical oncologist and a professor of medicine at the University of North Carolina. I conduct clinical trials, I see patients, and my research group has done work around patient-reported outcomes for many years, and

I'll describe some of the work by us and others as pertains to the questions posed to the panel today.

I'm going to try to be practical towards the specific questions that the panel will be addressing per my charge from CMS, so I will begin with some general comments and some evidence around the collection of patient-reported outcomes data in oncology clinical research in general, and then I'll turn very specifically to the tools that CMS selected for evaluation today.

These are my disclosures. I conduct research largely funded by the Patient-Centered Outcomes Research Institute and the National Cancer Institute. I have received funding for the development and testing of patient-reported outcome tools.

So, I'm an oncologist, and when I sit with a patient and we make a shared decision about treatment, one of the very first questions that they ask, in fact one of the first questions that we both ask, is how will

the patient feel with the product, how did patients like them previously deal with the treatment, how did they feel with the treatment? And without this information, we have an impaired ability to make informed

decisions. People like to know what to expect, right?

If somebody's going to go through a medical procedure or medical treatment, it would be nice to know how it would make them feel and make them function, and without this information I would argue that we have an incomplete understanding of a product's characteristics and a limited understanding of the longitudinal patient experience that's missing.

Now, you know, some have argued well, we collect this kind of information in clinical research, clinicians really understand their patients' experience, this is often documented either in symptom forms or in toxicity assessment, and this graphic shows some data from my group from a number of years ago where

we did a very simple tracking of about a thousand patients enrolled in clinical trials who were self-reporting their own symptoms and simultaneously their clinicians were systematically reporting on those patients' symptoms. You can see the cumulative incidence of the patient-reported symptoms in orange, the

cumulative incidence of the clinician-reported symptoms in blue showing that we, I and my colleagues, many of us in this room, substantially underestimate the symptom burden that our patients are experiencing by about half.

Now when thinking about adverse events, and Paul Kluetz from the FDA spoke about one of the key domains of patient-reported outcome data collection which is around symptomatic adverse events, you know, this information is currently collected from clinicians, right? In oncology trials we complete information using the CTCAE, the Common Terminology Criteria for Adverse Events. This information is not typically collected

from patients and you know, unfortunately, when it comes to symptomatic adverse events, this information it turns out is not reliable. As shown in this study that was led by Thomas Atkinson at Memorial Sloan Kettering Cancer Center, in which a really simple experiment was done, patients who were receiving treatment on trials were coming in to receive their therapy, and they were seen independently by two

different clinicians, one generally in the office for toxicity check and one in the infusion suite. This was generally within about ten minutes of each other, there were no treatments in between. Both were blinded to each other and completed CTCAE symptoms, and you can see down the middle of this graphic for a number of very common symptomatic adverse events reported with the CTCAE with the interclass correlation coefficients.

For clinical trial grade data, these numbers should be in the .8, .9, somewhere in that range, and you can see that these numbers are substantially lower, showing that if I and

my colleagues see the same patient on a clinical trial on the same day at the same time, that we often disagree with each other. This really raises questions about the reliability of clinician-reported symptomatic adverse event information. We as clinicians are creating a lot of noise, and this is because it is extremely difficult for one person to understand another person's experience with symptoms. And I would say that I myself as an oncologist have, I and my

patients have been a part of many of these studies and I really do know better than any of my colleagues, this is probably just dependent on the psychodynamics of interactions and the logistics of clinic visits.

Now, why is this important? Well, you know, here I show not a CAR T therapy but this is a table from Taxotere, a drug that I commonly use in my practice, and this is taken directly from the FDA label, this is the adverse reaction table from, that was the TAX-327 trial, one of the registration trials

for Taxotere, and we see that more than half of the adverse reactions are symptoms, right, highly subjective phenomena like nausea and taste disturbance and dyspnea, fatigue and myalgia. But you know, currently all this information is reported by clinicians and unfortunately, we have to really question the precision and reliability of this information. We can do better, as has been pointed out in a couple of the prior presentations.

So, what about the applicability of patient-reported outcomes to CAR T therapies? There are several domains of potential interest

that align exactly with those domains that the FDA previously presented to you as being of high interest to them in the evaluation of oncology drug products.

Number one, symptomatic adverse events. We can understand the short-term toxicity profile of these products. We can look at the longer term, right, the late effect of these therapies. Yes, it's wonderful that people are being cured who did not have such a

hope previously, but how do they feel, how are they doing, what can they expect? And in order to manage people's late toxicities we need to know what the late toxicities are and if we don't collect that information, we won't know.

What about earlier changing symptoms that might flag an impending morbid event, right, like CRS? There might be warning signs that come best from the patients if we're monitoring them.

Second physical functioning, if my patient is in bed all day, I'd like to know about that. And I think that future patients would like to know if prior patients had that experience.

And finally, change in disease-related symptoms, which Dr. Kluetz already discussed in detail.

Now, CMS identified a host of tools that they have asked the panel to score in various domains today. I won't comment on the quality of your grading questionnaire today, maybe you can extract from the presentations

about the psychometric properties of your own tool that you'll be using. I'm going to describe each of these tools briefly, Katherine already did this nicely, I'm going to expand on that a little bit, and I want to note that the information I will be providing was extracted through a structured review of the scientific literature and of clinicaltrials.gov data by Dr. Thomas Atkinson, who is here today. He has more hair than in the picture here, but he is more qualified than I am, I would say he is a psychometrician, he has been involved in the development and evaluation of patient-reported outcomes tools, I am but a humble oncologist conveying the message today.

So these are the tools, Katherine already went over these so I won't belabor

them, but I'm going to go through each of them in a little bit of detail now, and I would note that according to clinicaltrials.gov, there are 11 trials that have included patient-reported outcomes for CAR T. This does not necessarily include registries or postmarketing evaluations

that might not necessarily report to clinicaltrials.gov, but of the 11, two used the PRO-CTCAE, nine used the EORTC-QLQ-C30, and two used PROMIS. Again, this is from clinicaltrials.gov and that's the limitation of our evaluation, if somebody didn't post their trial, we wouldn't have seen it; hopefully everybody did.

So in addressing question number one, which just remind everybody here, question one posed to the panel, how confident are you that each of the following PRO assessments are valid and generalizable to the Medicare population, and panelists are asked to assign a score of one to five for their level of confidence.

We looked at two different elements of each of these tools. So you look, within the question there is, right, when we look at a PRO question, we try to tease apart the

meaningfulness, right? One of the domains of interests is what, is it valid, and another is, is it generalizable, so we tease this apart. So for part one, valid, we looked at what are

called the measurement properties. These are really common characteristics of measurement tools that help us to understand if, you know, if it's a good tool, if it's meaningful, if it's reliable.

So we looked at first what's called content validity. This is whether patients were involved in the development of the tool to understand if the terminology in the questions is widely understood by patients, and if that terminology actually maps to the concepts that you're trying to evaluate. If it's a pain questionnaire, are you actually measuring pain, are you measuring, you know, fatigue and, you know, other stuff?

Number two, reliability, right? This is test-retest. If you ask the same person the same question over and over again, will it give you the same answer because the responses in the questions are really clear to them or, you know, are they all over the map because the

question wasn't well structured and it's confusing, right? If you ask me, you know, one

day if my pain is intolerable and another day if my pain is excruciating, you know, that's not a good question, you know, I might feel both.

The third is what's called construct validity, this is whether the question, if it really reflects the underlying thing you're trying to measure. So if you're asking about sensory peripheral neuropathy, are you really measuring sensory peripheral neuropathies? This is generally done by looking at the correlations or the associations of the responses in the question to underlying anchors. So you know, for example if you're looking at peripheral neuropathy, you would expect to see different scores in people who are on taxanes or platinums or, you know, other products that might be expected to interact with neuropathy.

And then finally, clinical response, this is really change over time, do the scores change over time as expected?

So that's the valid part.

Then for generalizable, we looked at whether the tool's been used in CAR T. As I already explained, you know, it's a new product so there's not that much accumulated experience, although there are 11 trials, not bad, but we looked also at the use of these tools in the 65 and older population, just to look at general pertinence and prior use of the tool in the Medicare population.

Now question two is a little bit more nuanced, I would say, and I want to just take a moment on this before I launch into each tool. To remind everybody, question two asks, considering all PRO assessments in question one with greater than or equal to a score of 2.5, please vote whether or not those PRO assessments combined have available supporting evidence on each of the following desired characteristics, and there are eight of them.

So I'm just going to very quickly explain my interpretation, my interpretation may be different from other people's interpretation, but this is how we thought of this. A, the breadth of the measures in emotional, social and physical well-being. We

looked at whether the PRO tool includes those specific quality of life domains of emotional, social and physical well-being within the assessment of the tool, that's a little more straightforward.

B, quick throughput to apply to clinical studies. We thought this just means you can kind of get it up and going, you know, pretty quickly for a trial.

C, is it transferable to community practice settings? So we just simply looked, has it been tested and used in community settings in trials?

D, the measures are not sensitive to differences in age, so we looked at whether there's evidence that age alone does not sway the scores, right, so if you have people of differing ages and, you know, you put age into your model, it doesn't, you know, mess with your results, your findings.

E, the same question but for line of therapy and so similarly, we looked to see if line of therapy sways scores.

F, the measures are not sensitive to comorbidities. You know, this is more

challenging, because you can look at whether comorbidities themselves impact scores, right, the way we do for, say, case mix adjustment and quality assessments in clinical practice but you know, this could be a problem if the comorbidity has symptoms, right, the patient with severe rheumatoid arthritis, they may have pain from their rheumatoid arthritis and they might be reporting on that pain, so if you have a question that is agnostic to the etiology of the pain, you might pick that up. That's a little more challenging and nuanced, and I'll address that a little bit as I go.

G, the measures are generalizable to study of combinations of therapies. We simply looked, has the tool been used with combination therapies and comparing combinations versus single agents.

And then finally, used in net benefit analysis based on symptom burden and well-being. To be honest, I found, I wasn't sure exactly how to interpret this. The way that we used it for our analysis, we looked to

see if the tool involved some sort of global metric for the overall impact on the patient of

the patient-reported outcomes. I think there's another way you could interpret this, which would be, has the tool been used in net benefit analysis of the value of a treatment, and if that's the case, almost all the tools would be a yes. That's not how we interpreted it.

Okay. So I'm going to get right into it. So the PRO CTCAE, the patient version of the CTCAE, as a disclaimer. When I was on faculty at Memorial Sloan Kettering, my group worked under contracts to the National Cancer Institute to assist in the development of this tool. We do not currently derive any funding for this, this tool is the intellectual property of the NCI, so I may have intellectual interest in this tool but no financial interest.

So again, as Paul Kluetz mentioned in his FDA presentation, this is what we call a library, a PRO library. It's a bunch of individual items. You can pick and choose from

the library; if you want to measure fatigue, you choose that item; if you want to choose sleep disturbance, you choose that item. It was developed for the explicit purpose of

patient adverse event reporting, really symptomatic adverse events, it only includes symptoms, not other kinds of AEs. You wouldn't ask a patient about retinal detachment, for example. There are 78 adverse events included, it is mapped directly to existing standardized lexicons for adverse event reporting, specifically CTCAE and MedDRA, which are commonly used in non-oncology trials. This is free use, no license is required. There is an adult version in use and a pediatric version that's close to done.

So, these are the questions that I alluded to earlier. Looking towards your question one, areas of valid and generalizable, if you look at the measurement properties for whether we think it's valid, it is, the tool does well across all of these categories, confidence, validity, reliability, construct

validity and clinical responsiveness. It's in use in two CAR T trials and it's been used in numerous trials in the Medicare 65 and older population.

In looking at question two per your specific questions, for the breadth of

measurements in emotional, social and physical well-being, this is really nonapplicable for this tool, this is an adverse event tool, it's not looking at quality of life per se in terms of social or emotional or physical well-being. I mean, it alludes to that, right? Adverse events can impact these things but it's not directly measuring them, so that's not applicable.

It has been, it's ready to go, you can drop it into a clinical trial tomorrow so, you know, there's quick throughput. It's transferable to the community settings, it's been tested extensively in that setting. The measures are not sensitive to differences in age or line of therapy. Comorbidities don't impact the scores but again, you know, if you

have a patient who has a lot of symptoms or side effects from their arthritis drug, that could be picked up. It's generalizable to combinations. It does not have a measurement of global benefit. The tool can be used to evaluate benefit if that's your interpretation, but there's no overall general global measurement of net benefit in the tool.

Now as to the M.D. Anderson Symptom Inventory, this is a very well established tool, it's been around for a long time. It has 19 items, 13 specific symptoms, six questions about how the symptoms interfere with functioning. This is a symptom questionnaire, not a quality of life questionnaire. It is widely widely used in clinical trials. It also performs very well in all of these question one areas around it being valid and generalizable. It hasn't been used in CAR T but it's been used in many trials in the 65 and older population. It also does very well on the question two items. It does allude to emotional, social and physical well-being in that it looks at the

impact of symptoms on these areas. It does very well in all of the other categories, quick throughput, community settings. It's not sensitive to differences in age, therapy, or comorbidities, with the caveat that I mentioned earlier. It has been used in many studies with combinations and again, benefit, it does have a question as to the overall impact of symptoms on well-being so again, depending on your interpretation of this, I would say probably

it's a yes there.

The EORTC QLQ-C30 is also a very very well established, well traveled tool. It has 30 items, five for physical functioning, 14 symptoms, and multiple multi-item scales for cognitive, emotional, physical and social, so it has it all, right? It's got the physical functioning, it's got the symptoms, it's got the quality of life domains. It's not designed for adverse events reporting, however, so you know, that's not a purpose of it, but for efficacy or effectiveness, this is a well traveled tool.

It also does extremely well across all these valid and generalizable categories, although the content validity wasn't initially tested because the tool was developed before content validity was a prominent expectation of this kind of tool development, this was done subsequently. It performs extremely well in your question two items, really across all the categories.

I won't belabor these because you have my slides, except to highlight H again, the net benefit. It does ask about overall quality of

life, it does have sort of these overall health status questions, so by the criteria Dr. Atkinson and I applied, I would say it probably is a yes there.

The University of Washington Quality of Life Tool is not a well-known tool. This was included in the CMS packet and I was asked to comment on it, and we did do all the searches on it. It has six items for physical functioning, six for psycho-emotional function. It doesn't really do as well in question one.

Content validity was not established. Reliability, not in English, only in Spanish and Chinese. It has had construct validity and responsiveness assessed. It's really been used in a very limited number of trials, both in the under and over 65 populations. It has not been used in CAR T. And again, it really doesn't perform so well in question two. You know, it probably could be used quickly in clinical trials, it just hasn't been used in that many clinical trials. It's unclear, you know, whether age, line of therapy or comorbidities would impact the scores.

PROMIS, the Patient-Reported Outcomes

Measurement Information System, is a brief, precise, fixed or tailored tool. It was developed with NIH funding. This is also publicly available. It includes physical, mental and social well-being, as well as pain, fatigue and sleep. This tool was meticulously developed. It offers short forms. It offers a sort of computerized adaptive testing approach, kind of like if your kids took the SAT, you

know, depending on how you answer one question, there's a subsequent question to get a more precise estimate of what your actual score is. But there are also single items, it's, you know, really flexible. There are adult and pediatric versions. It does extremely well in all of your question one categories. It's being used in two CAR T trials currently. It's been used in many studies in the 65 and older population as you can see above. Its measuring properties really have been pristinely tested. It also does extremely well in your question two items really across the board. Again, you know, I raise that question about age and net benefit. You know, again, it does have overall quality of life items in it so I guess, you

know, I would probably say yes, depending on how you interpret that.

You know, the real limitation here is that there are only a small number of domains that you can measure with PROMIS, and so if you want, you know, a wider number of things to be measured in a trial, you're going to need

PROMIS plus something else, but for what it measures, it really is excellent, in my opinion.

The ESRA, I was asked to comment on as well. This really is not well known. As alluded to in Katherine's presentation, this is really not a well-known tool. I'm sorry, this is not a PRO measure, it's an electronic questionnaire system, so it's not really a PRO. It happens to include three PRO measures in it, the QLQ-C30 which you've heard about, the PHQ-9 which we haven't talked about, which is actually an excellent measure of psychosocial distress, anxiety, depression, which is commonly used to assess depression, and the Symptom Distress Scale, with is really kind of a lesser used symptom scale. It's not really been well tested, it's been very few trials

and, you know, because it's been used so seldom, because it's not a PRO, I didn't even go into evaluating it for questions one and two, it wouldn't do well because it just hasn't been tested in that way. To me it's not

applicable to these questions.

The FLIC, this is a PRO tool but also not really well known, it's an old measure. It had 22 items, it has physical, emotional, social function, well-being, pain and nausea. We could actually only find one cancer trial using this tool, so really not well traveled in the oncology space. And so again, we really didn't really go through the 1.A -- I'm sorry, the questions for one and two for this, because there's just no data to evaluate, it really wouldn't perform well, again, in our opinion.

This graphic unfortunately, didn't come over well when conveyed to CMS. It may have come over better in the size it was printed, or maybe it was censored.

(Laughter.)

The double question marks from me were smiley faces and the other ones were sad faces, but maybe CMS felt they should be a little

milder. But in our opinion, the ones with these double question marks are tools that are well established, well tested and perform well,

and if I were designing a trial, with the caveats that Paul mentioned in his FDA presentation, we want to make sure that the tools are appropriate to the domains of interest. These are tools that I would be comfortable considering, but the frowning faces, not so much.

All right. So just in the last four-and-a-half minutes before I finish up, there are some additional questions to the panel. Are there other PRO assessments to consider? I would say yes. One in particular that I'd like to highlight called the FACT GP-5, and this is a single item that asks people if they are bothered by the side effects of their treatment. This is a global assessment of side effect burden. This is a very helpful companion to the PRO-CTCAE, right? Just to remind you, the PRO-CTCAE is the tool where patients answer individual items about their individual symptom side effects, right? Do you have sleep disturbance? Do you have

taste disturbance? Do you have myalgia? This

is a global to go along with it. This is a five-point response scale, it's well developed, there's broad interest in using this, it's been alluded to in numerous past FDA and EMA presentations.

Are there additional desired characteristics besides those in question two? Yes, I think so. First the general, what we call measurement properties, all these things that Dr. Atkinson and I actually commented on in our responses to you, content validity, construct validity, reliability, sensitivity or responsiveness, these are really key measuring properties of an assessment tool, and really both need to have been tested and demonstrated to perform well for a good tool.

Prior testing in populations with cancer. The availability of language translations, this is essential not just in the U.S., but particularly outside the U.S. for international trials. And then, you know, I'd say really importantly, does this include items that are salient to the CAR T population? There really needs to be evaluation in this

population, probably qualitative with interviews, asking patients about what's going on with them very broadly so that we can understand what are the outcomes that are salient to this population, so we can then say is this the right PRO instrument to use?

And this is really on the sponsors, right? The sponsors spend a lot of money developing their measurement tools, conducting these trials. This is an essential part of understanding the patients' experience. The sponsors should be going out to patients in their trials asking what they're experiencing so they can substantiate the PRO metrics in their trials and particularly in their registries. I think in the real world, not just in the registration trials, this information needs to be collected.

All right. In conclusion, patient-reported outcomes provide valuable information about the patient experience and about the characteristics of products that cannot be well captured in any other way. There are well developed available patient-reported outcome tools that can be used

readily in CAR T trials that could be used tomorrow. Yes, we can do more work to hone it down, to get more specific to figure out what exactly would be best to measure, but these tools are shelf ready in many cases, but we should do further work to really hone down and understand what are the outcomes of interest.

Assessment of physical function, symptomatic adverse events and disease-related symptoms should be considered in any given trial of oncology, including in this population. Thank you very much.

DR. ROSS: Great, thank you, Dr. Basch, right on time, and to Dr. Atkinson for his support of this presentation.

So now we are turning from our, to the scheduled public comments portion of our meeting. Each speaker will be given six minutes to speak and we have one, two, three, four, five, six speakers, because one was unable to attend. And we are, as each speaker comes to the podium, I ask that the next speaker comes to the chair to keep us moving efficiently, and just as a reminder, to

25 disclose, introduce yourselves and to disclose

1 conflicts of interests. And our first speaker
2 is Dr. Kathryn Flynn.
3 DR. FLYNN: Hi. So, just a note that
4 we submitted slides before we knew how long we
5 would have to talk, so I will be skipping over
6 some slides, but they are all available of
7 course online. So yes, I am Kathryn Flynn, I'm
8 an associate professor of medicine at the
9 Medical College of Wisconsin, and I am also as
10 of November last year, now senior scientific
11 director for patient-reported outcomes at the
12 Center for International Blood and Marrow
13 Transplant Research, the CIBMTR. So I am here
14 representing the CIBMTR, CIBMTR paid for my
15 travel to attend the meeting. I don't have any
16 personal financial disclosures related to
17 CAR T, but CIBMTR as an organization receives
18 federal funding from NIH, HRSA and the Navy,
19 and as you heard earlier today, has a cell
20 therapy registry contract with Kite.
21 So CIBMTR, for those of you who aren't
22 aware, collects and maintains clinical outcomes

data on all allogeneic transplants as required by U.S. law. The centers also voluntarily submit data on auto transplants, and worldwide

centers additionally submit data voluntarily. So related to blood and marrow transplant research, we, the registry has information on nearly a half million, 475,000 patients that are included in the database.

