UNITED STATES OF AMERICA

DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOOD AND DRUG ADMINISTRATION

+ + +

CENTER FOR DEVICES AND RADIOLOGICAL HEALTH

+ + +

PUBLIC WORKSHOP - DETECTING CIRCULATING TUMOR DNA FOR CANCER SCREENING

+ + +

March 9, 2020
9:00 a.m.

White Oak Campus: The Great Room
Conference Center
10903 New Hampshire Avenue
Building 31, Room 1503
Silver Spring, Maryland

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

WORKSHOP PARTICIPANTS:

TIMOTHY STENZEL, M.D., Ph.D.
Director, Office of In Vitro Diagnostics and Radiological Health
CDRH/FDA

REENA PHILIP, Ph.D.
Director, Division of Molecular Genetics and Pathology
Office of Product Evaluation and Quality
CDRH/FDA

SOMA GHOSH, Ph.D.
Branch Chief, Division of Molecular Genetics and Pathology
Office of Product Evaluation and Quality
CDRH/FDA

ABDE ABUKHDEIR, Ph.D.
Scientific Reviewer, Division of Molecular Genetics and Pathology
Office of Product Evaluation and Quality
CDRH/FDA

JEFFREY SEIDMAN, M.D.
Medical Officer, Division of Molecular Genetics and Pathology
Office of Product Evaluation and Quality
CDRH/FDA

ANAND PATHAK, M.D., Ph.D., M.P.H.
Medical Officer, Division of Molecular Genetics and Pathology
Office of Product Evaluation and Quality
CDRH/FDA

PRAKASH JHA, M.D.
Medical Officer, Division of Molecular Genetics and Pathology
Office of Product Evaluation and Quality
CDRH/FDA

WEI WANG, Ph.D.
Mathematical Statistician, Biostatistics Team IV
Division of Clinical Evidence and Analysis 2
Office of Product Evaluation and Quality
CDRH/FDA

ERIK BLOOMQUIST, Ph.D.
Team Lead, Division of Biometrics Team V
Office of Biostatistics
Office of Translational Sciences
CDER/FDA

SUDHIR SRIVASTAVA, Ph.D., M.P.H.
National Cancer Institute

CHYKE DOUBENI, MBBS, M.P.H.
Mayo Clinic

STEPHEN GRUBER, M.D., M.P.H., Ph.D.
City of Hope

DANIEL HAYES, M.D.
University of Michigan

CHETAN BETTEGOWDA, M.D., Ph.D.
Johns Hopkins University

BARRY KRAMER, Ph.D.
Scientific Consultant, Division of Cancer Control and Population Sciences
National Cancer Institute (NCI)

BARBARA McANENY, M.D.
New Mexico Oncology Hematology Consultants, Ltd.

GEOFFREY OXNARD, M.D.
Dana-Farber Cancer Institute

JULIA BEAVER, M.D.
Office of New Drugs
CDER/FDA

PAM GALLAGHER, Ph.D.
Scientific Reviewer, Division of Molecular Genetics and Pathology
Office of Product Evaluation and Quality
CDRH/FDA

THEO deVOS
Vice President, Clinical & Scientific Affairs
Epigenomics

JOHN J. SNINSKY, Ph.D.
CellMax Life

GIRISH PUTCH, M.D., Ph.D.
Chief Medical Officer
Freenome

ERIC FUNG, M.D., Ph.D.
Vice President, Clinical Development
GRAIL, Inc.

ANNE MARIE O'BROIN LENNON, M.B.B.Ch., Ph.D.
Interim Director, Division of Gastroenterology and Hepatology
Professor of Medicine, Surgery, Radiology and Oncology
Johns Hopkins University

DANA CONNORS
Senior Scientific Project Manage
Foundation for the National Institutes of Health

JONATHAN COHEN
Small Biotechnology Business Coalition

KATHRYN LANG
Vice President, Outcomes and Evidence
Guardant Health

COLIN BEGG, Ph.D.
Memorial Sloan Kettering Cancer Center

DONALD BERRY, Ph.D.
Berry Consultants

STEVEN SKATES, Ph.D.
Massachusetts General Hospital

CONSTANTINE GATSONIS, Ph.D.
Brown University

GENE PENNELLO, Ph.D.
Division of Imaging Diagnostics and Software Reliability
Office of Science and Engineering Laboratories
CDER/FDA

INDEX

INDEX

M E E T I N G

(9:00 a.m.)

DR. PHILIP: Good morning, everyone. We are going to start the workshop. I welcome Dr. Tim Stenzel, Director, Office of In Vitro Diagnostics and Radiological Health, to welcome everyone to this workshop.

DR. STENZEL: Thank you, Reena, I appreciate the invite to kick this meeting off. This is an important meeting, we can walk and chew gum at the same time. As you might imagine, I'm pretty busy with other items lately. I had intended to attend as much of the conference today as I could, however, I'm going to be limited just to a few minutes this morning, but I will await word on the outcomes and the participation and the success of this meeting in advancing this important topic of how do we better screen for cancer with the new technologies that are coming out.