And we are now in the process of implementing an e-PRO system that will be available for use by the registry and the affiliated trials network, the BMT CTN. So, I will skip this one if I can. No. There we go.

So we looked last year at the BMT CTN studies that have collected PROs, and in 18 trials performed since 2004, half of those had included as a primary or secondary outcome a patient-reported outcome measure. Many different measures have been used in these studies, most commonly the SF-36 and FACT-BMT. But we were looking to make some recommendations going forward and thinking about the implementation of this e-PRO system,

of what to recommend and so -- I'm having trouble with this. I have to press it really hard, I guess. Okay.

So we had a couple of recommendations, first to use the same core measures in all

research studies of HCT patients, use a system that's free and easy to access, try to ensure a low burden for the patient who's of course undergoing a difficult treatment, using a single versatile measurement system for core concepts supplemented with additional measures as necessary. And so thinking about the registry context, the core system that was recommended in this article by Brown and Shaw was PROMIS. Even pressing really hard, I'm having some difficulty there. Okay. I don't know if it needs new batteries potentially.

DR. ROSS: Don't worry, you can have another minute.

DR. FLYNN: Okay, thank you.

So we've already hear about PROMIS, I'm not going to go into detail there, but it met those recommendations that we were hoping

for. Okay.

So just to reiterate a point just made in the last talk, really the most appropriate PROs to collect in cell therapy are unknown, so there really is some foundational qualitative work that needs to be done. We can probably make some good guesses about some of the

domains that will be, that will need to be measured, but to get into more specifics, there does need to be some additional work done, I think. However, once relevant constructs are identified, there are absolutely multiple available high quality measures that can be used, and can choose the appropriate measures and time points at that time.

Centers need a structure and process to systematically collect PROs, and so what I'm going to do with my remaining couple minutes here is just describe the components of our CIBMTR e-PRO system. So as you can see here, the e-PRO system is the integration of electronic patient-reported outcome collection with our existing systems for collecting other

information, clinical information. So in the bottom right we use Salesforce to track our studies, participants, time points, activities. At the bottom left is our integrated data warehouse where the clinical outcomes data from multiple sources are stored for research retrieval. Top left as I mentioned, we did identify PROMIS measures as that core system, but certainly other measures can be added as

necessary, and so certainly for some of the trials within the BMT CTN already we're adding items from the PRO-CTCAE for those specific studies.

And then to the right, note that we're using Qualtrics as the patient interface for administering patient-reported outcomes, so a very flexible user friendly system for patients to complete those PROs.

So, this system was developed with funding from the Navy grant, our partner, the National Marrow Donor Program, and our pilot e-PRO study just started this summer. It is a six-site pilot trial where we're examining

quality of life and PROMIS measures in patients as part of the CMS MDS study. This is just cross-sectional to explore the use of our system, but certainly longitudinal studies will be feasible as well.

There is, just a note here, this is just a brief overview of kind of the study procedures, but to note that significant planning and effort is required to manage this central coordination of multisite PRO data collection in terms of following patients at

multiple sites and getting their, you know, being able to contact them directly, when previously through the registry they are only contacted by their local center, and so for the CIBMTR directly to contact them is new.

And then the last thing I wanted to mention is related to this. We've recently organized a multidisciplinary working group of about 30 or so people with expertise in many different fields as part of a late effects task force. And again, this is in the context of BMT, but our goal is to develop a strategy for

the collection of late effects in patients that are reported to the CIBMTR. So of course it's a very heterogeneous population who's receiving transplants, and so focusing on which populations we should focus on to get kind of routine PRO collection, what domains we need to focus on, what measures to use, what time points, these are all questions that we're answering within the context of this task force, and we have a nine-month time frame, we started this summer and we're going to present our recommendations at the Transplant and Cellular Therapy conference which, in February

of 2019. That's it.

DR. ROSS: Thank you. Right on time. Our next speaker is Karen Chung, the senior director of health economics and outcomes research for Juno Therapeutics.

DR. CHUNG: Good morning, everyone. Again, my name is Karen Chung, senior director of health economics and outcomes research at Juno Celgene. I have been involved in patient-reported outcome strategy analysis,

communication, for over 15 years in the pharmaceutical industry, and I'm currently employed by Juno Celgene and do have stock options with them as well as other companies.

Celgene is developing investigational CAR T-cell products which are not FDA approved, and any data we discuss today is subject to change. CAR T-cell agents are novel agents which fulfill an unmet need in patients who have not responded to front line therapy, including Medicare patients. They have limited effective treatment options as well as limited survival. CAR T-cell therapies have been administered across sites of care and as novel therapies have a long-term follow-up to

continually assess efficacy as well as safety. And while AEs are specific to each CAR T-cell therapy, AEs are being increasingly identified very quickly and managed very efficiently. And while PRO measurement is important as it represents the patient voids, it is very complex from the clinical trial perspective and even more so from the clinical practice

perspective.

Celgene is developing two CAR T therapies which have the potential to significantly transform patient outcomes. JCAR017 is a CD19-directed CAR T-cell therapy for non-Hodgkin's lymphoma. bb2121 is a B-cell maturation antigen-directed CAR T-cell which is currently in clinical trials for multiple myeloma, and the other was for non-Hodgkin's lymphoma. Each CAR T-cell therapy has a unique targeted patient population, safety profile and manufacturing process. As the science of CAR T is rapidly evolving, we urge CMS to provide flexibility to consistently ensure patient access across all these disease states.

While we strongly support the incorporation of the patient voice into

clinical trials, we firmly believe PROs should not be a condition of coverage due to the significant barriers in the clinical practice.

And again, while we don't think PROs are appropriate for coverage, we did want to take a look at the question that CMS had asked

the panel to consider, and of the seven instruments that were delineated, we feel that four of the seven instruments could be appropriate for clinical trials involving the Medicare population.

The first is the PRO-CTCAE which Dr. Basch has mentioned. It does cover a wide range of symptoms and so for symptom assessment, it is a very useful tool.

The MDASI, or M.D. Anderson Symptom Inventory, covers a wide range of symptoms.

The EORTC-QLQ-C30, which we implemented in the JCAR017 and bb2121 trials, is a comprehensive instrument that assesses symptoms, functioning, as well as health-related quality of life.

The last instrument is PROMIS, which is basically an item bank, which also covers various symptoms as well as functioning.

This next question is really considering all these instruments together, and together, we feel that they have to have the breadth of measurement specifically in

emotional, physical as well as social well-being. They can be applied and have been applied to clinical studies and can be used in the clinical practice setting as well.

We didn't, we felt that they were sensitive to differences in age, lines of therapy, as well as comorbidities, and felt that they were also generalizable and can be used in combination therapy trials.

From end to end, PRO implementation in clinical trials involves significant resources in terms of both budget as well as head count. We need to support instrument selection, licensing, site training, data collection, analysis, as well as interpretation. PRO assessment in clinical practice is typically even more challenging due to the lack of infrastructure. Institutional barriers could include the healthcare provider burden, the additional FTEs that are necessary to coordinate administration and data collection,

and the lack of consensus on which is the most appropriate patient-reported outcome tool to

use. And then there's the, following the scoring, the expertise needed in scoring and analysis as well as interpretation.

Perhaps even more notably are the patient barriers, and so we're asking these Medicare patients who are typically very sick, third line and beyond, to respond to these questionnaires. They may have poor performance status and they may also face technology barriers as we move to more electronic platforms to collect this data, so it's something that they might not have the experience to really manage to do well.

So while patient-reported outcomes are key measures in hematology and oncology trials, including the CAR T-cell therapies, there are important considerations, which includes the wide range of tumor types and stages, also the broad areas of concepts. You know, are we interested in physical functioning, disease-related symptoms, adverse events, or health-related quality of life, you know, which do we focus on. And due to the diverse nature

and range of symptoms across and within tumor types, as well as the administrative burden, assessing patient-reported outcomes with validated instruments is complex.

Celgene has incorporated relevant PRO assessments in CAR T-cell clinical trials to complement clinical safety and efficacy data, which we feel is very important. However, while we feel it's very important in the clinical trial setting, we don't feel they should be a condition of coverage.

DR. ROSS: Great, thank you very much.

DR. CHUNG: Thank you.

DR. ROSS: Our next speaker is Dr. Surbhi Sidana, from the Mayo Clinic.

DR. SIDANA: Good morning and thank you for this opportunity. I am a hematologist/oncologist and I am not a PRO expert, but I'm leading two studies of PROs, including one of CAR T, and I just want to speak to the panel of the challenges we have faced in trying to design and lead the study. So, here are my disclosures, and ASBMT is paying for my travel to this meeting.

This data has already been shown so I

will not belabor this data anymore. However, CAR T-cell therapy is a novel therapy which has shown exceeding promise in patients who did not have other treatment options before. It has unique side effects, and some of the side effects we are not even aware about in the long term.

There is, the process for assessing PROs has already been discussed in detail and so I want to focus on the approaches of assessing PRO in patients with CAR T-cell therapy. We have conducted several studies in the last couple of years in hematology which have used various methods of assessing PROs. So let's focus on the challenges of conducting the study, and this is from my personal experience in conducting the study.

So what is an optimal outcome that we should use and what instruments should we be using? Seven instruments are being asked, you know, you're rating seven instruments today. In my study I'm using a completely different instrument because on my clinical judgment I thought that was a better instrument, along with some of the instruments we're reviewing

today. So even though we have validated instruments, not everybody agrees that those instruments should be the same in different studies.

Second, how do we account for missing data? A lot of patients who are undergoing CAR T-cell therapy will have side effects, get in to the ICU, and these patients potentially will have significant missing data leading to bias. A lot of times patients come to referral centers like Mayo Clinic for their treatment, and then they go back to their local doctor. So if we are going to use long-term data, we might miss patients who are now gone from the referral center.

And then the third thing, do we just collect this data or do we do something about it? As a doctor it's challenging. You're asking patients to give their symptoms and then you feel you're ethically obliged to do that, this also keeps the patients engaged. However, there are problems with that. It requires a huge infrastructure. It also requires

consensus to say when are we going to intervene. For example, if you ask a patient

for pain, do we intervene for a pain at seven out of ten, eight out of ten or nine out of ten? Is seven different than eight? And similarly for other symptoms as well. That will also require a lot of resources that centers and the community will not have present.

The other thing that is challenging, we want to assess how is the patient's quality of life in respect to the side effects they experienced initially, and that's problematic because right now all the different CAR T trials are assessing toxicity differently, Grade 3 CRS in one trial is not the same as Grade 3 CRS in another trial. The management of toxicities at my institution is very different from management of toxicities at another institution, so this is going to impact how we interpret this data and what this data means.

And then as many people have already

22 alluded, CAR T-cell studies are currently being
23 conducted in various hematologic and oncologic
24 malignancies and currently are approved for two
25 diseases, ALL as well as non-Hodgkin's

1 lymphoma. We expect that soon they will be
2 approved for other diseases like multiple
3 myeloma, and the short-term toxicity has really
4 varied across different trials based on what
5 instrument, what construct and what disease.
6 For example, a lot more CRS was seen in
7 non-Hodgkin's lymphoma than was seen in
8 multiple myeloma, so how can we put all of
9 these patients together with different diseases
10 which have different symptoms, different
11 constructs, and say we're going to measure all
12 of these the same?
13 And then, what is our benchmark? As
14 has been shown before, these patients with
15 non-Hodgkin's lymphoma previously did not have
16 many treatment options, their median survival
17 was six months, and now it's not being reached.
18 So how do we decide what's reasonable quality
19 of life or what's reasonable physical function

in these patients? How do we compare them to historical controls or even how do we compare them to their baseline what is reasonable?

So I think there's a lot of room for study at this point. We are conducting pilot studies at my institution and several other

institutions to address what's the feasibility, where is the missing data, how can we do this better, and do we need specific measures specific to CAR T-cell therapy? And then in the context of a working group, we need to come up with a consensus before we design a larger-scale study. I think at present we need at least 12 months to come up with a consensus based on preliminary data from our study and the studies being done at other institutions. Thank you.

DR. ROSS: Thank you, Dr. Sidana. Our next speaker is Dr. Cori Abikoff, the medical director for CAR T at Novartis.

DR. ABIKOFF: Thank you very much for allowing me to speak today. I'm Cori Abikoff, I'm a medical director for the CAR T program at

Novartis Pharmaceuticals Institution. My expertise is in pediatric stem cell transplant as well as adult and pediatric apheresis. I am a paid employee of Novartis.

Kymriah, the Novartis CAR T product, is the first FDA-approved gene therapy product on the market. It is currently approved in two indications, both pediatric and young adult

relapsed or refractory ALL, as well as adult relapsed or refractory large B-cell lymphoma. It's been extensively studied in clinical trials, both for validated clinical outcomes as well as PRO data, as was previously presented by my colleague, Dr. Ilia Ferrusi. It also continues to be studied in the outpatient, in the commercial setting under a risk evaluation and mitigation strategy.

As was previously discussed, chimeric antigen receptor therapies essentially are a living drug, which allows the patient's tumor to be targeted by the patient's own immune system through a process of gene modification. This is a complex process that requires that

the patient's own immune cells be removed, gene modified, and returned to the patient in a setting which has a degree of complexity that means that the timeline must be observed due to the significant burden of illness in these patients.

Novartis has chosen to study a population of patients who have significant burden of illness. Although pediatric ALL is not a common condition, it is the most common

cancer of childhood, and relapsed and refractory ALL represents the most common cause of childhood cancer death, falling only behind accidental injuries and inflicted injury, whereas diffuse large B-cell lymphoma is a more common illness and one that is more likely to affect the Medicare population.

In both cases when the disease is relapsed and refractory, there are incredibly limited treatment options, and these usually require incredibly toxic therapies that in order to reach standard of care with even acceptable outcomes requires the use of a stem

cell transplant.

In the JULIET trial where we treated patients with diffuse large B-cell lymphoma, you can see that approximately a quarter of our patients were over the age of 65 and these patients were heavily pretreated, with more than half of them having received three or more prior chemotherapies and having been refractory or relapsed to those therapies, and almost half of these patients having already received a standard of care therapy of autologous stem cell transplant.

Unlike the data that's previously been shown regarding complete responses as low as seven percent, the JULIET trial had a best overall response of 52 percent, complete response rate of 40 percent. This is really unheard of in this population. And when we look across the groups again, you can see that the patients aged 65 or older had a 59 percent overall response rate, consistent across all subgroups with the overall response in our trial.

But more importantly is not just the response, but the ability of these responses to be sustained, and you can see that in patients who were complete responders, there was a 95 percent overall survival at one year and 78.5 percent of patients were relapse-free during this time point.

In addition because of the living nature of this drug, patient response is not determined by their initial response, but in fact 54 percent of patients will progress from a partial response to a complete response over time frames as long as nine to 12 months.

These are not benign therapies, and

certainly we acknowledge the adverse events that need to be followed. Here in the JULIET trial you can see that adverse events greater than, at Grade 3 or higher, included 23 percent of patients with CRS, and 18 percent of patients with neurological toxicity. We also evaluated toxicity such as infection, and longer-term toxicities such as hypogammaglobulinemia.

It is important to understand that Novartis too has begun collaboration with the CIBMTR in order to provide a registry which will follow 2,500 patients, including at least 1,500 patients with diffuse large B-cell lymphoma, for 15 years after their therapy. This is in accordance with the FDA guidelines and includes an incredibly robust amount of information, including patient-level characteristics as well as disease characteristics, and the efficacy and short- and long-term safety information that can be followed for these patients. By partnering with the CIBMTR, we choose a leader in registry data for cell therapy, and one that all of our sites are familiar with. By doing so, we

believe this will encourage early and robust use of this registry data, and encourage and ensure that the real world data that's collected really reflects the patient population who is being treated with Kymriah. In addition to this by partnering with the CIBMTR, the data is not only owned by Novartis

but it actually belongs, in fact belongs in the purview of CIBMTR, allowing access to that data and the analysis sets that can be considered to be done by CIBMTR and their research networks, as well as Novartis and health authorities.

As a clinician not far out from being part of the care provided to patients who would be receiving Kymriah, I am not, the importance of treating patients and including them in decisions about their care is not lost on me, but Novartis does urge CMS to leverage the existing data as well as the robust mechanisms for further data collection in order to make decisions about how best to approach payment decisions. Thank you.

DR. ROSS: Thank you, Dr. Abikoff. The next speaker is Dr. Merav Bar, assistant member of the Fred Hutchinson Cancer Research

Center.

DR. BAR: I am Merav Bar, I'm an assistant member at the Fred Hutch in Seattle and I'm a transplanter, and I also take care of patients after CAR T-cell therapy, and I'm also

part of the long-term follow-up team for patients after transplant, and we are now building also our long-term follow-up for patients after receiving CAR T-cell therapy. And today I'm mainly focused on question number four regarding timing of evaluations of PROs in patients after CAR T-cell therapy and mainly for the long-term follow-up of those patients.

My disclosure, I have no personal financial or intellectual conflicts of interest. However, I just learned after I submitted this slide that a member of my family has shares in Bluebird.

For long-term follow-up of patients after CAR T-cell, most patients participating in CAR T-cell studies have been followed only for a short period of time, most studies for one or two years after they receive treatment. And the two commercial CAR T-cell products have only been approved in the last year by the FDA.

Therefore, the data regarding those patients is also limited in time. So, currently there is only limited data regarding the long-term

effects of those treatments.

Main concerns regarding long-term effects of CAR T-cells are prolonged B-cell aplasia with a hypogammaglobulinemia, acquired infections secondary to that, subsequent malignancies, and also new incidence or exacerbation of neurologic or autoimmune disorders.

There are objectives of a long-term follow-up after CART T-cells, which are to identify and mitigate the long-term risks of patients receiving treatment, and capture delayed adverse events.

There are several challenges in long-term follow-up of patients after CAR T-cell therapy. Most of them are the heterogeneous patient populations, the variety of the constructs of the CAR T-cells product. Although currently the two approved products and also for most of the products that are under investigation target the CD19, in the future we will see more products with different

targets, that they will affect the toxicity and

the safety profiles of those products.

There is a transition of care of the patients; most of the patients come to big centers in order to get the CAR T-cell therapy. However, after a short period of time of a month or two they return back to their referring physician, so it is a challenge to follow them for the long term.

Although there is very good responses that have been reported with the CAR T-cell products, there is still a relatively high rate of relapse of those patients and therefore, the patients are subsequently exposed to other treatment which will affect how the patients are feeling, their quality of life and side effects that you would see in the long term. And additionally, patients have multiple comorbidities that will affect the PROs.

And there are also specific challenges when you are talking long-term quality of life after CAR T-cell therapy. So for example, there is no validated instrument for quality of life and we see that there are different options that can be used, there is lack of

uniformity between centers. So although there is a number of centers that incorporate PROs into evaluation of patients after CAR T-cell therapy, there is no uniformity, and also, we don't know what optimal study design is.

In addition, other people here also reported about the significant resources that are indicated, so we need the resources in order to build the questionnaires into electronic forms, to follow-up with the patients after leaving the treatment center back to their referring physician, and we need a lot of resources in order to collect the data and then to analyze the data.

In our institution we right now are studying a pilot study to evaluate a patient after CAR T-cells and the objective is mainly feasibility, and we are using mainly the PROMIS Global Health and PROMIS-29, which have been validated in the transplant setting. And currently as I said, there is a variability between centers and there are only a small number of studies that are currently ongoing, and we support a collaborative work group in order to provide recommendations for the

instrument to be used, unify the study design, harmonization of the data, and potentially define a multicenter study between institutions. So currently, we think that efforts should be made in order to incorporate the PROs in CAR T-cell studies. However, we don't feel that PRO should be mandated for payer reimbursement for CAR T-cell therapies.

DR. ROSS: Great, thank you, Dr. Bar. Just before, I want to confirm that Dr. Heather Jim is not in the audience because she wasn't able to get here today. Good.

So our last speaker will be Dr. Gunjan Shah, hematologic oncologist at Sloan Kettering, who's representing the American Society for Blood and Marrow Transplantation.

DR. SHAH: Hi everyone, thank you for allowing me to speak with the committee. I am a bone marrow transplant physician and also work on cellular therapies, as well as part of the health-reported outcomes program at MSK, and I am receiving travel funds today and am speaking on behalf of our program as well as the ASBMT.

25 So you've heard a lot today over the

1 last several hours about what patient-reported
2 outcomes are and the differences with the
3 different scales, and we agree with a lot of
4 the comments already presented.
5 What I'd like to do with my time today
6 is present to you how we have used several
7 different scales and changed over time and
8 incorporated them into different trials, as
9 well as how we are converting these into a
10 standard of care approach across our entire
11 service, as well as for the CAR T-cell
12 patients, in terms of how to capture these by
13 paper surveys and our conversion to an
14 electronic process, and whether we're going to
15 use them for research and clinical care, and
16 how that works.
17 So, I present this today just as a
18 review article that was done in Transplant
19 looking at 114 studies, and you've learned
20 today along the way of how many different
21 patient-reported outcome measures there are,
22 and why they can be used in different ways, and

how they do tend to cluster around certain symptoms and certain assessments that can be used at different times.

On the upper right you can see a picture of sort of the different subscales of the MDASI that are disease-based, and what we've used over the last five to ten years in many of the transplant trials, specifically the autologous transplant trials, has been the MDASI myeloma scale. And what we've been able to do in that and the reason we use it is it's been able to be done at several time points through the first 30 days, and you've seen today that there are differences in kind of the scale of early toxicities and sort of later recovery. And what we've done is been able to look at changes over time using an area under the curve method, and so being able to condense a lot of that information into one data point that can be compared, especially in intervention studies where you're really trying to affect the system burden as opposed to just collecting some of this information.

On the bottom right, you've seen this already today, is the PRO-CTCAE, and we've incorporated this into more recent trials and used the symptom bank in a way to actually incorporate similar questions to the MDASI to

see if patients really answered the questions the same way. We also in our long-term maintenance trials and microbiota trials have specifically taken out the questions that are related to diarrhea and constipation and other GI symptoms, and have been able to correlate those with the collected stool samples.

On the upper left you see the PROMIS score that's also been described many times today, and the reason I present this here is that we are in the process of converting from the MDASI over to the PROMIS scale to better be generalizable across centers and as you've seen, you know, the plans from the CIBMTR and several other centers that have presented today, and so in an effort to be able to combine data, we are switching over to this scale.

The bottom left, you can see sort of what the paper version of a survey looks like, and sort of a scale system of this as being a five-point scale versus some others being ten-point scales.

Our informatics colleagues and surgical colleagues, using a grant from PCORI,

have converted the MSK system from a paper format to what they call MSK Engage, or an online system for collecting some of this information, and we're going to adopt this over to the transplant service and cellular therapies.

On the left side you can sort of see a particular patient's symptoms over time, and this is going to be available in the clinic, that you can look at a particular patient, convert it into their electronic record, and sort of follow them over time for a particular patient. Partly this is important because we are, and our institution has determined that it is important to act in some way on this information in real time, and so you can set

criteria of if you are above a particular score, that they will send a message both to the patient to call the office, but also to the office practice nurse to call the patient and determine if further things need to be done about it.

On the right side you can kind of see information sort of that was presented by other colleagues today of how do we present that

information and what do we do with it in terms of both a research and clinical following over time. And so we have software where you can aggregate this data across trials, across patients, and present data in a very interesting way to be able to look at both intervention trials, as well as just following over time.

And so we're going to be incorporating all of this into our proposed new plan going forward.

And so, we know in the CAR T-cell space that patient-reported outcomes are still in development and too early to mandate in

terms of coverage. However, we do agree that these are important to capture and study in both the clinical trial and commercial setting, which is what we are embarking on as well now, that we are going to use the PROMIS scale, PROMIS-29, and do weekly assessments, and follow that with monthly assessments for the first year using our electronic system, and be able to capture whether this is partly feasible and partly their scale over time.

One of the interesting things in this

and part of the discussion in our switching from MDASI to PROMIS was the time frame of all of this, that the MDASI scale was in a 24-hour recall period versus the one-week recall period of the PROMIS scale. There are sort of pluses and minuses obviously on both sides of this, but one of the things that, we think that some of the missing data can probably be accounted for by having this every seven day scale, that there are those days where you were in the ICU or you weren't able to answer some of the questionnaires on any sort of every 24-hour

scale, but over the last week be able to aggregate some of that data, and potentially account for less missing data with that.

The other sort of further along questions that have been asked by the committee in terms of timing and feasibility, we do agree that the three- to six-month window seems to be the most reasonable option because of the patients going back as has been described by other people, and we do think that the use of technology can allow for more collections over time, and we look forward to working with CMS and the rest of the people who have discussed

today about doing this over time. Thank you.

DR. ROSS: Thank you, Dr. Shah. That concludes our scheduled public comment period.

We have had one individual sign up for the open public comment period and they have been told that they will have one minute at this front mic to make comments, and that is Mallory O'Connor. Please introduce yourself, and make sure to disclose your conflicts of interest.

MS. O'CONNOR: Thank you. My name is Mallory O'Connor, with the Biotechnology Innovation Organization. BIO is an industry trade association, so we do represent manufacturers of CAR T-cell therapies.

And I will be very brief here today, but thank you for your time. The Biotechnology Innovation Organization appreciates the opportunity to provide comments to the MEDCAC during this meeting on the state of evidence for CAR T-cell therapies.

BIO is the world's largest trade association representing biotechnology companies, academic institutions, and state biotechnology centers and related

organizations. We appreciate the committee's focus on developing better understanding of the patient experience and PROs in cancer clinical studies and care. BIO believes that patients must be involved in decision-making regarding their care and that patients and patient advocacy organizations play a vital role throughout the drug development process as they

know what desired outcomes, risks, and other considerations are most appropriate for their disease states and the diseased states that they serve.

We believe an open stakeholder dialogue on PROs is an important and useful exercise across many therapy areas, but we have significant concerns around the use of PROs in governing coverage decisions, particularly for this new therapy area serving vulnerable Medicare beneficiaries. It is critical to ensure that Medicare patients are able to receive timely access to the highest standard of treatment for their health condition.

We therefore urge MEDCAC and the Agency to move forward cautiously in the NCA process and not to incorporate PROs into

coverage determinations for CAR T. BIO's position is detailed further in written comments submitted to MEDCAC in advance of this meeting, and in response to the NCA. Thank you very much.

DR. ROSS: Thank you very much.

So, that concludes the morning session of the formal presentations and both the scheduled public comments and open public comment period. We are running a half an hour ahead of schedule, which I was told is a good thing, that will allow people to get into the cafeteria before the CMS lunch rush.

People are asked to return to this room in 60 minutes, by 12:30, so you actually have 63 minutes to eat lunch.

MS. ELLIS: Excuse me. When we come back from lunch, if all of the presenters could please sit in the very first row where it says reserved, for the second half? Thank you.

(Luncheon recess.)

DR. ROSS: If people could start coming in and taking their seat, I just want to remind all presenters to take an assigned seat in the front row.

MS. JENSEN: All right, we're going to get started because I want to make sure that everybody is able to get out on time to make their flights.

So for the panel and for the speakers, this is the time for the panelists if they have any questions, that they can ask any of the speakers those questions. We have an hour, so hopefully we can keep our answers succinct as best as possible, so that we can get through everybody's answers and all the panelists' questions, so that they will be able to answer our 23 questions at the end of the meeting. All right.

DR. ROSS: Great, so at this point I'll just open it up to the committee to see if anyone has questions for the presenters.

DR. GOSS: I have a couple of questions, one is for the panel members or for the speakers. Do either of the existing CAR T therapies that were approved by the FDA have a labeled claim for PRO outcomes? We heard that the FDA has a very clear set of standards for PRO outcomes. Have either of those products had a labeled claim that reports PRO data, or

do they have PRO data reported as part of their clinical trial endpoints on the label?

DR. GO: Hi again, Will Go from Kite. We do not have any labeled claim to my knowledge in our USPI for PROs.

DR. GOSS: Okay, thank you.

DR. ABIKOFF: Novartis also does not have any labeled claim with regard to PROs within our U.S. label.

DR. GOSS: Do you have it in other labels?

DR. ABIKOFF: Within our European labels.

MS. ELLIS: Excuse me. Could you please state your name for the record?

DR. ABIKOFF: Sorry. Cori Abikoff, from Novartis.

MS. ELLIS: Thank you.

DR. ABIKOFF: Within our European labels we do.

DR. GOSS: Can you --

DR. ABIKOFF: I can't speak to the specifics.

DR. GOSS: Can you suggest why it's not in the U.S. label, versus an EU label?

DR. ABIKOFF: I don't have access to that specific information.

DR. GOSS: I have another question. In terms of the CAR T trials. What percent of the patients, where both sponsors mentioned that in the pivotal trials PROs were used, what percent of the patients failed to complete scheduled assessments at scheduled time points when PROs were used, and how did you address that in terms of responder bias?

DR. ABIKOFF: I'm going to actually ask Dr. Ferrusi to respond to that question.

DR. ROSS: I want to just remind speakers at the mic, because I've been told the same, please speak up so everybody can hear and the mic picks it up. Thanks.

DR. FERRUSI: Thank you. Ilia Ferrusi, with Novartis. I don't have the exact percentage and what I can tell you is that in the JULIET study analyses of the PRO data, we focused on patients who did have a complete response or a partial response there because that's where we had data to analyze.

DR. GO: Will Go from Kite. In our pivotal ZUMA-1 study it's a single-armed design

so we did not do any prospective PROs in cohorts one and two of the pivotal study, which was the data that was used for the labeling of the USPI. We then incorporated PROs as exploratory endpoints in additional cohorts of ZUMA-1, such as in cohort three. This has not been reported out yet, so we don't have that data on hand, but this is obviously one of the challenges that we, as other speakers have said, in terms of collecting missing data.

As I said on the podium, ZUMA-7, our randomized controlled phase three global trial we are collecting PROs prospectively and it is a secondary endpoint.

DR. ROSS: Thank you.

DR. CUYJET: Aloysius Cuyjet. This question is for Dr. Basch, am I pronouncing that correctly? First I'd like to thank you for a very cogent presentation of the different PRO tools. Anytime I see seven of anything, I know one of them is not an ideal tool to provide the information. So what I'd like to ask you, what suggestions might you have in terms of improving the patient-reported outcomes process, since we have seven different

instruments to look at? What would you to do to come up with one or two ideal instruments?

DR. BASCH: All right. Ethan Basch for the University of North Carolina, so do you mean in this particular population or in general?

DR. CUYJET: Well, I haven't seen any -- it's a whole area for discussion, so I'm assuming, and I'm taking to -- I'm going back to my experience at Rutgers Medical School where we had robust end of life care, so patients would make decisions based on how much pain they were having, how much sleep they got, who in their family they spoke to, so I'm sure there's diversity in genders, there's diversity driven by cultural backgrounds, ethnicity, socioeconomic status, education status, there's a whole list of variables that we consider in how patients report outcomes, and I'm clearly not an expert in that field. So if you had to come up with an instant, what additional questions or parameters would you want to look at?

DR. BASCH: Yeah, I think it's a nuanced question, I'll do my best, and

Dr. Kluetz from the FDA may also have some insights on this.

So in terms of putting together a tool that would give us insights about how people feel with this therapy, you know, I go back to something that I mentioned and Dr. Kluetz did as well, that physical functioning is very important. Now physical -- you know, a lot of people talked about oh, we don't know what tools we can use yet, we have to go back and, you know, start at first principles. That's not the case for physical function, physical function is physical function, right? I mean, I see patients getting all kinds of therapy with all different diseases, and physical function is pretty uniform, there are excellent tools which are already available, some of which are on your list.

You know, the EORTC QLQ-C30 has very good physical function, PROMIS has very good physical functioning. I think those are ready

now and in an assessment I would absolutely include them, number one.

Number two, I would measure, I would let patients self-report their own side

effects. We know that this is, you know, it's not that it's underreported, it's just that we miss a lot of stuff and we misattribute. Patients know better than we do as investigators, so I'd absolutely include symptomatic adverse events. And to figure out what adverse events are important in a given trial, that's really dependent on the products that are being tested and what's known about those products, and hypothesizing over time as we accumulate experience, you know, we start to know, okay, which ones should we ask, and those gets loaded into a form. So now you've got a form that's got physical function and a bunch of side effects, right?

And then the third, I think, which is more challenging, and Paul Kluetz can comment on this, is disease-related symptoms. I think that's a little more challenging in this

context but that could be considered, I'm going to put that aside for a moment.

And then the final piece is overall quality of life, and that includes some of the domains we talked about, you know, emotional or social functioning, and we already know that

and that stuff is generic too, that crosscuts diseases. And so I think you could put together a tool, you and I could do it on the back of a piece of paper like after the meeting, we could just, you know, put down those domains and those actually would probably be pretty reasonable as a start from where we are today, okay?

Now that said, I think it would be useful to take a step back and go to the population and really talk to people to see what symptoms and things are really an issue to them, and then we could go to Version 2.0. But you know, I think we are ready now to measure things that are meaningful to people and most likely will detect signal.

DR. ROSS: Dr. Shah? Oh, I'm sorry.

DR. SIDANA: Surbhi Sidana from Mayo Clinic. Just as a comment to that, you know, we are also using PRO-CTCAE, but the challenges, there are 78 questions, and I had to, based on my clinical judgment, pick which 20 of them. Now my colleagues who are using PRO-CTCAE may pick another 20. And right now my patients are filling out a questionnaire

which is taking them 45 minutes for 20 questions.

Are they all of the right questions? I think that is where the prelim data comes in, like which questions exactly, and we'll know which questions are changing over time, talk to patients who had CAR T, okay, what was important to you, what symptoms did they have, what is important, so I think we need that data.

And I think one thing which none of us talked about is a lot of these people get neurotoxicity, like about up to a third can get that, we are testing questions for cognitive function, did they recover cognitive function?

If they had neurotoxicity, did they still have cognitive impairment at six months, 12 months, I think that's important to address because it may be subtle and we need to pick it up. Thank you.

DR. CUYJET: Let me ask one other question before you go. I haven't heard anything -- you mentioned that people have to come to certain centers because not everybody's providing CAR T therapy. So if you're talking

a Medicare population on a fixed income, what about ancillary considerations? How do people factor in financial burdens, ancillary costs in terms of their decision and how they're making decisions to commit to a new therapy where the outcomes may or may not be desirable? There are considerable side effects to take into consideration, and there's some economic considerations that may impact the family members or the members themselves. Is that part of the assessment?

DR. SIDANA: That's not part of our assessment for this study. We are doing

another study where we are looking at people enrolling in trials or not, and a lot of people don't enroll in trials because coming back and forth to a center is more money, it takes time, somebody has to take time off from work. But I think it's an important question to ask. We are not collecting that information right now but it is important, especially if you're going to mandate someone collect questionnaires or come back for follow-up to a referral center, but who is paying for that, you know, who's paying for the caregiver to take time off. I

think those are challenges and I think they need to be addressed.

DR. ROSS: Dr. Kluetz, were you going to stand up?

DR. KLUETZ: Hi, this is Paul Kluetz from the FDA, and I just wanted to address a couple things. The first was a little bit about labels, you know, FDA labels versus European labels and what's the threshold for data regarding those two different ways of communicating. Europe definitely has a

different threshold for what to put in their labels and how to put it in, they have different regulations, et cetera.

For our labels, especially if you're making a claim of treatment benefit saying our drug reduces pain, our drug improves health-related quality of life, it needs to be statistically tested, substantial evidence, and that's not frequently done, they're typically not incorporated in the statistical hierarchy and tested in that fashion. But we have many examples of using descriptive PRO data in labels to further describe a therapy, and so I was just jotting down some of the more recent

examples.

For safety, which I think kind of is interesting in this context, crizotinib, which is a really important lung cancer therapy, was known to cause ocular toxicities through normal clinician report, and ocular toxicity is somewhat unusual, so they wanted to get a little more information about how that was actually affecting patients so they did

incorporate a patient-reported outcome specific to that and in the label it notes that yes, there was a lot of ocular toxicity, but patients did not feel that they were bothered by it, and there were several other facets of it that were from the patient that really gave, I think, a lot more information about that toxicity.

There's several efficacy examples and one where we added, did actually have a lot of flexibility in what we would normally accept, would be the Hemlibra label as far as improvements in function and joint pain, and that was, it was statistically tested but the instrument had some flaws, so we do put this data in labels.

I would say on the other question, which is what should we do if we could tailor something right now, I agree with Dr. Basch, I think physical function is a very, as I mentioned, disease-agnostic type of measure that's going to be pretty applicable. There is some finesse in there because you do want to

make sure you have, you're where you need to be in your scale because if your baseline function is very high, like in the female adjuvant breast cancer trial where you have young women that are actually functioning very well, you might want to add a couple higher functioning items on there to capture that level. So there's some finessing, but I think physical function is important.

I do think wearable devices in addition to PRO in that physical function domain is going to probably be something that's going to be very valuable in the future as well.

Then finally for the value of symptomatic adverse event reporting by patients, one of the things that we're looking at that I think is going to help, especially in

single-armed clinical trials, is we have a very hard time understanding what's actually disease and what's actually treatment-related side effects. And the way FDA does it currently is we don't look at the attribution that the

physician gives to the AE, we just assume it's due to the drug because we don't really know how else to do it. So you'll see in phase one trials and early accelerated approvals like 80 percent fatigue, very high levels of fatigue which, you know, is probably, some was there at baseline. What you will do with these PROs is that you will get a baseline measure, and then it will be systematically assessed, and so you can take baseline into consideration. We're looking at ways to say we're not going to call it a drug-related adverse event unless it goes above what it was at baseline, and I think Amy Ludek from Mayo has done some work in that, so we're exploring that, we think that could be valuable to sort of cut through some of the fog that we see in these single-arm trials where you really want to talk to your patient about what they might experience. You know, it looks relatively significant if there's high levels

of symptomatic side effects that may or may not be attributed to the drug.

MR. FRANKEL: Can I just follow-up on

that point? One of the things you mentioned on a slide, you categorized besides the patient-reported outcomes, you had, I think you called it observational reported outcomes, and you noted that that may be both from the caregiver, for example?

DR. KLUETZ: Yeah.

MR. FRANKEL: Do you really view that as being two separate measurements? Because I imagine, certainly with a pediatric population, and we're discussing an elderly population that is very ill and is undergoing this therapy. They're typically going to be accompanied by a caregiver, loved one, their spouse perhaps, who will be able to provide insight for a PRO that they may not be able to do on their own, so it would seem to be inherently part of a patient-reported outcome rather than a separate category. Am I correct with that?

DR. KLUETZ: Yes. It's a subtle point. I think what you might be referring to is what we call proxy reporting, where it's

someone other than the patient filling in the

same questionnaire that the patient was supposed to fill in. We don't actually, FDA is not a fan of that, our outcomes assessment staff doesn't like that. Rather, for infants or those who are faced with a brain tumor or major dementia that is unlikely that they're going to be able to fill out the form themselves, they would look for observable signs that the care provider can record. And that's a little different because you don't get that non-observable nausea type of pain thing that you can actually observe. So in those cases you get diarrhea, you get activity levels for kids, and so that's kind of the way we look at it, observational-reported outcomes need to be observable signs.

MR. FRANKEL: And how do you tease out things like financial toxicity as it's phrased, or general anxiety because they're grappling with a serious illness, versus that being specific to the therapy involved?

DR. KLUETZ: Yeah. I tried to make it clear that there's no perfect way to tease that out completely. Symptoms are probably the

closest to the drug effect, as I said, and even within symptoms, teasing out whether it's a drug-related symptom or a disease symptom, or even a symptom of a comorbidity is unclear. Now that one thing that we tend to do is to hold PRO to a higher standard than we do any other clinical trial measure. We know that CTCAE also suffers from the same challenge, so yes, I think teasing that out is a challenge.

MR. FRANKEL: How much do you think that biases the actual measurement?

DR. KLUETZ: Which part of the bias?

MR. FRANKEL: Well, in the sense that there can be an increased, let's say whatever they're specifically measuring, let's say anxiety, and you can say whether it's related to the drug. Do you use a baseline comparative to other patient populations to be able to say well, this is something that we see consistently with other therapies in patients who are undergoing therapies for serious illness, and we can actually deduct that from our overall evaluation, this is actually set aside from that benchmark?

DR. KLUETZ: It's one of the reasons

why we don't typically label things like anxiety in a cancer trial. It may be obviously where, you know, anxiety is the actual disease, but there's so many non-drug influences to anxiety, sleep, for instance, because there are so many nondrug influences. Financial toxicity we don't look at at all, because drugs aren't even being paid for in the clinical trial. So some of those concepts that you're referring to are used a lot in NIH trials or in postmarketing trials to understand the patient experience once the drugs are marketed, but for our premarket, those we look at a little bit less, and focus more on the disease treatment-related symptoms.

DR. ROSS: Dr. Gottschalk?

DR. GOTTSCHALK: I have one question. Right now we're looking in the CD-19 space, we're probably going to measure a lot of outcomes which are confounded by the treatments where the patients have already been treated, so what is the value of getting PROs in the setting right now when we will hopefully move these therapies more in the outcome setting,

more than an autologous transplant or lymphoma,

or instead of an allotransplant for children. And so I was wondering, you know, Dr. Basch or Dr. Kluetz, how do you adjust for that?

DR. BASCH: Well, I would just say in response to your question, and also your prior question, that --

MS. JENSEN: Can you identify yourself?

DR. BASCH: I'm sorry, Ethan Basch, sorry. You know, many of these PRO tools have been evaluated in populations with advanced disease who are highly symptomatic, heavily pretreated, with multiple comorbidities, and have been able to delineate very clearly between arms when there's, you know, when there's no real effect there. And so there are many examples of, despite the challenges that you allude to, where these tools perform extremely well, and that's because some therapies really improve the way people feel and some therapies really worsen the way that people feel and you know, many therapies do a

little bit of both in different ways, and these tools are able to detect that. So I would argue that in an advanced population or in a

heavily pretreated population, it's perfectly appropriate to use these tools. In fact, those are the settings in which these tools are most commonly used.

Now that said, I think yes, you might get a crisper signal in an adjuvant setting or in a healthy population as you move therapies, you know, more up front, but I don't think that that's a reason not to use it later on. In addition, you know, you can collect a lot of information that's hypothesis generating for earlier.