I have been interested in this topic for a very long time, dating back to the '90s, when I opened a molecular diagnostics lab at Duke University and was -- and also director, medical director of the cytogenetics lab. So we, of course, were screening in both those labs for blood-based disorders, primarily leukemias and lymphomas, doing Bcl-1 and Bcl-2, testing any of the translocations that are present in leukemias and lymphomas, monitoring, you know, several advances to FISH. At that time, Mayo Clinic started counting thousands of nuclei in order to increase the sensitivity and accuracy of FISH in blood-based testing.

Then some of my attention turned to collaborating with the breast cancer SPORE at Duke, who was interested in early detection of breast cancer. Of course, at that time there were various technologies like rare cell detection; cell search was in development and had been probably commercialized not long after that. There were other cutting-edge technologies at that time to detect rare cells and do molecular analyses on those rare cells. Our thinking at that time, of course, was if we could detect the rare cells, capture them, we

could determine their origin. At that time Abbott, among many others, probably had been doing mining of the InSIGHT database looking for tissue-specific markers of RNA expression and they identified many of them and filed patents, which is all publicly known, I imagine. They invited me and the breast cancer SPORE group at Duke to partner with them and co-develop these technologies. We were successful in the first round of EDRN and applications to join the EDRN. Our research effort focused on breast cancer, of course. Abbott had identified expression -- tissue-specific expression markers across a number of different tumors. We eventually commercialized the markers that were under development at that time through Abbott.

And then I ended up going to a company called Asuragen, where they had a lot of interesting microRNAs and we were -- we filed a number of patents around early detection of various cancers using microRNA detection in blood and until this day, you know, when now we're seeing the newer technologies absolutely parallel sequencing, next generation sequencing, there's still a very active, I believe, rare cell detection effort out there. And there's other technology, digital PCR. I became aware in the mid-2000s of BEAMing technology out of Johns Hopkins, a lot of the early work in cell-free DNA was done at Johns Hopkins. And then we were all fascinated to see Dennis Lo's publication on detection of circulating fetal cells in maternal circulation, which I think kind of just captured everybody's imagination about what we could do.

So this is a very exciting field, we are very interested here at the FDA in advancing technologies, getting them to the marketplace as soon as possible, that will help patients. You all who are participating in this conference online and in person will play an incredibly important role in achieving these ends as soon as possible, so I do have a charge for you that I wrote this morning. It is, my charge to you today is to find ways to quickly advance these cutting-edge technologies based on sound science and full validation to deliver better

care and outcomes to patients while ensuring and ushering adequate safety. We do look forward to the outcomes, the many outcomes of this successful workshop that we're holding today. We stand ready to work with you to advance these technologies in a very adaptable, flexible manner and don't hesitate to reach out to us and engage with us early as we support these efforts.

So with that, I turn it over to everyone else who's going to be helping with this conference, and thanks for participating.

(Applause.)

DR. PHILIP: Thank you, Dr. Stenzel, for coming to this workshop in spite of your very busy schedule. And thanks for everyone who is calling in and also especially to those who are here. Some of our panelists are also calling in, so there will be a little bit of a task for the moderators trying to moderate the session with panelists being on the call.

I don't have to give this slide to those who are here and those who are calling in because, you know, you are here because of your expertise in this area, but I thought I should at least project the numbers. This is from the American Cancer Society. In 2020 there will be 1,800,000 new cancer cases and 600,000 related cancer deaths projected to occur and -- but the good news is, in U.S., the mortality is going down over the past 25 years and that's partially due to screening or earlier detection.

As you all know, there are some existing recommended cancer screening methods, those such as colonoscopy for colon cancer, a mammogram for breast cancer, Pap tests and HPV testing for cervical cancer, and for the low-dose CT, that's for high risk patients for lung cancer. Those are the currently recommended screening approaches.

And, you know, throughout the workshop you'll hear these terms, so I just wanted to let you know how FDA interpret these terms right now. Screening is for a test to determine if a cancer might be present in an asymptomatic individual. And then when you hear the

term pan-cancer, what we are -- how we are interpreting it, that's a test that can identify all cancer types, whereas if you hear the term multi-cancer, that means that's a test that can detect several different types of cancer, but not all cancer types.

And here is a position paper from WHO. As you can see, the screening is looking at precancerous lesions, and, in early cancer, so that's in asymptomatic individuals, whereas when you see early diagnosis, that is in already diagnosed cancer patients.

So my task is to let you know why are we here today. You know, we have looked at single cancer screening tests, but multi-cancer screening tests are coming and it will be tasked with regulatory challenges.