DR. GOTTSCHALK: I think that was not my question. The question was, you know, side effect profile will be probably different. You know, for example, giving therapy in a patient who has a history, there is probably more expansion, more neurotoxicity, et cetera. So then if you have a very validated PRO set of data but you haven't measured every

pretreatment therapy with CAR T, and so then the question is how does this data look like when the patients are not so heavily pretreated?

DR. BASCH: Do you want to take that?

All right.

DR. GO: Will Go from Kite. So, I totally agree with you. I mean, this is where the, I think a challenge that we're all facing across industry as well as our academic partners and patient standpoint, you're exactly right. Let's just take CD-19 as an example, right? In our trial, in the pivotal trial and effectively third, fourth, fifth-line patients, two-thirds of them already had B-cell aplasia because they had so much prior rituximab. And as we are, you know, continuing to look at the B-cell aplasia, which is one of the long-term questionable side effects, about what that means for patients, how is that going to go over time? You're exactly right.

As we get to earlier lines of therapy, potentially we might see fitter T-cells, fitter

patients, and that's why, again, I defer to ZUMA-7, because why? That's a second line therapy with a randomized controlled trial where we are going to be looking at that with some classic PRO measurements.

DR. ROSS: Can I -- I wanted to ask a question, and I think Dr. Shah is one of the

people who actually raised their hands. So, we've heard a bit about how, you know, this therapy is so effective, kind of like why do we need PROs. We also heard among the comments from the panel that the PRO should only be used as part of randomized controlled trials. I was hoping that some of the clinicians who've used PROs in practice, not research, could talk to some of the, not just the challenges which we heard more about, but the successes of how they've been used and how they've informed clinical decision-making.

DR. SHAH: Gunjan Shah from Memorial Sloan Kettering. So, I think that while I can fully understand your questions of sort of timing and duration of looking at these PROs,

that specifically to what we can do with them even now is, we expect even if we continue to use them in these later line settings with several lines of therapy, that there will be several iterations of these CAR T-cells, and we expect that future ones will be better than the ones now.

And one of the things that we've been doing with the autologous transplant as part of

looking at all of this is, essentially you have a therapy that's safe enough that what you're really researching is how to decrease the symptom burden and how are you actually making a difference, that these are your primary outcomes, you know, it's safe to give, it's effective, we know that this works, but how do you make it better for the patients, how do you make them not need to be in the hospital or not be in the ICU, that kind of thing.

So some of these measures are really for that, and so I think that partly to answer your question, having these at the baseline of sort of the first generations of these drugs

being used commercially and, you know, on trials, it's helpful to then sort of inform the studies of the future.

In the autologous transplant setting, you know, one of the studies and one of the only studies that's really shown to make a difference has been an acupuncture study that we did with our integrated medicine colleagues at MSK, and were able to show a difference in their patient-reported outcomes as a primary, of decreasing fatigue and changing their

symptom burden, and so I think that having this information is valuable over time.

DR. KLUETZ: May I?

DR. ROSS: Yes.

DR. KLUETZ: Paul Kluetz. Just one comment about late stage versus early stage. You know, most of our single-armed trials are multiply refractory, our dose finding trials particularly, and there's actually been some interest in using sort of side effect bother and side effect PRO to help better find dose, so that's one possible, actually a pretty good

utility for that.

And I'd also argue that it's still important to measure safety and it's very important to measure safety in that setting. For instance, we know that in second and third line multiply chemo-treated patients, we're going to see a lot more neutropenia with another cytotoxic agent. And so I think we'll see, it's important to understand that toxicity profile and I think, I look at it as complementary to how we're looking at safety as well.

With things like health-related

quality of life and physical function, I may have to agree with you that maybe that's not the right spot for those more broad net benefit kinds of questions, but for safety, I think it's actually a pretty important use.

DR. BASCH: Ethan Basch. I'll just comment briefly on the real world use of PROs. So, our group and others have done many registries. We currently have a large national U.S. trial, or study I should say, real world

study supported by PCORI, in which patients receiving systemic cancer treatment for advanced disease at 50 community practices around the U.S. are self-reporting their own patient-reported outcomes on a weekly basis throughout their entire treatment trajectory. The compliance rate is 96 percent, meaning that if you look at the average proportion of patients who self-report every, at any given week, it's 96 percent. 80 percent of those are self-reporting on their own, and the additional 15 or 16 percent, they actually get recovered by somebody calling them if they don't self-report, so it's augmented by having a central person in addition to collect the

information.

I would also mention, there's been some questions about informative missingness when patients are hospitalized or have severe toxicities, and in those settings we do use proxy reporting, so we will use a caregiver or clinician who will provide the information and that's generally used in sensitivity analyses,

so that we understand the reason for the missingness, but again the missingness is extremely low, and these are patients with advanced disease, often close to death.

DR. ROSS: Dr. Perissinotto, and then Dr. Goss.

DR. PERISSINOTTO: So, one, I appreciate Dr. Sidana for mentioning the potential cognitive side effects that happen to be particularly important to our Medicare beneficiaries. So my question is for Dr. Go and any of the panel members in terms of the trials with the reported neurotoxicities if we know the extent of the variability of the toxicities, if there is any cognitive assessments that were done at baseline or the follow-up, and what the long-term sequelae are.

DR. GO: Will Go from Kite. I'll comment first and then I'm going to ask our FDA colleague to comment as well. I think it's very challenging in terms of neurocognitive behavioral testing. What we did in ZUMA-1, the pivotal trial, we incorporated a mini-mental

status exam, which is not obviously a great office tool. We chose that because in previous FDA-approved products like blinatumomab from Amgen, they also used it as well, so that is what I would say is a very blunt tool to look at that. Obviously, we are exploring possibilities of other more complex neurocognitive testing, but this, I agree with everyone here that as CAR T's go to other disease states, different lines of therapies, that this will be something that I think we would want to as a community to continue to support, and we at Kite Gilead will definitely keep supporting it.

DR. KLUETZ: Paul Kluetz with the FDA, and I think it's an excellent question because I think it's, I like these targeted questions that are getting at things that we know that are happening, can we further describe and

characterize the effect. Cognitive testing using a, is a clinical outcome. Patient-reported outcomes are obviously challenged. If you're cognitively impaired,

filling things out can be challenging, although there are some cognitive scales.

There is interest in, again, looking at technology, so are there different types of gaming types of situations where you have certain kinds of, almost a performance outcome where you're filling in certain things on an iPad, and there are some interesting things that are coming out with that, but they're, we haven't seen that arrive at the Agency.

DR. PERISSINOTTO: Thank you.

(Pause.)

DR. ROSS: Dr. Goss, and then Dr. Lamon.

DR. GOSS: I had a couple of questions. Dr. Basch, I appreciated your presentation because it was really very helpful. There were a couple of other -- there was a question that I just wanted to clarify. The way our question is asked, it's not asked specifically about CAR T at this point, it's

just PRO, and in one of your conclusions you made comment about the utility for CAR T, and I

just wanted to make sure that I'm understanding the question correctly, number one, and number two, to know if that would change how you're thinking about the issue of PROs if it were specific to CAR T.

And I also had a question about, kind of pragmatic, so our question two has to do with, you know, transferable to community practice and, you know, quick throughput to a trial setting, and I was trying to go through the data that I had available. With the exception of the presentation on the FACT, which wasn't one of the measures we're looking at, in none of them did anyone report what was a minimally important clinical difference. And so I would be interested in our general assessment of the experts out there about in which of these measures do we have kind of a defined clinically important difference that we could use as a benchmark.

And also, there was some lack of information about the cost of licensing, for example. So, EORTC I think has a licensing

arrangement, you know, and as mentioned, it's a strongly validated measure, I would agree, but I'm just curious if anybody has any details on those types of practical implementation limitations, because I think that may be relevant to how we think about this.

DR. BASCH: We did --

DR. ROSS: Dr. Basch, please --

DR. BASCH: I'm sorry, my apologies. Ethan Basch from University of North Carolina. Yeah, so we did report on which tools were used in CAR T trials really just as a matter of information, but the basis for particular use in community practice or how widely we use the tools for generalizability came from use in the Medicare-aged population, and I did show that as a separate item for each individual tool, and that was the basis of that, not the use in CAR T.

DR. GOSS: Okay. Any thoughts on the minimally important clinical differences, and whether or not there are any of them that have really well-established guidelines or some that you feel that may be missing as well?

DR. BASCH: Well, I and some others

can comment on this as well. So, you know, in FDA lingo, this has been sort of changed to view a score that represents a meaningful change, so for all of the tools that we gave a smiley face to, there have been evaluations of what is a clinically meaningful score change, with the caveat that the PRO-CTCAE is, you know, about adverse event reporting that's generally descriptive rather than, you know, comparison of proportions, hitting a certain score threshold.

DR. ROSS: All right, so I know there are a number of questions here. Dr. Lamon was next, and let's just try to keep the questions as short as we can so we have enough time.

DR. LAMON: I have a question for Dr. Snyder and anyone else who wants to answer. I really liked the graphic presentations you did on the issues of getting clinician engagement, but I'm thinking about all the technological issues, and my impression is that the ability to do the PRO measurements is technology and that we have more information systems. How are you getting the information on those graphs, are they in real time, and

what's the interface with the electronic record that you're using at Hopkins, or any other records if anyone else wants to comment? I think that's limiting clinician involvement and putting a wedge between collecting data and using it, and do we have it in real time to use it in real time?

DR. SNYDER: Claire Snyder from Johns Hopkins, thank you for the question. For the purposes of our research we made up the data so it was really easy to get.

(Laughter.)

However, the rationale behind the research was work that our group had done at Johns Hopkins and my colleague Michael Brundage had done in terms of clinical trial data where we wanted to show the data to patients and clinicians and we didn't know the best way to convey all the information we wanted to, how is the patient doing over time, what's an important difference, what is statistically significant, what does the doctor need to pay attention to? They're not going to learn all

about these questionnaires, we need to make them immediately interpretable and intuitive.

So, the reason that we had to do the research that we did is that there is a huge increase in the collection and use of these data in clinical practice, so our team at Hopkins started doing this in 2005. I would say we were some of the pioneers in the U.S., I feel like we are now almost obsolete, but the work done by Ethan Basch and others has moved this so far forward where he is, for example, doing this study in 50 community practices.

A colleague of ours, Roxanne Jensen, who's now at the National Cancer Institute, did a review of e-PRO systems in 2014 and even then in cancer care alone, there were 33 unique systems meant for clinical practice. The big challenge now is getting the data in the electronic health record. With funding from PCORI, a group of us, including some folks here, developed a users guide for how to integrate patient-reported outcomes into electronic health records. It is freely

available on the PCORI website and it walks step by step through all the considerations involved. It does not provide one right answer but a range of options and their relative

advantages and disadvantages. So I think increasingly, there are tools that are going to get us there. Thank you for the question.

DR. ROSS: Dr. Shah, do you have a quick response?

DR. SHAH: Yes, just very quickly, Gunjan Shah from Memorial Sloan Kettering. So, I briefly was able to show you some of the figures from our MSK Engage platform that's being created and sort of in use on the surgery side and being transferred into a more long-term use for the transplant and cell therapy side. And you know what, the way it's working right now and what we're hoping to continue is that you can actually pull it up in the office, that you can pull up an individual patient and show that patient, here's what you've reported over time, and with one click you can actually decide to include that in

their electronic record, and so that it can be, you know, part of their record over time, but also pulled up in sort of a dynamic fashion to intervene on if you so choose to, but also see, you know, which things are higher at which visit, which ones are worse today, which are

better today, and look over time.

We on the clinician side can then also say here's your entire panel of patients with the same disease, or answered the same survey, and then have more aggregate data also built in to be able to look at.

And so I think it's kind of important to be both ways, sort of aggregated across the population, but also to include the patient in showing them what they reported along the way also.

DR. ROSS: Thank you. Dr. James, you've had your hand up the longest.

DR. JAMES: All my questions have been answered by the last two.

DR. ROSS: Great. Dr. Feinglass?

DR. FEINGLASS: For our FDA colleague,

Dr. Kluetz, how often does the result from a PRO assessment tool become a deciding factor for a binding FDA decision?

DR. KLUETZ: Thank you for that softball, this is Paul Kluetz.

DR. FEINGLASS: You're welcome.

DR. KLUETZ: Paul Kluetz from the FDA. So, I think it's a really important question,

it's something I talked about over lunch and that is, are we using patient-reported outcomes to further characterize how a therapy affects the patient in the totality of data, and then we organize that in a qualitative or a quantitative risk-benefit determination, which is what we do at FDA, mostly qualitative right now, yes, we do that all the time.

We wrote a recent New England Journal of Medicine article on the use of metastasis-free survival, which is a new endpoint for nonmetastatic castration resistant prostate cancer so it was a novel endpoint, and in this particular case patients normally don't get a therapy and they're usually asymptomatic,

and so it was like sort of a maintenance therapy question so we were really quite concerned about the tolerability, this was an important part of our decision, because we knew that the benefit was there, that it was pushing back metastatic disease, but how tolerable was it? And so in that case we did use, looked very carefully at this overall side effect bother question and different side effects, and made sure there was no significant signal there

in addition to the normal CTCAE data, and so that weighed in.

I think the bigger question is, have we ever used it for a negative nonbinding decision, and I think that's obviously what everyone is really concerned about, and that's not to my knowledge. We've used it for positive, important positive decisions. For instance, Jakafi, as I said, it was a key secondary endpoint that moved the regulatory decision from an accelerated approval because it was a surrogate endpoint as a primary endpoint, to a regular approval because the

secondary endpoint was a symptom improvement, a clinical benefit that was meaningful to patients.

DR. ROSS: Dr. Civic, I think you were next.

DR. CIVIC: Yeah. One of the questions we're asked is how long to measure, sorry, a PRO, to be able to identify a valid treatment effect and, you know, we're looking at late toxicity but also, I think it was Dr. Abikoff talked about late benefits, that there wasn't a response until, in some patients

until nine to 12 months, which makes it seem like we should be measuring PROs for at least 12 months. Does anyone want to comment?

DR. GO: Will Go from Kite. So yeah, similar to other trials and in our pivotal trials, number one, we've actually seen that with a single dose of CAR T, as well as at the NCI, and we'll hear Dr. Yang comment as well, that we've seen conversions from stable disease to PR to complete remission as late as over 12 months, and this is why -- and without any

other intervening therapy. And so this is why, and again, I am not a PRO expert, I'm a hematology oncologist, but if I were to design the PROs, again, that's where the challenge lies, because you're going to start seeing potentially late converters as far as 12 to 15 months.

DR. ABIKOFF: Cori Abikoff from Novartis. I agree, it was my point that we do see these patients progress over time and that is one of the things that differentiates CAR T therapy from other therapies, and I also am not an expert in PROs, but I think that this along with the questions that have been raised about

things like neurologic toxicity, these are still fairly young technologies and they've been studied for a fairly short period of time, so understanding what those late effects are and how that impacts PRO measurement as well as understanding the immediate effects and how that affects PRO regimen, are still things that we're trying to understand, and why we are actively utilizing them in our current and

future clinical trials, because they will help us to answer those questions.

DR. BAR: Merav Bar from the Fred Hutch. Regarding the long-term follow-up for PROs, I think there is two sides of it. One is the one that patients might respond later but on the other hand, there is still relapsed disease or progression of disease after and a lot of patients that we are looking at receive subsequent therapies that may also affect how they feel, their quality of life, and symptoms. So there are two groups of patients that, one may respond later, but on the other hand there still are patients who will have progressive disease and relapse after, either because of interim therapy, they have symptoms of disease

progression or because of subsequent therapies, so these two things need to be taken into consideration as well.

DR. ROSS: Okay. Dr. Garrido, I think you had your hand up next.

DR. GARRIDO: So, from Dr. Snyder's presentation, we saw that individuals,

including clinicians and researchers with quite substantial education aren't so great at reading graphs and interpreting changes in PROs. So I'm wondering, either in your own personal experience in working with patients with limited literacy or education, are people able to understand just the questions themselves, not even the changes, or have these been evaluated in people of limited literacy or education?

DR. SIDANA: Surbhi Sidana, Mayo Clinic. While I don't have the exact answer you are asking, you know, I had a patient who was filling out a similar questionnaire in our study. He did not have neurotoxicity but his heart rate was fast, but he had not slept because of all the alarms going off in the ICU, and that patient had to read a question three

times on that questionnaire to understand. Now I don't know what to do with that answer, do I even trust the answers the patient gave? So yes, I mean, those are challenges, not only of patients understanding questions, but even

well-educated patients who are having side effects of treatment, you know, being able to answer them in the state that they're in.

The one more point I would like to make from before is, I think it's important to study late effects because as you know, for allogeneic transplant, we found out, you know, there are late effects like chronic graft versus host disease that impact quality of life. Now we don't know any about CAR T yet, but who knows what's going to happen when these people are like three years out, four years out? So I think it's important to study them, we just don't know what they are right now.

DR. PERISSINOTTO: Can I just add to the question about low literacy also? Because I think you'll be able to answer this if some of the PRO measures have looked at multilingual and multiethnic populations.

DR. BASCH: Yeah, absolutely, so --

thank you, Dr. Snyder. I'm Dr. Basch, Ethan Basch, and yeah, I need to get like a sticker on me or something to me as a reminder, which

speaks well to your question, right, I need to be prompted.

So, a couple things. First, you know, in looking at Claire's evidence, which I think is, you know, terrific studies about interpretation of the graphic, you know, we haven't applied that level of scrutiny to clinicians, for example, in interpreting waterfall plots or Kaplan-Meier curves, or all the different graphics that we are expected to interpret in journal articles or in drug labels, right? So I mean, people have trouble digesting data. You know, I told Claire that personally I like the USA Today, I like a simple graphic, like I can get that, so I think there's something to simplicity in understanding graphical displays. But I think that, you know, as Paul alluded to, we sometimes apply a greater level of scrutiny to these patient measures than we do to the metrics that we all take for granted every day, and I just want to caution us not to be, not to

apply a higher level of scrutiny.

Regarding your question, so there have been many many PRO studies done in patients with low education levels, low health literacy levels. In a study that my group conducted that was reported last year at ASCO and in JAMA, we had a very large arm of patients who had never used a computer before and they were using a computer and they, that population had low literacy and almost universally had less than high school education, and they were universally almost able to self-report, and actually that group saw greater benefits from reporting PROs and having information conveyed to the clinicians for management of symptomatology.

So I mean, as far as language, there have also been many studies done in groups speaking other languages. I'd say all of the tools with the smiley faces have been linguistically adapted into other languages using a pretty, I'd say a pretty rigorous translation process that often involves both cognitive interviews of people and if done well, includes people with different levels of

literacy and education as well, so I think for the good tools, it's generally pretty good.

MR. FRANKEL: A quick follow-up to that. Do you regularly, I assume this may have come up when you evaluate these tools, to ask the patient how burdensome they find the tool that they're answering? So, is that every single tool you have that question and you have the data from there to be able to say well, this tool, we have a very negative response and this one -- and I assume that would be true for, as the patient progresses through treatment they may have different responses to that as time goes on, and what do you see with those terms?

DR. BASCH: So, I'm sorry, maybe you can restate that; what it the thing you're interested in knowing?

MR. FRANKEL: The patients' feedback of how burdensome they find the tool that you're actually using to measure their feedback.

DR. BASCH: Yeah. So we've done a lot of that, others have, I think Claire has too, so we've done a lot of work with how burdensome

people find questionnaires. You know, there are a few people who find these questionnaires to be burdensome, but just like they find going to get their CAT scan burdensome, and their liver biopsy burdensome, you know, not that a PRO instrument is similar to a liver biopsy, but part of the things people do as a part of trials or care is burdensome, but may have value.

The vast majority of patients are very enthusiastic. In multiple surveys that we've done, on average, about 94 percent of people say they'd recommend doing this to others, they'd do it again, they find it highly valuable, it improves communication with the care team, they feel that they're an active participant in care, an active participant in the clinical trial enterprise, and people feel engaged, people like doing this. I'd say that in some of the settings where we do studies where we ask people the same questions week after week after week, you know, there are people who push back, like couldn't you come up with a few new questions or like, you know, I

25 already told you I don't have fatigue, why do

1 you keep asking me about fatigue? And this is
2 where we're starting to use technologies to try
3 to make things a little more user friendly, but
4 in general people don't find these things
5 burdensome at all, in fact quite the opposite.
6 You know, most people are delighted to be, you
7 know, a part of what we're doing.
8 DR. ROSS: Dr. Flynn.
9 DR. FLYNN: Yes, Kathryn Flynn from
10 Medical College of Wisconsin and CIBMTR chair.
11 Just one additional point. I can't speak for
12 all of the measures, all seven measures, but
13 certainly for the PROMIS measures, one of the
14 stated goals in developing those was to
15 evaluate every single item in people with low
16 literacy, so every item at a minimum had at
17 least two people with less than a ninth-grade
18 reading level evaluate the item through a
19 cognitive interview, I think the PRO-CTCAE also
20 had cognitive interviews specifically targeted
21 to people with low literacy, so for those
22 meticulously developed measures, I think you

can have confidence that most people will understand them.

With those modular approaches, of

course, that's where, you know, taking into consideration how many different domains, how many different questions you're choosing, and testing that again to make sure in that particular patient population, you're not asking something that people can't complete.

But then another question you had asked earlier about licensing fees, also, both PROMIS and PRO-CTCAE do not have licensing fees associated with them, so that's not a burden.

DR. CHUNG: Hi, Karen Chung from Juno Celgene. Just addressing, again, the literacy levels in most of these instruments, the four of the seven that would, you know, move forward, they are built to be at a fifth grade, you know, kind of education level, so hopefully, you know, we're trying to take care of the literacy by making sure that the language is really understandable.

With regard to understanding the

outcomes, you know, some of the analyses we really try to do so it's understandable to clinicians as well as patients include responder analyses so they know, well, this is the proportion of the patients in the clinical

trial who had a clinically meaningful improvement or, you know, worsening, or stabilized. So those are the kind of metrics we feel, you know, help them really understand the outcomes more than kind of what is the mean change from baseline, you know, and the other kind of, you know, modeling that we do on the PRO data.

So it's all trying to be, you know, very concrete in the level of change and filling out the difference between responder or minimally important difference, and a lot of people have done different analyses around that. You know, there's anchor-based, distribution-based, and for the EORTC-QLQ-C30 we felt very comfortable using that because there have been solid MID research done out there by (inaudible) and so that's what we're

using to identify our responders.

DR. ROSS: Dr. Cheng, you had a question earlier?

DR. CHENG: Yes. Go ahead.

DR. FERRUSI: Sorry for the delay. I saw a nice lineup of people and I thought I would wait to see what they had to say.

DR. ROSS: Just introduce yourself.

DR. FERRUSI: My name is Ilia Ferrusi and I'm from Novartis.

A lot of good points have been covered here. Standard practice when developing instruments is to develop them at no more than eighth-grade reading level, and I did want to address one component, whether all of the items are relevant, I can't remember who asked the question, but for instruments that are developed as standalone instruments, so I'm not talking about something like an item bank where you pick and choose, but something like the FACT-G for example has been developed, and has domains within it.

When cognitive debrief is done, so a

first draft of the instrument has been developed and the cognitive debrief is taking and sitting down with a patient in that population, that's a really important part. You're talking to real patients who have the disease condition of interest, and you ask them to work through the items and tell them how they're interpreting this, how they understand the response options. You also would go

through a practice of asking is this relevant to you, do you feel that any of these items are repetitive, and that's a very purpose-driven process that we go through to ensure that we're not asking too many questions and the fit is just right.

So some instruments like, the instruments that, Dr. Basch has actually summarized their development, and he talked about content validity, if you saw a smiley face or checkmark next to content validity, that's some of what he was talking about.

DR. ROSS: Thank you. Dr. Cheng.

DR. CHENG: Joe Cheng. I just, I

still need some clarification as far as what the concerns are about collecting patient-reported outcomes, and I guess my question really is, there seems to be a lot of concern about using PROs in following how patients do. Do you have another suggestion, then, for collecting quality added life years, or how do you really assess things like minimally clinically important difference, and then really, how do you risk adjust without collecting this data, the results of your

patients? And then how do you then coordinate whether this is related to an episode of care versus fixed time points?

And I guess that's what I'm saying, because all the concerns about PROs seem applicable through all of medicine, whether it's a stroke, or spine, or any tertiary center would seem to have the same concerns that you have about follow-up patient care. I'm just still trying to figure out how does this apply directly to CAR T, and are you saying that we shouldn't be collecting any of these PROs for

anything we do, or quality added life years are not as important? I guess I just want some clarification on that.

DR. SIDANA: Surbhi Sidana, Mayo Clinic. I think it's very important to collect these data, that's why we are doing them. I think what's not clear is exactly which ones. Again, we don't want to burden our patients too much but we also want to get the answers right, what is important to collect and then how frequently do we need to collect it? Do we collect it every week for one year, do we collect it every month for two years, like when

are we seeing the changes? I think that's the finesse we need to get right, but it's very important to collect.

And I think the third part no one really talks about is who's going to pay for it, because right now I'm doing a study that has only 30 patients we need to collect. It takes one patient one hour per questionnaire, each patient will fill out seven or eight questionnaires, so that's a lot of time for the

coordinator. And once that patient goes home, someone has to call that patient up, or if they're filling it electronically and they don't answer, someone will be asking that question over the phone to ensure completeness. And if they've gone away from my practice and now they're seeing a local clinician and if there's a symptom, even if I see it, what do I do? Say they say they're having severe pain on that question. Now I'm not following them on an everyday basis, so that creates an ethical dilemma as a clinician, I don't know what the right answer is, but I think it's very very important to collect them, but in some way as a community, and we're already talking about

forming a working group. How do we answer these questions, like what do we do about the data we get, and who pays for it, and how do we collect it in a standardized manner so that we are collecting things that are important.

DR. ROSS: Just in interests of time, try to keep your answers moving along. There's a long line.

DR. CHUNG: Karen Chung, Juno Celgene. I completely agree that patient-reported outcomes are important and I think it's important to assess them in kind of a systematic way, and so that's why in clinical trials, you know, we have very good kind of follow-up to all these rigorous schedule of assessments. If they go off study, we have one last assessment. I think the concern is really if we had it in the real world that would be great, but I don't think the infrastructure is there. I don't think there's, you know, a way of getting the data systematically and cleanly. I mean, we have learned from a lot of trial and error in clinical trials a lot of issues with data, you know, getting the data collection right. And so I think to, you know, have the

general practices pulling this data together in meaningful ways so that we can use it is still, we're a little bit far away, you know, with regard to that and all the other issues with regard to instrument selection and analysis, and all the logistics around it.

DR. GO: I just want to give a clinical perspective as a former transplanter, as a former allogenic stem cell transplanter. CIBMTR has been obviously the biggest group that has been for all, mandated by law. That took them almost 20 to 30 years before we could understand GVHD scoring, and so I think if it takes 20 or 30 years to even get GVHD scoring right, our opinion is it's going to take a long time to really get PROs right, and this is why from Kite Gilead, we don't believe that right now it's warranted in terms of coverage analysis.

DR. BASCH: Ethan Basch, University of North Carolina. Thank you.

I really, I have to say I came here today, I was very very surprised, as you might be, to hear the reticence on behalf of some stakeholders to collect this information that

cannot be gathered in any other way in a population that we are bringing back to the clinic all the time, harvesting from, reinfusing, scanning, et cetera, et cetera. We

are spending a lot of resources on this patient population and to not collect patient-reported outcomes, which is essentially handing somebody a questionnaire, to me frankly seems rather absurd.

There's a many-decade experience administering questionnaires to people in trials and in the real world with very high rates of compliance. There are all different kinds of ways to do it, it can be done on paper, it can be done with a telephone survey system, it can be done with an i-Phone or Android system. This is done all the time. There are hundreds and hundreds and hundreds of registries in oncology patient populations with 90-plus percent compliance rates using electronic devices all over the world now, and to say that feasibility is a barrier to me is simply refuting an enormous amount of accumulated knowledge and ability.

To the 45-minute or hour-long

questionnaire, I mean, that seems very unusual to me. Our questionnaires that we use

repeatedly take between five and ten minutes long, and we often ask people, to your question, did you find the questionnaire burdensome or too long, I mean, it's really never an issue. There's some trials that have longer questionnaires that are spaced out maybe every three months, but again, I mean to me, compared to what we are asking patients to do in order to receive these therapies, this is minuscule, so I don't really see the barriers.

DR. CHENG: Can I ask a follow-up to that question?

DR. ROSS: No. Well, I just wanted to allow her to speak, and Dr. Yang has been waiting for a long time. I want to make sure everyone gets a chance to ask.

DR. FERRUSI: Thank you, Ilia Ferrusi from Novartis. You know, I think many valuable viewpoints have been expressed here. What I would like to add is that PROs generally, yes, are a great thing to measure to understand ultimately how the patient's experience is going. But what, I want to bring us back to

principles and make sure we're focusing on why we're asking for PROs, what is the research question, what is the context in which, because the answer to that question, which measure to use, is going to vary depending on what you want to measure and what the context is.

So in broad strokes, it is hard to answer that question and our position, I would like to clarify, is simply that we are not comfortable with PROs being required as a requirement for coverage or access to a medication.

DR. ROSS: Great. Dr. Yang, do you still want to ask your question?

DR. YANG: This is a question addressing the fact that almost everything we've talked about here today is about capturing acute or on-therapy toxicities, or under-appreciating them. The main difference in my experience with CAR T, especially with CD-19, is it's a one-time treatment, and at the back end patients who are responding or doing well, which is almost half of those patients or more, have a paucity of any interventions or requirements at that point, and are we

capturing that? So do any of the people who have PROs associated with their studies have questions such as how many people have gone back to gainful employment, how much more care have they required in the last year or two, and how often do they think about their disease, how often do they have concern or anxiety about their disease, because this can be a one-time treatment and then a walk away.

DR. GO: Will Go from Kite. So, we are looking exactly into that, Dr. Yang, in terms of the work productivity and activity impairments in Version 2.0 in our randomized Phase III trial. I think that's the biggest thing we're doing, so we are actually looking at that in all of our trials since this was mandated by the FDA for 15-year follow-up, so we are going to get adverse events, look at the B-cell aplasia, the use of IVIG, as well as some of these other PRO and back to work products.

DR. FERRUSI: Ilia Ferrusi from Novartis. To answer your question, no, we are not collecting return to work, but the work productivity, activity impairment questionnaire

is a very good tool for that. I would say that we are using, again, the FACT-Lym, which has physical, social, emotional and role functioning, so as a component of role functioning, we can certainly look at a return to normal activity, and we are continuing to collect that data 12, 18, 24 months after their administration of CAR T in JULIET.

DR. ROSS: Mr. Frankel, you get the last question.

DR. BAR: Sorry. To answer this question about the long-term follow-up, so yes, an effort has been made and is continuing to be made to learn about those long-term effects. Currently we don't have the data, CAR T-cell clinical trials started maybe about five, six years ago so the data we have right now is limited, and I think in the first few years the most excitement was about whether the treatment works or not, what was the response rate, and people paid less attention to more long-term effects and quality of life. However, now when we know that maybe there is approximately a

50-percent response rate and long-term response, so people are paying more attention

to those quality of life questions, and we are planning to follow-up patients at least yearly for 15 years from now according to the FDA requirements, so we are making an effort to learn that, but we still don't have data.

And the thing that I would like to say here is that effort has been done, and we will make even more effort to learn those questions. The question is if we need to make this a mandatory thing when we make the decision whether or not to reimburse patients for such treatment.

MR. FRANKEL: This question is for Dr. Basch and Dr. Kluetz. You advocate for PROs to also be given to patients who were receiving the standard of care until now. So in other words, as a patient, I think that many would be interested to know how are patients faring in terms of their observation of their own outcome when they receive CAR T therapy in a specific instance, and how are the patients

who did not undergo the therapy and have a, let's say three-to-six-month survival on average, how did their feedback look? And that way you could actually compare those two groups

of patients, and I think that that would probably influence many patients much more than if they only saw receiving the therapy and they saw the drawbacks there, let's say, if they were looking at the advantages and disadvantages, and they could actually compare that to the alternative. Because I think without that, the patients are really at a very weak position to really have a fully informed decision.

DR. KLUETZ: Paul Kluetz from the FDA. So I think one of the problems, one of the issues is context which I was talking about a little bit earlier, and that is, is this a single-armed trial or is this a randomized trial. I mean, you won't have that -- comparing to a historic control is obviously going to be very challenging in this field right now given the heterogeneity of the tools

that are used, and assessment frequency and things like that, and so really when you compare it to the standard of care you're talking about a randomized trial much like the one that was actually presented as, I guess, the second-line trial that was presented.

Now you could do that, and in fact that's the majority of what we get at the FDA in oncology, is randomized trials, and they do ask the same questions of both arms, and that does help to give you a comparison of how well they may feel or function on one arm versus the other.

MR. FRANKEL: And how about moving forward? So in other words, does that, for whatever reason they're not eligible, or they opt not to go through CAR T therapy? Maybe they're concerned about certain toxicities involved, but capturing the data from those patients so that the patients in the future who have to decide between the two could have that at their disposal.

DR. KLUETZ: Yeah, that may be outside

of more of a regulatory question but it is an interesting question, and I don't know how you would design that, but it doesn't seem like something you would normally see in the regulatory setting.

I did want to actually add one more point to the point of, have people ever used at the FDA patient-reported outcomes to make a

negative decision? Let's remember that in oncology we have objective tumor-based measures, and survival is our primary efficacy measure, and we always have. In many other therapeutic areas that's not the case, so I don't want to speak for the entire FDA by saying we don't use patient-reported outcomes in a very important way to make key efficacy decisions, because that's actually not true. There are many therapeutic areas where the disease manifestation is only a symptom and that's the only thing to measure, an analgesia being an obvious example, and in those you need to show that patient-reported outcome is improving, or that therapy is not going to show

any efficacy.

DR. ROSS: So at this time --

DR. BASCH: I just want to respond to the question briefly.

DR. ROSS: Please introduce yourself first.

DR. BASCH: Ethan Basch from the University of North Carolina.

So, the most valuable comparative data will be from a prospective randomized

controlled trial, that's one of the reasons why it's really important for, you know, sponsors in their discussions with regulatory authorities, to really think about these outcomes and pick them right at the very beginning, so we can really understand in that context because, you know, we have a little bit more equipoise in that setting.

I think your question really alludes to real settings, to registries and postmarket surveillance, I would guess. You know, I do think there's value in having comparative data after a drug is on the market in order to do

comparisons, especially if that information was not really fully characterized pre-approval, or if there are not long-term outcomes prior to marketing. That said, there are limitations. Obviously there are many dimensions of selectivity, patient and provider selectivity, and so these populations will inherently differ, those who did and didn't get the therapy of interest, in this case CAR T. And so if that was done, then there are methods of balancing those differences in observational data, they just have to be done very well.

DR. ROSS: So, thank you to the presenters again, and speakers, for continuing to answer our questions. So I let us go about ten minutes over, this was obviously a very rich discussion, and many of the panel members had questions.

We're now supposed to transition to the period where we have an open panel discussion. I will just note that we are not precluded from asking the speakers or presenters additional questions, but if you are

asked, I would request that you keep your answers very short. But this is really an opportunity now for the panel to further discuss the area, to think about in anticipation of the voting which is going to be in an hour from now, what further information we need or that we still feel uncertain on. Dr. Goss. Oh, and then -- go ahead.

DR. GOTTSCHALK: I would like to circle back to two things. One of these is duration of follow-up. You know, some have mentioned the FDA mandate of 15 years, but that really comes out of the gene therapy arena to look at the risk of insertional mutagenesis

after the transplantation of genetically modified T-cells, so the question is really, how long should we really follow-up these patients?

And the other question is, or kind of comment is, right now there's no clear proof test to track the commercial products, and I would encourage the companies to develop those because in the PRO assessment if something

comes up, of course we want to know, what is the precursor, are there some measurable CAR T-cells, and that is not right now available outside the research setting, so I think that probably is another key thing you really need to assess the safety involved in the long-term outcome of these cells.

DR. ROSS: Can I just ask, Dr. Gottschalk, are you asking that question to the panel to say clinically, what's the appropriate time?

DR. GOTTSCHALK: What is the appropriate time, how long should we really follow these patients?

DR. CHENG: So basically from what I understand and from what I heard, like

Dr. Abikoff mentioned, that 54 percent of patients went from partial to complete. I assume the symptomatology would also follow the difference between a partial versus complete remission in nine to 12 months, which means it would seem to me you would have to follow at least 12 months in order to get -- and that was

a question that was asked before, so if it's a question about the three choices that are listed there, it would have to be at least 12 months or up to 24, in order to see whether or not the patient symptoms would follow the response rate.

DR. GOSS: Actually I have a contextual question because I mentioned it before, but I was wondering if Tamara could clarify it for us. The way these questions are asked, they're not asked specifically about CAR T, I just want to be sure that's correct. So we're asking about PROs in the Medicare population, and we're asking about some specific measures, and then we're asking about, you know, ability to implement. But nowhere does it say specific conditions and nowhere does it say, you know, specific treatments, so

we might have to think more broadly if we're putting a time frame. I understand for CAR T, you know, six, 12 or 24 months might be appropriate, but for other situations it may be longer, and so it may affect how we answer

these questions. I just want to make sure I understand the questions.

MS. JENSEN: Do you want to add to this, Joe? So, I do think it's broader than -- yes, we didn't specifically say CAR T, so is it generalizable, but I'll also look to the team to see if they want to add to anything. Okay, I'm good. Yes, you are absolutely right.

DR. CHENG: If that's the case, then it makes some of these questions challenging, like the length of duration of follow-up, because if it's not disease-specific, the duration will then obviously change.

DR. GOSS: And again, most of these measures are PRO oriented, or I should say oncology oriented, so there's an implication there, but it's not, it certainly wouldn't be relevant for cardiovascular disease, but the way we're answering some of these questions in that general sense, CMS could apply these

recommendations, I guess, more broadly. I just want to make sure we know what we're voting on.

MS. JENSEN: Correct. So, you know,

the national coverage determination that's open is CAR T, but yes, some of these answers could, depending on what happened, could be used, we might be able to use these more generally as we move forward in other types of technologies.

DR. JAMES: And I'd just like to put forth a question I have for CMS. The selection of the PROs is one that you have judged based on oncology. There's a whole host of others out there. AHRQ has developed a whole series of CAHPS measures that are used for making judgment on the quality of care that is being done to patients from their perspective. And the National Quality Forum also contracts with CMS in looking at PROMIS for the development of quality-based measurements. Are any of those in play or are those future developments?

MS. JENSEN: Those are not in play for this MEDCAC.

DR. CHENG: I would actually, then, just kind of think that we are looking at this specifically for CAR T, because for example if

you look at PROMIS, PROMIS goes from everything

from, you know, the PROMIS-10 which you can crosswalk to EQ-5D-3L for example, as a historical control to these other disease-specific measures, so I think when we're looking at this, unless we put it in the context of oncology and specifically CAR T, it would be very challenging to make heads or tails of how to answer it, because you can't compare PROMIS, for example, to MDASI outside of a specific context.

DR. ROSS: Yes, I think we should be encouraged on oncology for sure, including CAR T. I would keep us, we should not be thinking outside of the oncology space.

DR. GOSS: Just a comment, or really thought that I had that I want to share with the other panel members is particularly when you think of a situation like CAR T, I was an observer at a MEDCAC a month ago on a completely different therapeutic area, and one of the presenters got up and said, you know, one of the most important things for a patient that they want to know is what can I do to stay independent.

So on one level, PROs, everything that is local and specific to an individual patient is important to them, and you know, being functional and not being a burden on their families or their caregivers is very important, and it seems to me that the patients who got into the CAR T trials didn't get there by chance, there is significant selection bias where patients sought out treatments, they had nothing, you know, they felt they had nothing else to lose, but not every patient with a cancer actually feels that way, so some patients are willing to forgo treatment and toxicity in order to be able to have peace, you know, for whatever time they have left.

And so I think there's a -- and the industry team I think did a very nice job of presenting your studies, except I don't think your findings from your trials are generalizable to Medicare per se because of that, number one. And so I think your notion that well, we believe in PROs but we're going to measure them in trials, I think is great and is important, helps the regulators make decisions, but it doesn't generalize to what

Medicare has to deal with in terms of whether or not these should be more broadly available. And so I think it's important if you're not going to support this type of notion for going forward in some really systematic way, I think you'd be well advised to Phase IV studies to include additional PROs to help inform these questions that will inevitably come up again, because I think, you know, the population you've studied is a very slim narrow part of the population that could eventually be trying to seek out this treatment, and I think that's a concern.

DR. CUYJET: I just have a comment to make and I think one of, part of this conversation in order to be used as a brainstorming operation on how to do things better, it was mentioned that physical activity is a very important monitor for improvement. In my past experience we used telemedicine in experiences with heart failure in Medicare patients, and usually you don't just have heart failure, you have diabetes or hypertension, or an abnormal lipid profile, and if you can get

patients to invest -- it doesn't make any sense

to invest in the heart failure and not take care of your diabetes and not take care of your other comorbid conditions. So I think we ought to start thinking about the mobile technology that's emerging as an opportunity to track patient improvement independent of pure patient-reported outcomes which can be very subjective depending on time of day and how I'm feeling and how much pain I'm having. But there may be a more, a better tool to improve outcomes over a period of time, and it's stuff that can be transmitted electronically, it doesn't require -- you can decide whether you want to monitor on a weekly or monthly, or bimonthly basis, it's entirely -- I think we ought to start thinking about how going forward we can track better patient outcomes and responses more easily with better information.

DR. PERISSINOTTO: I just want to add to what you said because, or to both of you actually, because my biggest challenge now as a clinician in geriatric and palliative medicine

is exactly this question. When my patients go to see their oncologists or their surgeons, and they're trying to understand the risks and

benefits of consenting to these procedures, and most of the time the data that's presented is around survival, it's around dying in the OR and very narrow-based things. Yet what my patients want from me is to know what is my quality of life going to be like afterwards and am I going to walk, what is my cognition going to be like? So these tools, whether we use them to approve drugs, or we use them in what part, it is important to know how is this going to inform them, and help me as a clinician in assisting them in their decisions.

DR. GOSS: Yeah, I think a shared decision-making model would be really important here.

DR. PERISSINOTTO: A novel idea.

DR. GOSS: And you know, honestly, and I don't know how this would play to the PRO experts, but if you look at the PRO and getting some kind of time trade-off, and giving the

vignette of what, you know, if you think about what cytokine release syndrome looks like and explain that to a patient, you know, here's your chance of survival but here's what you're going to have to go through before you're

feeling better that might be even more relevant because that has to do with the decision to treat or not to treat, which is different than what do I look like nine months from now. So just a thought, because it's a different set of concerns, but it could be very important to patients and to providers.

MR. FRANKEL: I don't want to harp on it, but when you treat those patients in a geriatric population, when it's presented to them, do you think that it's crucial for them to see the alternate paths? So in other words, if you hone in on one potential therapy and you discuss the risks versus benefits, and they say well, they don't want to have these types of potential adverse events, and then I think a key part of that discussion has to be well, if you don't do this therapy, these are the

quote-unquote adverse effects of not doing anything and it's not exactly a pretty list either. So I think if you don't give that list in a very clear and transparent way, then the patients are not really making an informed decision, they're making a very biased decision because they're only seeing the drawbacks,

they're not seeing the optimal potential outcomes and the risks, in this case death, and a death that could potentially have a very challenging period of time until that point in the next few months.

DR. PERISSINOTTO: Yeah. I think if you really look at a shared decision-making model, you're not really starting with the risks and benefits, you're starting with what are your goals and what are you hoping for, and if you start from that point, then you back in to the risks and benefits of treatment versus not treatment. So I think that absolutely you have to, you know, weigh the cases of, for example, you have metastatic GI cancer and you can go through a surgery and chemotherapy and

have significant toxicity and end up with, you know, a pouch after the surgery, and without that treatment you will have a bowel obstruction, so it is looking at how you will die. It is also looking at limited life expectancies, and as we heard with these trials, you're looking at people already with limited life expectancies, and you do have to weigh those, but it is starting from the start.

What we don't often do as clinicians is saying what are you hoping for, because if someone tells me I don't want to prolong my life and I want to focus on the quality, then that's a different thing than saying I want to prolong my life at all costs regardless of side effects.

MR. FRANKEL: Do you think that that answer can change depending on the data that's provided to them, so if a person says --

DR. PERISSINOTTO: Yes, absolutely.

MR. FRANKEL: Right, so that's what I'm saying that may be critical here, because we're dealing with a patient population where

education is key and that's what the PROs are all about, it's to be able to educate the clinician and the patient alike. And if you're only collecting and emphasizing the data of the risks versus benefits of the new therapy and not very clearly articulating the alternative course, then I just think that patients are, I mean in the context of patient advocacy, most patients in my experience want to live and they want to live with good quality of life, that's ideal.

Then the question comes, well, if you can't have that, then what's the best alternative? And many times if the best alternative is survival, it's well, how's that survival going to look, is it going to be painful next few months and death in one, let's say for example. Is it going to be a painful next six weeks and then survival with a restoration of quality of life, perhaps with CAR T therapy.

DR. JAMES: We're addressing the whole area of patient preference, which is really not

addressed in PROs, but is the next step up from that, because you can get informed information and share that with the patient, but without understanding what the patient's goals and what the family goals are, you don't have that preference.

DR. ROSS: Yeah, and I'll just note that in shared decision-making, it's not treatment yes-no, it's treatment path A versus treatment path B, and PROs are aspects of information that help inform those goals of care, they're not actually the shared decision-making themselves. So we're talking

about information that can inform the patient care plan in terms of what their goals are, what their objectives are, if quality of life is more important than mortality, or whatever the tradeoffs may be.

DR. CHENG: I think that's the disconnect that I'm seeing here, is that we're talking about the quote-unquote real world application and real world assessment versus the clinical trials and the inclusion-exclusion

criteria, because we know that when we treat patients in a clinic we don't follow exclusion-inclusion criteria the way we do in these clinical trials.

And so maybe getting back to one of the discussion points was are there other PRO assessments, I guess the question I would pose to the group is, are these too specific for cancer per se, and should we be looking at this as a simple EQ-5D to say look, all we want to care about is what's the quality of life here of a treatment, something that's easy to do. EQ-5D, I think it's hard to argue that that's an onerous add, but yet would give us a general health assessment whether or not going for a

treatment, or any type of treatment, whether it's CAR T or lifelong IVIG, et cetera, how much effect it would really help. Because that's something that we could then talk to our patients about, the whole idea of the quality of life here.

So I guess that's a question. I know we're being asked to talk about these PROs, but

one of the concerns from everything we heard is that these are just too onerous to get on a regular basis for the data that we're getting out of it, and should we take a step back and just say for example for PROMIS, let's start off with a PROMIS-10, let's start off with something modular that we can build up, but still gives us the idea that, is this treatment really helping somebody, or are we looking at administrative or other variables that the patients may or may not care about?

DR. GARRIDO: I think there's a tension between finding a scale that provides useful enough information but that is still going to be sensitive to changes after receiving a treatment. We don't want measures that are too specific related to very specific

adverse events that are only going to occur in a subset of patients or a subset of therapies, but if we go to too global of a measure, will we see any meaningful change in that after receiving some type of therapy, whether it's CAR T or something else. I don't know the

answer.

DR. PERISSINOTTO: And also I think that it was mentioned a couple times before, you know, in surveys you have patients that say oh, I remember three words from last time, I don't have problems with cognition if I remember them from last time, but certainly that's part of it. But I do like one of the things that I think Dr. Basch said in terms of the additional characteristics of maybe having some general health assessments and part of that would be dealing with function and physical health, because I think I mentioned earlier, it is clearly a struggle for all of us in how we measure cognition in a more reliable way, both in terms of adequate measures and then being self-reported.

DR. CIVIC: I have kind of a related thing, a little bit of a committee process

that, you know, we're looking at these instruments and we may or may not want to add more to our list at this point, but we've also talked about how this is a developing field and

that there aren't necessarily, you know, there might be better instruments developed in the future or CAR T specific instruments. So it's like choosing some, you know, one, two, three, four, or seven of them now, probably that's not going to preclude the addition of other instruments as they get developed, but it's not entirely clear.

DR. ROSS: Well, I can let CMS answer that. I think because it's part of the discussion questions that they are looking for our advice on things that they should be considering in the future as well. Is that correct, or not exactly?

MS. JENSEN: No, I think -- I mean, that's -- I don't think this is the end of this conversation, and so this is what we have for today.

DR. ROSS: Dr. Yang.

DR. YANG: You know, I think we can either make these PROs too specific or too

general. If you make them specific, you have the advantage of them being applied to the

treatment you're talking about. If they're too general, you put the burden on patients to decide their global assessment. And if they're nauseated at the time they're filling out the questionnaire, they're not thinking about the surgery they need next week or the IV they might need next week, they're thinking about this problem right now, so I see that as the problem in both directions.

And so -- and the other problem I have is when you're talking about metastatic cancer, for instance, the outcomes for solid tumors are all the same, so you're just discussing how much intervention, quality of life and other issues, but if you're talking about a potentially curative treatment, who fills out the questionnaire for the patient who dies, and what do they put down? So I don't really know how you can globally assess, then, the impact of the treatment if the other alternatives, if one of the possibilities is you could get over this cancer.

DR. GARRIDO: Related to that, we have

our question about the optimal duration, or how confident we are about whether we can get meaningful results if we look at a six-month trajectory, or a 12- or 24-month trajectory of PROs. I'm concerned about long-term monitoring of PROs and survival drop off, especially if we end up doing some kind of long-term follow-up of a therapy versus standard of care using a registry. So if we have patients who aren't able to answer questions either because of an adverse event or due to differential mortality in the two groups, it's going to make it very difficult to isolate these after the treatment, even with the best practices in observational data analyses.

I run into this all of the time in palliative care research where one of the main goals is improving quality of life, we're not trying to improve survival, but it's, the people who are getting palliative care versus not, no matter what we do to try and make comparable treatment groups, they're so different that it's really hard to isolate the effects of palliative care.

Just something to take into account as

we're thinking about meaningful durations for looking at these measures.

DR. CHENG: And I guess I would just answer, you know, if someone passes away, certainly functional outcomes are pretty irrelevant, so I don't think that's really a good point. But I think one of the things we're really talking about is just the challenge of postmarket surveillance of any treatment, and I don't think that's something that we can say isn't needed or is too hard to do, because the durability of any treatment is going to be pretty important irrespective of the field. And so I think from a larger standpoint, we do need to look at ways of assessing what is the durability and the long-term outcomes for our patients, and whether or not it's a short-term gain or long-term gain does depend on whether or not we want to put our patients through this overall.

So as a surgeon, if I do a surgery for, you know, for a metastatic tumor, then sure, I can get them through it and they'll do fine for six months and still pass away, but boy, is that worth it if they have

postanesthesia issues like postoperative cognitive issues, et cetera. And I think that is the question that needs to be answered here, which is, is there a surveillance tool, you know, that we can use to assess whether CAR T or other treatments have the durability of effect, or is it something that we follow for three to six months, it seems okay, and then in two years durability starts waning, and whether or not that's worthwhile, or is it the IVIG that helps keep it from getting there?

DR. ROSS: And I also want to emphasize, particularly in the realm of postmarket surveillance, we're not necessarily just thinking about these PROs for patients who lived versus died and how to then assess the missingness, but you know, quite often this type of information as new therapies come to market and other therapies gets tweaked, this happens quite commonly in the medical device space, you know, that the devices themselves improve over time, you use this type of information to better understand symptom burden

with those sort of, you know, iterative product over time, and comparatively across products.

MR. FRANKEL: In terms of the neurologic toxicities, which really goes hand in hand again with the question of how long to capture the data, I think that it was mentioned by Dr. Go and Dr. Ferrusi about 14 months or so, that Dr. Go mentioned 14 months in terms of seeing a complete response when there wasn't until that point. But what about, in terms of neurotoxicity, how long did you see that at that point at 14 months, what percentage of the patients that had neurotoxic effects did you still see at that point along the line?

DR. ROSS: If you get a question directed to you, you may stand.

DR. GO: Will Go from Kite. So yeah, we're still exploring that in all of our studies, so we don't really have all the data right now, but in general we only had at that point in time when we get a cutoff that we will then file with the FDA as well as will be publishing in a journal, we had one patient

with grade one memory impairments. So that's sort of the work that we're doing, but again, these are sort of crude measures as well, and so as I said before, we're trying to figure out

how to do this because we are very interested in PROs, as well as neurocognitive testing, so we're exploring those opportunities right now.

MR. FRANKEL: And how do you, did you adjudicate which neurotoxicities observed were related specifically to therapy versus just because of hospitalization that you see in an older population?

DR. GO: Right, where's my FDA colleague? Oh, he's gone, all right. I'm going to tap him in in a second here. So that's exactly right, and so obviously we do have attributions in our clinical study to, is it related to the CAR T therapy, is it related to disease, is it related to the cytotoxic conditioning chemotherapy.

MR. FRANKEL: Or is it delirium because of an in-hospital experience?

DR. GO: Correct, so we don't have it

specifically, so all we ask is, is it related to CAR T, yes-no, and then in our new trials is it related to disease, yes-no, and that's the only thing that we really have, it's very crude and rudimentary, but this is exactly the question to clinically, and as I used to

practice, I mean, I get delirium in the ICU with all the beeping, you know, when I was an ICU resident, so that's --

DR. GOSS: Was the neurotoxicity measured with a PRO measure or was it usually Barthel or something else?

DR. GO: So, this is why -- sorry to interrupt, but this is why the second time we did a mini-mental status exam, because one, that had already been tested in blinatumomab prospectively, but obviously you can't even do a mini-mental status exam because you're in Grade 3 neurotox that means a mini-mental status exam's a zero. And that's why, you know, rudimentary we went from a 27 to 30, which is roughly normal, the patients who had Grade 3 neurotox went to zero and then came

back to roughly 27 or 30.

This is the challenge. We didn't do any proxies, because obviously that's another challenge to collect that. And then to your point, though, this is why I think it's challenging, especially in the neurotox setting. What we try to do for consistency, number one, we use a CTCAE 4.03, we do not

have, we collected all of it, we provided all of it. And this is a challenge because some of the neurotoxicities were at the time of death and clearly with patients who had progressive disease, so this is why this is a challenge, because as a lot of people know, how do people die of leukemia and lymphoma and fascial diseases and progressive diseases, and a lot of times the patients are in an impaired neurologic state.

And I'll tap in my FDA colleague.

DR. KLUETZ: Paul Kluetz from the FDA. The issue of attribution, I can't stress, is one of the most challenging factors in evaluating clinical trial data because of all

of the situations that you've just mentioned. Disease can cause it, treatment can cause it, comorbid disease can cause it, and many times it's very complicated and challenging. In fact, this is why we don't like disease-free survival as an endpoint. Even though it would be nice and clean, when patients die, it's very hard to determine whether or not it was due to disease or due to something else.

And so what, the way we look at

attributions in a randomized trial, if it was a randomized placebo-controlled trial, even better, but we hardly see those much anymore, so in single-armed trials we just assume that for now, until we get more data, that it is at least possibly related to the drug.

DR. BAR: Specifically regarding the neurotoxicity, so there is some data from our institution, and definitely patients that are undergoing the CAR T-cell CD-19, they do have neurotoxicity, patients who develop CRS are at high risk for developing neurotoxicity, and there has been a trial that was published a few

months ago from our institution trying to understand the mechanism that caused the neurotoxicity.

There is no clear answer but there is some direction showing probably that there is some permeability of the blood-brain barrier that caused increased toxicity. However, what we found was that the neurotoxicity is usually short term, and even patients that develop neurotoxicities, patients with CRS neurotoxicity, it is usually short term and patients do recover within a number of weeks.

So when we started to look at longer-term data on those patients, we did not see the patients that had short-term neurotoxicity have some cognitive defects later, its early data, and we didn't study that very systematically, but from the data that we have, even though they had high risk of neurotoxicity if they developed CRS, it was short term and with no long-term cognitive effects.

DR. ROSS: Dr. Yang.

DR. YANG: You know, when I think about the issue of mandating a PRO, I think of you have a purpose for that, you know how to use that information if you're going to mandate its acquisition, and I wonder how I would use that information if I were a clinician and had an infinite database on PRO information, I could present 13 percent nausea incidence to a patient, five percent severe, or I could say, you know, 87 percent of patients don't have nausea, and I could say the same thing about almost every complication. And then I would also have to integrate that with, you know, you have a 30 to 35 percent chance of having a

durable complete response. So I find this, the information is definitely helpful, definitely useful, but I don't know how I would specifically apply it in a uniform consistent fashion, if I had it all.

DR. OLSON: I can respond to that to a certain extent as a patient, specifically as a patient who reported outcomes with one of the CAR T clinical trials since I was in one

unfortunately about, almost eight years ago. There was one patient, actually two patients treated before me. We had no idea what was going to happen, but fortunately I had two patients just ahead of me, and I was warned that I was going to get sick and what the symptoms were going to be and what to expect, and that really helped because when I started getting sick I went yay, it's working. But it takes a little of the scary out of it to know, okay, somebody else got treated this way, I'm reacting the same way, it makes me feel better.

And again, you know, whether it's percentages or just general information of here's what to expect, especially in clinical trials where, you know, the trial I was in, the

only animals that had been treated were mice before the three of us, so there's not much data, but as that data grows, they will feed it back to the patient who is considering a clinical trial, and I think that is really important.

And another piece of that is that I'm

part of the LLS First Connections program, so I provide to a certain extent the patient-reported outcomes, a lot of CAR T patients that we have now, to approved drugs, I'm getting probably a connection one or two times a month, and what the patients want to hear is what do I expect, what's going to happen, I've read this. And of course you have to be careful, you're not their doctor, but at the same time it's so comforting to them to hear somebody else that's been through this and they survived, and to know what they're going to expect, you know, when they go into those things, okay, you know, Doug told me that's going to happen.

I literally just yesterday got an email from one of my First Connections patients that I had talked to probably three months ago,

and she sent me a note. She said I want you to know I went through my CAR T therapy and it was really a battle, and she had a lot of neurological effects, she said they knew how to treat them, she was rough, but on the other

side she's in complete remission, and it was really worth the fight, but she knew all the stuff going in. So really, it takes the fear away.

And then I have one more comment since I have the microphone. We were talking about duration of follow-up. CAR T-19 is creating a whole new group of patients that haven't existed before. A lot of us don't have B-cells. I get my IVIG once every, right now I'm getting it every other month, I was getting it every three months, and we're feeling our way along, but to be able to continue, I'm almost eight years out as I said, but I'm still without B-cells, and there's a whole bunch of folks coming behind me, so I think long-term follow-up is going to be important.

And just one more comment about PROs and clinical trials. I get a little bit worried when I hear some folks expressing the

fact that it may make it difficult to get some clinical trial started or that it's going to slow down enrollment or whatever, and I

certainly would caution CMS with regard to how it gets the requirement for PROs in clinical trials, how it gets applied, such that it doesn't get in the way of patients getting enrolled and being able to participate in the clinical trials, because right now it offers so much help and hope to patients.

DR. ROSS: That was very helpful, thank you. Other questions from the committee, or discussion points that they want further considered?

DR. CHENG: I think, you know, when we talk about PROs and clinical trials, I think, you know, there's a number of what I would call disconnects because we're seeing a number of societies and national organizations develop their own registry effort to collect patient outcomes, whether it's Neurosurgery with QOD, or the Society of Thoracic Surgeons, et cetera, and so it seems that some of the concerns that were brought up before, with for example data acquisition I think Red Cap is a fairly cheap

or free tool. And so I think as we move

forward, I think PROs are going to be something that is going to be captured, like in Washington State where we capture scope over in Seattle on a regular basis, irrespective of whether it's a trial or not, and I think the idea of understanding what is the quality of the care we provide patients is going to be important, not just for oncology but just for medicine in general, and I'm saying that the tide is going in that direction where we have to be able to show the benefits of anything that we do in medicine, and whether we like it or not, the PROs are probably going to be the best way to do that, because you can't do a randomized controlled trial for every single question we have in medicine, not realistically.

DR. ROSS: Dr. James.

DR. JAMES: One point that Dr. Basch raised that I think we need to consider, and that is as we sit and talk in terms of what is being recorded by physicians on adverse effects versus what comes out from a PRO, there's a gap, and how do we explain to our patients that

gap between what's being reported to the FDA and what patients are reporting.

DR. ROSS: Dr. Feinglass.

DR. FEINGLASS: I think everybody on this panel, industry included, would be the first to say that the patient's view is important, and at the end of the day the patient comes first. None of us are here for any other reason than that, or I hope we're not. But I think the other piece surrounding PROs in general is the heterogeneity of the field, which in some cases the PRO is constructed to be different from another PRO on purpose, so I think what the panel has to make a decision on at the end of the day in answer to the questions from CMS are not specific to CAR T, they are specific to, are PROs useful in the arena of clinical research, and how do they inform the decisions that we are going to make while we see patients, while we conduct trials, while we design treatments.

So one of the things I want to make sure we all remember at the end of the day is not only the number one thing, that the patient is at the end of it, the second part is as

we're considering the tools, they're heterogeneous on purpose in some cases, and how are we going to use that uniformly, are they generalizable, are they not generalizable, and I think what we've heard many of the presenters say today is they are meant to be used in very specific cases, they are meant to be used with care, they are not applicable to everything, and I think as we consider the questions, we need to keep that in the back of our minds.

DR. GOSS: I was just going to say a couple last thoughts, and I agree. I mean, the patient effectively is critical, and I think it's valuable that CMS is actually asking these questions and addressing this issue. I remember a number of years back, so some of the data we can get from clinical trials that is very useful, and obviously it's almost a standard, and probably is a standard for FDA to require PRO endpoints in, or PRO data in clinical trials. And there's still, even with that, there's still some gaps, so there's opportunities to fill gaps.

My recommendation to CMS is to keep

25 asking these questions, and to be adaptable and

1 flexible because the field is in motion, it's
2 evolving, and I think there's valuable
3 information here that will guide decisions made
4 by patients, decisions made by payers on, you
5 know, what's valuable and important in
6 treatment and technology. And you know, I
7 think overall, we would be well served to
8 remember that. When we don't have complete
9 clinical information, PRO data can at least
10 provide good color and give guidance.
11 So, I remember 15 or 20 years ago, CMS
12 issued a coverage determination or an NCD for
13 treatment refractory seizures. The important
14 question was, well, it doesn't cure the
15 disease, why would we pay for this, and the
16 answer is because it showed a significant
17 reduction in the events, and there was a strong
18 correlation between the reduction in events and
19 patients' quality of lives. So there is a way
20 to bring it back to patients, and that's really
21 important for us to remember.
22 So even if we don't have a perfect

solution, it's worth trying to improve the field and make incremental gains as we go, rather than throw our hands up and say there is

nothing to do.

DR. ROSS: Okay. Do any of the committee members want to make any final comments during this discussion period?

MR. FRANKEL: I echo a point that was made a little bit earlier, that I would be hopeful that CMS would, when evaluating PROs in general, are not necessarily specific to CAR T therapy because I think it's broader than that. Dr. Basch had noted that he was skeptical of the concerns of it being a barrier to implement PROs. On the other hand, I can't help but notice that that wasn't the position that was being suggested by multiple stakeholders, both in the background materials we have, the presentations today, and anecdotally. I've heard such a sentiment before, and I would hope that there wouldn't be any barrier to access for patients because ultimately, as was just said, the patients ultimately are the focus

here, and if there was a potential barrier for a hospital or clinician to providing the CAR T therapy for a patient, or whatever therapy that might be due to the lack of resources to implement the PRO, whether the CMS would have

some kind of pathway in place, that that type of concern could be processed and addressed so that those patients wouldn't be detrimentally affected by a PRO being implemented, and that you would just get the gains from PRO, not that kind of unfortunate unintended trickle down consequence.

DR. ROSS: I think it's an important point to be cautious. I would be very surprised if there was any hospital or facility, a place that could perform CAR T and couldn't collect PROs, it's just --

MR. FRANKEL: That's basically what was presented.

DR. ROSS: I understand. And I just wanted to say, Dr. Goss, to my knowledge, and I thought about this, I do not think PROs are required as part of an oncology approval or any

other FDA regulatory action. Our FDA colleague has left us, but I just wanted to make sure that was correct.

So, we've basically chatted for an hour, we're a little bit ahead of schedule, but I think now is the time when we're going to get ready to call a motion to vote. Is there

anything formal that has to happen?

MS. JENSEN: So, not necessarily formal, but I just want to go on record. We are planning on doing this vote different than we have done in the past, not in the voting, but just that they're not going to record it on their phones or with an electronic device. We're going to, the panel will be saying their name and their vote, we will record it, you will see it behind us just because, we're doing this because we thought we might run out of time and there are 23 questions.

I also wanted to go on record to say the official vote is the piece of paper that the panelists give us, so when we are done with this meeting we will take those papers, we will

compare with what we have here and make sure that it's accurate before we post it on our website.

So before we continue, I want to make sure the panel is okay with moving forward and how we're going to vote, and that you say your name and give us your vote, we'll record it. It's supposed to be put on behind us, are they -- okay, good. So, go ahead.

DR. GOSS: One question on the ballot.

MS. JENSEN: Sure.

DR. GOSS: So question number -- are we going to answer each question and go through the vote on each question, because question number two really is contingent on the vote on question one, so is that an average score of 2.5 for my scoring, or the average of 2.5 for the group scoring is required before we would vote on number two?

DR. ROSS: The group scoring.

DR. FEINGLASS: So we will be going through them one by one.

DR. ROSS: I think it will be easier

to go one by one. I'm going to just read the questions from the beginning to make sure we're all on the same page, give everyone a chance to just think them through, and --

DR. YANG: One other clarification.

DR. ROSS: Yes, of course.

DR. YANG: With respect to section five, question B, the how confident are we that any of those studies in these populations, you're talking about usual care versus a protocol-driven intervention. Is that a

randomized trial you're talking about predominantly?

DR. ROSS: Correct, that is my understanding of the question.

DR. YANG: Okay.

DR. ROSS: So, on May 16, 2018, CMS opened a national coverage determination on CAR T-cell therapy for Medicare beneficiaries with advanced cancer. As part of this NCD analysis, MEDCAC will review the evidence specific to PROs. We are seeking recommendations from the MEDCAC panel regarding

how existing PRO assessment tools should be incorporated into future clinical studies, including future clinical studies on CAR T-cell therapy.

I think just as a side note, we've discussed future clinical studies in the oncology space and I think we've come to that as an agreement or expectation that we're talking about oncology studies specifically, including CAR T-cell therapy studies.

The MEDCAC will focus on specific PRO assessment tools and important characteristics of a PRO assessment tool.

Then we are going to assess whether the scientific evidence supports a specific number of outcome assessment studies, design characteristics, study duration, and suitable controls for applying PROs to health outcomes research. This meeting will explore these challenges. And just to note, MEDCAC panels do not make coverage determinations but CMS benefits from their advice.

So, voting questions. For each voting

question, please use the following scale identifying your level of confidence, with a score of one being low or no confidence, and five representing high confidence, so it's a scale of one to five, and I'll go one by one.

Question 1.a. How confident are you that the PRO-CTCAE, the Patient-Reported Outcomes Common Terminology Criteria for Adverse Events, is valid and generalizable to the Medicare population?

DR. CUYJET: Al Cuyjet, I'm going to vote three.

DR. CHENG: Joe Cheng, vote four.

DR. CIVIC: Diane Civic, four.

MR. FRANKEL: Naftali Frankel, three.

DR. GARRIDO: Melissa Garrido, three.

MS. ELLIS: Can you excuse me one second?

DR. ROSS: Can we start from the beginning?

DR. CUYJET: Al Cuyjet, I voted three on question 1.a.

DR. CHENG: Joe Cheng, vote four.

DR. CIVIC: Diane Civic, four.

MR. FRANKEL: Naftali Frankel, three.

DR. GARRIDO: Melissa Garrido, three.

DR. GOSS: Tom Goss, three.

DR. JAMES: Tom James, four.

DR. LAMON: Joel Lamon, four.

DR. PERISSINOTTO: Carla Perissinotto, four.

DR. FEINGLASS: Shami Feinglass, three.

DR. GOTTSCHALK: Steve Gottschalk, four.

DR. OLSON: Doug Olson, four.

DR. YANG: Jim Yang, three.

DR. ROSS: Question 1.b, how confident are you that the M.D. Anderson Symptom Inventory is valid and generalizable to the

Medicare population?

DR. CUYJET: Al Cuyjet, I vote four.

DR. CHENG: Joe Cheng, three.

DR. CIVIC: Diane Civic, three.

MR. FRANKEL: Naftali Frankel, three.

DR. GARRIDO: Melissa Garrido, three.

DR. GOSS: Tom Goss, four.

DR. JAMES: Tom James, four.

DR. LAMON: Joel Lamon, four.

DR. PERISSINOTTO: Carla Perissinotto, three.

DR. FEINGLASS: Shami Feinglass, three.

DR. GOTTSCHALK: Steve Gottschalk, three.

DR. OLSON: Doug Olson, four.

DR. YANG: Jim Yang, four.

DR. ROSS: Okay, question 1.c. How confident are you that the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire, the EORTC-QLC-C30 core questionnaire, is valid and generalizable to the Medicare population?

DR. CUYJET: Al Cuyjet, three.

DR. CHENG: Joe Cheng, four.

DR. CIVIC: Diane Civic, four.

MR. FRANKEL: Naftali Frankel, three.

DR. GARRIDO: Melissa Garrido, four.

DR. GOSS: Tom Goss, five.

DR. JAMES: Tom James, five.

DR. LAMON: Joel Lamon, four.

DR. PERISSINOTTO: Carla Perissinotto, four.

DR. FEINGLASS: Shami Feinglass, four.

DR. GOTTSCHALK: Steve Gottschalk, four.

DR. OLSON: Doug Olson, four.

DR. YANG: Jim Yang, four.

DR. ROSS: Question 1.d, how confident are you that the University of Washington Quality of Life, UW-QOL, is valid and generalizable to the Medicare population?

DR. CUYJET: Al Cuyjet, I voted two.

DR. CHENG: Joe Cheng, two.

DR. CIVIC: Diane Civic, two.

MR. FRANKEL: Naftali Frankel, one.

DR. GARRIDO: Melissa Garrido, one.

DR. GOSS: Tom Goss, one.

DR. JAMES: Tom James, two.

DR. LAMON: Joel Lamon, two.

DR. PERISSINOTTO: Carla Perissinotto, one.

DR. FEINGLASS: Shami Feinglass, two.

DR. GOTTSCHALK: Steve Gottschalk, two.

DR. OLSON: Doug Olson, two.

DR. YANG: Jim Yang, one.

DR. ROSS: Question 1.e. How confident are you that the Patient-Reported Outcome Measurement Information System or PROMIS, is valid and generalizable to the Medicare population?

DR. CUYJET: Al Cuyjet, four.

DR. CHENG: Joe Cheng, five.

DR. CIVIC: Diane Civic, four.

MR. FRANKEL: Naftali Frankel, five.

DR. GARRIDO: Melissa Garrido, four.

DR. GOSS: Tom Goss, three.

DR. JAMES: Tom James, five.

DR. LAMON: Joel Lamon, four.

DR. PERISSINOTTO: Carla Perissinotto, five.

DR. FEINGLASS: Shami Feinglass, three.

DR. GOTTSCHALK: Steve Gottschalk,

four.

DR. OLSON: Doug Olson, four.

DR. YANG: Jim Yang, four.

DR. ROSS: Question 1.f. How confident are you that the Electronic Self-Report-Cancer, ESRA-C, is valid and generalizable to the Medicare population.

DR. CUYJET: Al Cuyjet, two.

DR. CHENG: Joe Cheng, two.

DR. CIVIC: Diane Civic, one.

MR. FRANKEL: Naftali Frankel, one.

DR. GARRIDO: Melissa Garrido, one.

DR. GOSS: Tom Goss, two.

DR. JAMES: Tom James, two.

DR. LAMON: Joel Lamon, two.

DR. PERISSINOTTO: Carla Perissinotto, two.

DR. FEINGLASS: Shami Feinglass, one.

DR. GOTTSCHALK: Steve Gottschalk, two.

DR. OLSON: Doug Olson, one.

DR. YANG: Jim Yang, one.

DR. ROSS: And the final, question 1.g, how confident are you that the Functional Living Index for Cancer, or FLIC, is valid and

generalizable to the Medicare population?

DR. CUYJET: Al Cuyjet, two.

DR. CHENG: Joe Cheng, two.

DR. CIVIC: Diane Civic, one.

MR. FRANKEL: Naftali Frankel, one.

DR. GARRIDO: Melissa Garrido, one.

DR. GOSS: Tom Goss, two.

DR. JAMES: Tom James, one.

DR. LAMON: Joel Lamon, two.

DR. PERISSINOTTO: Carla Perissinotto, one.

DR. FEINGLASS: Shami Feinglass, one.

DR. GOTTSCHALK: Steve Gottschalk, one.

DR. OLSON: Doug Olson, one.

DR. YANG: Jim Yang, two.

DR. ROSS: Great. So before we move on to the next section of questions, each panel member does have an opportunity to state for the record why they voted the way they voted, or if they want to explain any of the intention behind their vote.

MR. FRANKEL: On just PROMIS, the one trend that stuck out listening to the different stakeholders was, that was the common thread, I

think, from across the board, where it was either, even those that aren't very enthusiastic about PROs in general noted that PROMIS was recommended and it was in that context. So there was, if I'm not mistaken, that was, had the broadest consensus among the speakers and different stakeholders today.

DR. ROSS: Do any other panel members have comments?

DR. YANG: I think it's not only to win, but whether they're adequate in and of themselves that is deeply important so, you know, the range of your vote matters too.

MS. JENSEN: Can you state your name for the record for that last comment, please.

DR. YANG: Jim Yang.

MS. JENSEN: Thank you.

DR. GOSS: Just one last quick comment, Tom Goss. For the PRO-CTCAE, I was concerned about the respondent burden there for many items, and I was unclear on how it's useful. It sounded like people are using bits and pieces of it, and I think that when you cut

something up that was developed as a whole,
that undermines some of the validity

potentially.
DR. ROSS: Are we allowed to take
comments at this point in response?
MS. JENSEN: One. Go ahead.
DR. BASCH: It was actually developed
as a library, so each individual item is
validated individually, so it's not meant to be
used, so actually the purpose is for people to
use little pieces of it, you know, anywhere
between, you know, one and, you know, as many
as you want.
MS. JENSEN: What's your name?
DR. BASCH: Ethan Basch.
DR. ROSS: Thank you, Dr. Basch.
Okay. Four of the PRO assessments
were rated as a 2.5 or higher. That's the
PRO-CTCAE, the MDASI -- is that how you say it
-- MDASI, the EORTC-QLQ-C30, and PROMIS.
Whoever invented PROMIS, they had a good
thought in mind, marketing in mind.
So we now move on to question number

two, which is, considering those four PRO assessments with greater than or equal to 2.5, we're going to vote whether or not those assessments -- it says combined, but are we

considering them independently? I'm looking to the CMS team to make sure that the wording is right.

(Inaudible discussion.)

DR. ROSS: So it will be all four of those.

DR. FEINGLASS: Josh, can I clarify one thing?

DR. ROSS: It's Joe, but yes.

DR. FEINGLASS: Joe, sorry.

DR. ROSS: That's fine.

DR. FEINGLASS: So my clarification is on age, and one thing we didn't discuss before, I believe that many of these that we've now picked were designed for adults, and so when we're asking this question of not sensitive to difference of age, can we make an assumption there that we're not talking about pediatrics?

DR. GOSS: Actually I don't think so,

because one of the studies showed that even in the pediatrics, they were Medicare beneficiaries, some 25 percent of the patients had Medicare, presumably because they were disabled because of their illness.

DR. FEINGLASS: So the reason I'm

asking is because it potentially changes some people's votes, because if you're looking at who is sensitive to age, if they're only designed for someone over the age of 18, that impacts it. So can we make, for the purposes of the panel in voting, can we make an assumption that we're looking at focus on the Medicare age?

DR. ROSS: Yes, I believe we are making the assumption that we are considering the use for Medicare beneficiaries with cancer.

DR. YANG: The other wording, available supporting evidence, do you mean available or sufficient?

UNIDENTIFIED PANELIST: Adequate.

DR. ROSS: I think it fits our job to say whether it's sufficient.

DR. YANG: Should that word be available or adequate? Because available means any evidence.

DR. ROSS: Would the CMS team like to respond?

DR. SZARAMA: Any evidence.

DR. ROSS: Any evidence, okay. Thank you.

DR. CIVIC: And then like for A, are we adding them all up, or each one has to stand on its own?

MS. JENSEN: So it's a single vote.

DR. CIVIC: No, I know that, but is it additive or, you know what I mean?

MS. JENSEN: Well, it is how the panel wants to interpret it, the questions are the questions, but you're making a single vote, realizing you're taking the four that you've done 2.5 or higher and saying whether, yes or no collectively on that.

DR. ROSS: So conceptually it's a challenging exercise, to consider all four PRO assessment tools and whether any, yes-no, will

meet these criteria.

So, does the panel need me to restate the four that we're voting on, or is everybody on board? Okay.

So question A, the characteristic is the breadth of measures in emotional, social and physical well-being, yes-no.

DR. CUYJET: Al Cuyjet, yes.

DR. CHENG: Joe Cheng, yes.

DR. CIVIC: Diane Civic, yes.

MR. FRANKEL: Naftali Frankel, yes.

DR. GARRIDO: Melissa Garrido, yes.

DR. GOSS: Tom Goss, yes.

DR. JAMES: Tom James, yes.

DR. LAMON: Joel Lamon, yes.

DR. PERISSINOTTO: Carla Perissinotto, yes.

DR. FEINGLASS: Shami Feinglass, yes.

DR. GOTTSCHALK: Steve Gottschalk, yes.

DR. OLSON: Doug Olson, yes.

DR. YANG: Jim Yang, yes.

DR. ROSS: 2.B, quick throughput to

apply to clinical study.

DR. CUYJET: Al Cuyjet, yes, again.

DR. CHENG: Joe Cheng, yes.

DR. CIVIC: Diane Civic, yes.

MR. FRANKEL: Naftali Frankel, yes.

DR. GARRIDO: Melissa Garrido, yes.

DR. GOSS: Tom Goss, yes.

DR. JAMES: Tom James, yes.

DR. LAMON: Joel Lamon, yes.

DR. PERISSINOTTO: Carla Perissinotto, yes.

DR. FEINGLASS: Shami Feinglass, yes.

DR. GOTTSCHALK: Steve Gottschalk, yes.

DR. OLSON: Doug Olson, yes.

DR. YANG: Jim Yang, yes.

DR. ROSS: 2.C, transferable to community practice settings.

DR. CUYJET: Al Cuyjet, yes.

DR. CHENG: Joe Cheng, yes.

DR. CIVIC: Diane Civic, yes.

MR. FRANKEL: Naftali Frankel, yes.

DR. GARRIDO: Melissa Garrido, yes.

DR. GOSS: Tom Goss, yes.

DR. JAMES: Tom James, yes.

DR. LAMON: Joel Lamon, yes.

DR. PERISSINOTTO: Carla Perissinotto, yes.

DR. FEINGLASS: Shami Feinglass, yes.

DR. GOTTSCHALK: Steve Gottschalk, yes.

DR. OLSON: Doug Olson, yes.

DR. YANG: Jim Yang, yes.

DR. ROSS: 2.D, measures are not sensitive to differences in age.

DR. CUYJET: Al Cuyjet, with the clarification, yes.

DR. CHENG: Joe Cheng, no.

DR. CIVIC: Diane Civic, yes.

MR. FRANKEL: Naftali Frankel, yes.

DR. GARRIDO: Melissa Garrido, yes.

DR. GOSS: Tom Goss, yes.

DR. JAMES: Tom James, yes.

DR. LAMON: Joel Lamon, yes.

DR. PERISSINOTTO: Carla Perissinotto, yes.

DR. FEINGLASS: Shami Feinglass, yes.

DR. GOTTSCHALK: Steve Gottschalk, no.

DR. OLSON: Doug Olson, yes.

DR. YANG: Jim Yang, no.

DR. ROSS: Question 2.E, measures are not sensitive to line of therapy.

DR. CUYJET: Al Cuyjet, yes again.

DR. CHENG: Just a point of clarification. So this is a double negative, so we're saying it is sensitive to line of therapy?

MS. JENSEN: Correct.

DR. CHENG: Then no.

DR. ROSS: No, no, the measures are not sensitive to line of therapy. It doesn't matter which line of therapy they're receiving,

but PRO is still a valid assessment.

You're voting no?

DR. CHENG: I'm saying it's a double negative, so if I'm saying that PROs are sensitive to a line of therapy, the vote is no.

DR. ROSS: Right.

DR. CHENG: Then Joe Cheng, no.

DR. CIVIC: Diane Civic, yes.

MR. FRANKEL: Naftali Frankel, yes.

DR. GARRIDO: Melissa Garrido, yes.

DR. GOSS: Tom Goss, yes.

DR. JAMES: Tom James, yes.

DR. LAMON: Joel Lamon, yes.

DR. PERISSINOTTO: Carla Perissinotto, yes.

DR. FEINGLASS: Shami Feinglass, yes.

DR. GOTTSCHALK: Steve Gottschalk, no.

DR. OLSON: Doug Olson, no.

DR. YANG: Jim Yang, no.

DR. ROSS: Okay, 2.F, the measures are not sensitive to comorbidities.

DR. CUYJET: Al Cuyjet, yes.

DR. CHENG: Joe Cheng, no.

DR. CIVIC: Diane Civic, no.

MR. FRANKEL: Naftali Frankel, yes.

DR. GARRIDO: Melissa Garrido, yes.

DR. GOSS: Tom Goss, yes.

DR. JAMES: Tom James, yes.

DR. LAMON: Joel Lamon, yes.

DR. PERISSINOTTO: Carla Perissinotto,

yes.

DR. FEINGLASS: Feinglass, yes.

DR. GOTTSCHALK: Steve Gottschalk, no.

DR. OLSON: Doug Olson, yes.

DR. YANG: Jim Yang, no.

DR. ROSS: Question 2.G, measures are generalizable to studies of combinations of therapies.

DR. CUYJET: Al Cuyjet, yes, again.

DR. CHENG: Joe Cheng, yes.

DR. CIVIC: Diane Civic, yes.

MR. FRANKEL: Naftali Frankel, yes.

DR. GARRIDO: Melissa Garrido, yes.

DR. GOSS: Tom Goss, yes.

DR. JAMES: Tom James, yes.

DR. LAMON: Joel Lamon, yes.

DR. PERISSINOTTO: Carla Perissinotto, yes.

DR. FEINGLASS: Feinglass, yes.

DR. GOTTSCHALK: Steve Gottschalk,

yes.

DR. OLSON: Doug Olson, yes.

DR. YANG: Jim Yang, yes.

DR. ROSS: And the last question, 2.H, used in net benefit analysis based on symptom burden and well-being.

DR. CUYJET: Al Cuyjet, yes, again.

DR. CHENG: Joe Cheng, yes.

DR. CIVIC: Diane Civic, no.

MR. FRANKEL: Naftali Frankel, yes.

DR. GARRIDO: Melissa Garrido, yes.

DR. GOSS: Tom Goss, yes.

DR. JAMES: Tom James, yes.

DR. LAMON: Joel Lamon, no.

DR. PERISSINOTTO: Carla Perissinotto, yes.

DR. FEINGLASS: Feinglass, yes.

DR. GOTTSCHALK: Steve Gottschalk, yes.

DR. OLSON: Doug Olson, yes.

DR. YANG: Jim Yang, yes.

DR. ROSS: Again, I'd like to open it up to give panel members an opportunity to explain their vote or any of the information they want to state for the record.

Dr. Garrido.

DR. GARRIDO: This is Melissa Garrido. I used a very minimal standard, so if any of the PROs had any of the evidence, I voted yes.

DR. GOSS: Tom Goss. I would say the same thing. My assumption was that if in the aggregate either one of them covered it, then the answer had to be yes.

DR. JAMES: I'm Tom James with B and C. Specifically we've heard from some of the health systems that there were difficulties, but we heard from others that they have been able to achieve those, so that's why I voted yes, I think it's possible.

DR. ROSS: Any other panel members want to make a comment?

DR. FEINGLASS: One thing I neglected to state at the very opening of this meeting, which is probably obvious to all industry in here, but my comments reflect the all-industry point of view, they do not reflect any individual company's view.

DR. ROSS: Stated for the record.

DR. GOTTSCHALK: Steve Gottschalk. I just want to state for D, since I'm the only

pediatrician on the panel, I think they are age sensitive, and we need PRO measurements specifically for pediatric patients.

DR. ROSS: Okay. We have two discussion questions to address before we move on. Just to state to the panel explicitly, are there PRO assessments other than those listed in question one that have adequately stated evidence-based criteria and processes that you would want to raise, bring to the attention of CMS for further consideration? Then, are there additional desired characteristics other than listed in question two that you believe should be taken into consideration? They're not voted on, these are discussion questions for the panel members, if people have responses.

DR. GOSS: So, a couple quick things. I would say -- this is Tom Goss -- I think that the FACT has been used, and it has a number of condition-specific measures that I think have been validated in a variety of cancer types.

And I would also say that the EORTC has a number of tumor-specific add-on modules that I would encourage CMS to evaluate them as far as their utility for specific conditions.

DR. CHENG: I would just make a comment that we need to look at the PROs in a context of the presenting episode of care. So for example, someone made allusion to using, you know, CAR T therapy in the future for multiple myeloma, but if the patient, for example, had a pathological spine fracture with spinal cord compression or injury, they would certainly need a different type of assessment based on metastatic spine disease or their presenting episode of care, compared to using what we're talking about today as well.

DR. CUYJET: Okay, Al Cuyjet, I'll just make a comment, it might sound like a broken record, but I'm looking out at the audience, I might see a couple millennials and no Gen-Z around, so these patient-reported outcome tools have been developed by boomers and older. I think the technology is available to enable us to do a better job of collecting information, and I'll leave it at that.

DR. FEINGLASS: Shami Feinglass. The two things I'd add are from a diversity and inclusion standpoint in clinical trials. One

25 thing that was brought up by Dr. Basch, who we

1 now know when he stands up at the mic, are the
2 availability of language translations, I think
3 is really important. And as you look at
4 developing, those of you in the room who are
5 developing more patient-reported outcome
6 assessment tools, is there diversity and
7 inclusion in the people that you're looking at
8 when you're putting them, asking them those
9 questions, are those questions relevant to them
10 from a diversity and inclusion standpoint? So
11 to be specific, gender, cultural, where are
12 these people from, what do they identify as,
13 what are their languages, can they actually
14 answer your questions.
15 DR. GOSS: Tom Goss. I would just say
16 that I would also suggest that CMS evaluate
17 whether or not there are licensing fees for any
18 of the measures that we recommend, I think
19 there is some variability of some of them. And
20 I would also say that it would be important as
21 well that, for any of these measures that they
22 would consider, clearly the validity of

translations is important as already noted, and I think the -- there was another one, and if I think of it, I'll come back to it.

Oh, respondent burden. I think you should always have a sense of the time frame it will take to complete it, because the oncology patients may be fatigued or having other symptoms, so what seems like a short time, but it could be a long time, and certainly if someone were going through these symptoms and you were listing all of that, that would be, I think hard.

DR. GARRIDO: Melissa Garrido. I would add an adequate variation in the responses, so an absence of other floor and ceiling effects.

DR. ROSS: If we have no additional comments, we're going to move on to question three. How confident are you that each of the following assessment intervals are appropriate measurement periods for a valid PRO assessment?

DR. CUYJET: Al Cuyjet, question 3.a, answer one.

DR. CHENG: Joe Cheng, three.

DR. CIVIC: Diane Civic, three.

MR. FRANKEL: Naftali Frankel, three.

DR. GARRIDO: Melissa Garrido, three.

DR. ROSS: Pause, pause, pause, sorry.

So we're talking about 3.a, the variable event-dependent frequency interval.

MS. JENSEN: Yes, Garrido is three.

DR. GOSS: Tom Goss, one.

DR. JAMES: Tom James, three.

DR. LAMON: Joel Lamon, one.

DR. PERISSINOTTO: Carla Perissinotto, one.

DR. FEINGLASS: Feinglass, one.

DR. GOTTSCHALK: Steve Gottschalk, two.

DR. OLSON: Doug Olson, three.

DR. YANG: Jim Yang, four.

DR. ROSS: Again on a scale of one to five, how confident are you in the fixed time-dependency frequency interval?

DR. CUYJET: Al Cuyjet, four.

DR. CHENG: Joe Cheng, four.

DR. CIVIC: Diane Civic, four.

MR. FRANKEL: Naftali Frankel, four.

DR. GARRIDO: Melissa Garrido, three.

DR. GOSS: Tom Goss, four.

DR. JAMES: Tom James, three.

DR. LAMON: Joel Lamon, five.

DR. PERISSINOTTO: Carla Perissinotto,

five.

DR. FEINGLASS: Feinglass, four.

DR. GOTTSCHALK: Steve Gottschalk, four.

DR. OLSON: Doug Olson, four.

DR. YANG: Jim Yang, two.

DR. ROSS: Okay, question four, again a scale of one to five. How confident are you that a PRO assessment over the course of the following study duration identifies a meaningful durable treatment effect with a valid PRO? A, six months.

DR. CUYJET: Two, Al Cuyjet.

DR. CHENG: Joe Cheng, two, but specifically for CAR T.

DR. CIVIC: Diane Civic, two.

MR. FRANKEL: Naftali Frankel, two.

DR. GARRIDO: Melissa Garrido, three.

DR. GOSS: Tom Goss, two.

DR. JAMES: Tom James, two.

DR. LAMON: Joel Lamon, one.

DR. PERISSINOTTO: Carla Perissinotto, two.

DR. FEINGLASS: Feinglass, two.

DR. GOTTSCHALK: Steve Gottschalk,

two.

DR. OLSON: Doug Olson, three.

DR. YANG: Jim Yang, three.

DR. ROSS: Hold on one second. Okay, question 4.b, 12 months?

DR. CUYJET: Al Cuyjet, three.

DR. CHENG: Joe Cheng, four.

DR. CIVIC: Diane Civic, three.

MR. FRANKEL: Naftali Frankel, three.

DR. GARRIDO: Melissa Garrido, two.

DR. GOSS: Tom Goss, three.

DR. JAMES: Tom James, four.

DR. LAMON: Joel Lamon, four.

DR. PERISSINOTTO: Carla Perissinotto,

four.

DR. FEINGLASS: Feinglass, three.

DR. GOTTSCHALK: Steve Gottschalk, three.

DR. OLSON: Doug Olson, four.

DR. YANG: Jim Yang, four.

DR. ROSS: Question 4.c, 24 months?

DR. CUYJET: Al Cuyjet, five.

DR. CHENG: Joe Cheng, five.

DR. CIVIC: Diane Civic, three.

MR. FRANKEL: Naftali Frankel, four.

DR. GARRIDO: Melissa Garrido, one.

DR. GOSS: Tom Goss, four.

DR. JAMES: Tom James, five.

DR. LAMON: Joel Lamon, five.

DR. PERISSINOTTO: Carla Perissinotto, five.

DR. FEINGLASS: Feinglass, three.

DR. GOTTSCHALK: Steve Gottschalk, four.

DR. OLSON: Doug Olson, four.

DR. YANG: Jim Yang, five.

DR. ROSS: Great. It was my mistake,

I forgot to ask after question three so I'll do them together, questions three and four, I want to give panel members an opportunity to explain their voting if they would like to state for the record anything they took into consideration. That's questions three and four. Dr. Yang.

DR. YANG: Jim Yang. For question number three, I interpreted that as being based on the individual investigator in the study if you can pick the cogent times for intervals, versus automatic fixed times regardless of treatment. Is that a correct interpretation?

DR. ROSS: Well, my understanding, and other members can contribute, is that it's a fixed time interval as sort of prespecified at one week, at four weeks, at eight weeks, not necessarily that you could pick it.

DR. YANG: Not necessarily picked for every study.

DR. ROSS: Correct.

DR. YANG: But the other one, that would be something where the investigator would

decide what time intervals were the cogent ones, for 3.a?

DR. ROSS: Yes, the investigator would decide that this is the right time to ask the PRO.

DR. GOTTSCHALK: So for 3.b I interpreted it could be like for the first eight weeks it would be weekly, and then you would go to monthly intervals; is that correct?

DR. ROSS: That is correct.

DR. GOTTSCHALK: All right.

DR. CHENG: Yeah. I interpreted it with the variable event-dependent, it's just, that's the real world situation where the patient would come back to clinic at plus or

minus X number of days or weeks based on the follow-up time.

DR. GOSS: Yeah, I interpreted -- this is Tom Goss -- I interpreted that 3.a has, you define specific events and then you administer the PRO only when those events occur, and if the event doesn't happen you don't really need the PRO. So the occurrence of an event, say

neutropenia for example, as opposed to standardized set times, and these are representative set times, but in any given protocol for any particular study, the intervals would be defined based on the research question at hand. You know, it wouldn't always be weekly, it could be variable --

DR. ROSS: As long as it's fixed.

DR. GOSS: -- at three weeks or four weeks, 12 weeks, you know, 26 weeks, 52 weeks.

DR. ROSS: Right. Do people have any other comments they want to make about question four, or additional comments about three?

DR. GARRIDO: Melissa Garrido. My diminishing scores with the greater time lines reflect a diminishing confidence that we can

isolate a treatment effect from confounding factors over time.

MR. FRANKEL: On question four, my concern was just the lack of data that's available at this point in terms of durability, you know, it still remains to be seen on the

time tables that we're talking about if we're going to see positive or negative effects. So when we're talking about 14 months plus with dramatic potential responses, I just figured that a longer window of time at this point until we see data to say otherwise, is a prudent approach. But obviously, we're basing our opinions on a real lack of data, so I assume this will be reevaluated as more data comes in.

MS. JENSEN: Can you state your name just for the record, please?

MR. FRANKEL: Naftali Frankel.

MS. JENSEN: Thank you.

DR. ROSS: Any additional comments from the panel members for the record?

DR. CUYJET: I based my decision primarily on that slide that showed the longitudinal course for treatment over time, so

we have to monitor these patients over the course, there's going to be a lot of variation in this patient population and their responses, so we have to look for the responses.

DR. ROSS: Would you please just restate your name?

DR. CUYJET: Al Cuyjet, I'm sorry.

DR. YANG: This is Jim Yang, I would just like to clarify again. I am not assessing this integrating all units of times equally like, it was mentioned that with a longer time period the effects would diminish if equally valued and weighted, that's not the way I was interpreting it.

DR. ROSS: Great. We're going to move on to question number five, again, confidence on a scale of one to five, how confident are you that PRO assessments can provide meaningful results when studied with each of the following control populations, 5.a, patient him/herself, before and after intervention.

DR. CUYJET: Al Cuyjet, four.

DR. CHENG: Joe Cheng, four.

DR. CIVIC: Diane Civic, three.

MR. FRANKEL: Naftali Frankel, three.

DR. GARRIDO: Melissa Garrido, three.

DR. GOSS: Tom Goss, four.

DR. JAMES: Tom James, three.

DR. LAMON: Joel Lamon, five.

DR. PERISSINOTTO: Carla Perissinotto, five.

DR. FEINGLASS: Feinglass, four.

DR. GOTTSCHALK: Steve Gottschalk, four.

DR. OLSON: Doug Olson, five.

DR. YANG: Jim Yang, three.

DR. ROSS: Question 5.B, usual care versus a protocol-driven intervention.

DR. CUYJET: Al Cuyjet, four.

DR. CHENG: Joe Cheng, four.

DR. CIVIC: Diane Civic, four.

MR. FRANKEL: Naftali Frankel, four.

DR. GARRIDO: Melissa Garrido, three.

DR. GOSS: Tom Goss, four.

DR. JAMES: Tom James, four.

DR. LAMON: Joel Lamon, five.

DR. PERISSINOTTO: Carla Perissinotto, three.

DR. FEINGLASS: Feinglass, three.

DR. GOTTSCHALK: Steve Gottschalk,

four.

DR. OLSON: Doug Olson, three.

DR. YANG: Jim Yang, five.

DR. ROSS: And finally, question 5.C, historical control.

DR. CUYJET: Al Cuyjet, one.

DR. CHENG: Joe Cheng, two.

DR. CIVIC: Diane Civic, two.

MR. FRANKEL: Naftali Frankel, four.

DR. GARRIDO: Melissa Garrido, two.

DR. GOSS: Tom Goss, three.

DR. JAMES: Tom James, two.

DR. LAMON: Joel Lamon, one.

DR. PERISSINOTTO: Carla Perissinotto, one.

DR. FEINGLASS: Feinglass, one.

DR. GOTTSCHALK: Steve Gottschalk, one.

DR. OLSON: Doug Olson, three.

DR. YANG: Jim Yang, one.

DR. ROSS: Great, thank you. Does any panel member want to state for the record their thinking behind their votes?

DR. CUYJET: Al Cuyjet. I'll just use my experience as a clinical investigator in the

ALLHAT trial, you had to have elevated blood pressure to be enrolled whether you were on treatment of not. At the end of the study, 85 percent of our study cohort was at (inaudible) blood pressure, so I am a firm believer in protocol-driven interventions.

DR. CHENG: Joe Cheng. For historical controls, I think only a few of the PROs like PROMIS are able to be cross-walked to other historical things like EQ-5D, and so I voted down low because some of the other ones we chose would not have an easy crosswalk ability.

DR. FEINGLASS: This is Dr. Feinglass. I agree with Dr. Cheng on that.

DR. ROSS: Great. So I believe we have come to the end of our votes. We now have an opportunity for a final open panel discussion and I have only 20 minutes. Each panel member has an opportunity to give their final remarks in a maximum of two minutes if we could just go in order, and you can decline, you don't have to take advantage of this opportunity.

DR. CUYJET: This I think is my last MEDCAC meeting, I think I have to take a year

break, but it's been a very interesting experience.

DR. ROSS: Don't forget your name.

DR. CUYJET: Al Cuyjet.

DR. ROSS: I think Dr. Basch has left.

DR. CUYJET: But it's been great participating in all these discussions because it's such a wide variety of opinions regarding whatever the topic is that we discussed, and it's been very refreshing to be engaged and involved in it, so I want to thank the MEDCAC.

DR. CHENG: Joe Cheng. I echo that and thank you for all the insight that you've given me across the various spectra of this topic.

DR. CIVIC: Yes, this is Diane Civic. I really learned a lot today and am glad I participated. Just in terms of my own experience and the questions, I think I really, you know, put a lot more effort into answering the first set of questions and looking at the specific instruments, and I think, you know, the other ones were much harder maybe for a lot

of us, and based on a lot less data, but we all did the best we could.

MR. FRANKEL: Naftali Frankel. I just want to first thank everyone for the great presentations and the great discussion amongst the panel members. The only thing that I just wanted to mention in closing is that when we talk about patient-reported outcomes that it's really in the singular that we're talking about patients as individuals rather than a homogeneous population, the patients have independent needs and comorbidities and different responses. And it's very important, I think, that when discussing this general topic of patient-reported outcomes, we have to always focus on the patient as an individual rather than just as a population, and I trust that based on the conversations that we had today and the discussion that CMS will take note of that when evaluating PROs moving forward, that obviously, that it's going to be considered in that light for patients to be empowered with information as well as the

clinician through that transparent process, but the patients can learn from each other, but with keeping in mind both from the clinical side as well as the patient side, that

individuals vary greatly from each other. Thank you.

DR. GARRIDO: Melissa Garrido. Thank you to all of the speakers today for very informative and helpful presentations. I think improving PROs is a very worthwhile endeavor. I just think we should use extreme caution when trying to infer any causal relationship between PROs and the various treatments that may be considered.

DR. GOSS: Tom Goss. Thanks for letting me participate. It's been very interesting and I appreciate all the presentations made by the experts, they were very informative and helped us to really understand some of these issues in greater detail. I think our work is helpful but probably not sufficient, because there's some open questions remaining, so I hope CMS will

remain open to any additional information as it becomes available, but I love the concept of really including the patient voice in patient decision-making and assess access to treatment.

DR. JAMES: Tom James. This is my first MEDCAC, so I really appreciated the

presentations and the opportunity to be here. We all come with our own experiences to this kind of forum. As a primary care physician working with the insurance industry, I work with both individuals and populations, but my experience is in working with Picker Institute and we talk in terms of patient focus, not patient centered, because patient centered is what is being done to them, patient focus is their own preferences. This is a terrific first step for CMS moving toward patient preferences.

DR. LAMON: It's my pleasure being here. Reading these questions ahead of time put me out of my comfort zone, and I appreciate all the information. Just as an aside as a practicing physician, I trust that medical

education is still training physicians to treat one patient at a time, and all of this needs to come up to conform those decisions to that care. So I would make a comment to CMS or whomever, to say that leaning always on more data for people giving services, we need to lean on the electronic health record people to deliver a record that will allow a seamless way

that will allow us access to this data, so we're no longer in silos buying all this equipment that's replaced frequently because that is no longer adequate. We've defined what we need and now we must demand that it be provided for us.

DR. PERISSINOTTO: Carla Perissinotto. I want to echo the comment from my colleague here about more use of the EHR in information gathering. It's a privilege to be here today, I'm very impressed with just the breadth of expertise and I think that helped to have a very balanced discussion coming from multiple viewpoints. I also want to acknowledge that it's great to include someone who deals

specifically with older adults at the moment, so thank you for including me.

DR. FEINGLASS: I wanted to thank the patients that are on the panel and in this room. It's important to have your view, it's important to ground us with that view, so thank you for your time and your efforts. In addition, I found it very interesting when we were talking to our colleague at the FDA about the fact that at least in the oncology space

today, we've heard that no PRO has been used to drive a negative decision related to oncology at the FDA, so that was interesting.

Again, I think PROs have promise, real promise, no pun intended there. I think it is a field that has more development to happen in it. We are encouraged in industry by the development of the patient-reported outcomes. As you've heard, many in industry have used PROs in their trials, we think they have a purpose, and as we see going forward how these are used, we're certainly interested in seeing how this field moves forward, so thank you.

DR. GOTTSCHALK: Yeah, I would like to echo the other panel members' comments, I also really enjoyed being here, participating, and I would like to thank also the speakers. I probably have three comments.

First, my kind of take-home message is that PROs are probably not ready for prime time to be mandated for experimental therapies like CAR T-cell therapy. The second thing, I would really encourage that you really take advantage of CIBMTR. At least if you look in the stem cell transplant arena that really is the most

robust database to glean outcomes and the infrastructure is there, so that would be at least a starting place, especially since most treating physicians are transplant physicians, of CAR T-cell patients, so they're very familiar with the data requirements and the reporting requirements in this.

DR. OLSON: It's been a unique privilege to be able to participate in something like this today and I certainly learned a lot, and it was particularly

gratifying to hear so much focus on the patients and what that patient is experiencing, and it's, as I said, gratifying to hear that.

DR. YANG: I'd like to thank everyone who presented. I learned that PROs are extremely valuable instruments for acquiring information that cannot be acquired any other way. The follow-on is just as important, though, what interventions will eventuate and can we demonstrate that those have benefits back to the patient who generated those data, and that's the piece that I'm looking for still.

DR. ROSS: Then I will conclude by

just extending my appreciation to all the panel members and speakers who volunteered their time today. Chairing a meeting like this is actually quite exciting in many respects. It's the science of really two emerging fields coming together. The science of PROs has really exploded in the past decade, in no small part thanks to PCORI and the efforts of investigators who appeared here today, as well

as the science of cell-based therapy, which is due in no small part to the industry colleagues who are here, and the scientists at NIH who spent, you know, decades doing this work. I think both are now sort of coming to the cusp of actual clinical practice, which is exciting for us. And now as a general interest here among others and the geriatricians, we have to figure out how is this going to, how can we best generate evidence that's going to inform decisions not just in very specialized treatment centers but much more broadly.

So I appreciated the opportunity to help steer the conversation, keep everyone on time. Thank you very much.

MS. JENSEN: So, let me conclude on

behalf of CMS and the team, the national coverage determination team that's in the front row, thank you. Thank you for your participation, thank you for all of your comments, they are very appreciated.

And Dr. Cuyjet, let me tell you, you don't know yet this is your last MEDCAC,

because we might have scheduled another one yet and haven't told you.

DR. CUYJET: The sentiment won't change.

MS. JENSEN: We do appreciate all that you have done as well on your tenure here. And Dr. Ross, thank you for chairing this. This is your first MEDCAC ever, and we threw him into the deep end to chair it as well, and you have done a fabulous job, so thank you for that.

So just for next steps, very quickly, so this is part of our process, part of our national coverage determination process. I don't know if anyone has heard, but we opened up a national coverage determination on CAR T, so this is part of that process. You can go to our website to know, we have a tracking sheet of what the next step is, and our next step is

the proposed national coverage decision which is due in February, end of February, like February 27th, right? Many of you may know the date. So I think it's due, the proposed is due the end of -- there are several pending but I

think the end of February this one is due, so it will be public on or before that date, so that is the statutory due date and so we will meet that. The final, then, will be due 90 days after we make the proposed public, so those are our next steps.

Now we're going to take all this back and we're going to review everything that the panel has said as part of our analysis, this is one part, it is not the entire part, and we will then start drafting our coverage determination and make that public before the statutory due date or on the statutory due date.

So again, thank you very much, and anything else?

MS. ELLIS: I just need to collect the pre-score sheets from all of the panel members.

MS. JENSEN: So with that, we're concluded, so thank you very much. Safe

travels, everybody.

(The meeting adjourned at 3:10 p.m.)

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25