And throughout today you will hear some common themes in all different sessions: what is the good design to give a multi-cancer claim, what is the appropriate sample size to get there, and where is the appropriate study endpoints. Do you look at pooled cancer together or do you have to look at an individual cancer type? And how do you evaluate the benefit-risk for this multi-cancer test? Of course, we can't also ignore the pre-analytical and analytical factors that goes along with it, so we will also discuss that.

So the agenda today, we will have Session 1 which will talk about the technical considerations, and also we'll get into some of the aspects that I talked about, about study design, tissue of origin, those topics you will constantly hear in all three different sessions. And after the first session there will be a short break and then we'll go into the second session where it will be more the clinical validation considerations will be discussed and then there will be a lunch break and then there's a public comment session. We do have some requests for public comments and I'm expecting they are here in person and if you are here in person, please contact Dr. Gallagher, she's sitting here in the front. And the last session is going to be more into the statistical considerations, and after that, we will be adjourned for the day.

So with that, I turn over to the Session 1 moderators, Dr. Ghosh and Dr. Abukhdeir.

DR. ABUKHDEIR: Well, thank you, Reena. Well, I want to thank everyone for coming out here today and make the effort to be here, so I'm going to open up Session 1, which is the State of the Science and Technical Considerations and Translation of Results into Clinical Actionability. This session will be moderated by Dr. Soma Ghosh and myself, Abde Abukhdeir.

And I'm going to give you just a brief overview of how this session is going to be laid out. So I'm going to bring up the panelists in a second and introduce them and let them introduce themselves, and then Dr. Ghosh is going to give you an overview of how FDA conducts a review and then I'll give you an example of an FDA-approved cancer screening test and then finally, we'll open it up to the panelists to answer some question that we hope to discuss.

So before we jump into that, a few logistics. So if you'd like to purchase coffee or a snack, it's outside at the kiosk, you probably saw it on your way in. You may come out these doors and make two rights and you'll see the kiosk.

If you haven't already pre-purchased your lunch and you'd like to have a lunch, please do so now because you're not going to be able to purchase it during lunch and you can also do that at the kiosk.

And finally, if you need the restrooms, out the door over here, go to the right, all the way to the end, make another right and you'll see the sign for the restrooms.

And with that, I'd like to invite the panelists to come forward, please, and find your seats over here.

(Pause.)

DR. ABUKHDEIR: So for Session 1, we have Dr. Bettegowda from Johns Hopkins, Dr. Chyke Doubeni from Mayo Clinic. On the phone we have Dr. Stephen Gruber from City

of Hope, we have Dr. Daniel Hayes from the University of Michigan, and we have Dr. Sudhir Srivastava from NCI. And I'm just going to ask each of them to give a brief background on themselves and introduce themselves. Can we start with Chetan?

DR. BETTEGOWDA: Good morning. Is that better? I'm Chetan Bettegowda, I'm a neurosurgeon at Johns Hopkins. I have a clinical interest in taking care of individuals with brain tumors, but on the laboratory side, I'm one of the principal investigators at the Ludwig Center for Cancer for Genetics and Therapeutics, that's led by Bert Vogelstein and Ken Kinzler, and our laboratory group has had probably a couple of decades' worth of interest in this arena. It's a real pleasure to be here, thanks so much.

DR. DOUBENI: Good morning, I'm Chyke Doubeni. I'm a family physician by training, but also a clinical epidemiologist. As some of you know, I'm also a member of the U.S. Preventive Services Task Force, but I am here on my own -- Preventive Services Task Force. I have spent the best part of my academic career in the area of cancer screening, principally in the area of colorectal cancer screening, and I'm delighted to be here.

DR. HAYES: My name is Dan Hayes, I'm from the University of Michigan Rogel Cancer Center, I'm a medical oncologist with a principal interest in breast cancer and I've spent most of my career regarding tumor biomarkers in general, especially more recently CTCs and ctDNA. And I have a very real conflict I think I should declare and that is that I receive research funding from Menarini Silicon Biosystems. My institution has a patent related to CTCs that is licensed to them and I get royalties from that yearly. And I'm on the advisory board of a number of diagnostic companies that I'm happy to list if people would like to know more.

DR. SRIVASTAVA: Good morning, I just wanted to introduce myself, Sudhir Srivastava, I'm head of the cancer biomarkers program in NCI and in fact, I take you back to 2000 when we started EDRN, we started doing biomarkers in plasma. Then we know it as

liquid biopsy, but now the term liquid biopsy really is something that goes maybe a long time back and as you heard, Johns Hopkins has been a leader in this area for many, many years and in fact, they have been part of EDRN from Day 1.

So we have a number of programs that address liquid biopsy. One of the latest programs being is -- is called a Precompetitive Collaboration on Liquid Biopsy. That basically brings public-private partnership to accelerate progress in liquid biopsy for the sake of early detection. And so I'm looking forward to deliberate in here. The one thing, disclaimer I have, I have no conflict of interest except that any statement I make here is not endorsed or supported by the NCI or NHI federal agency. This is personally my personal opinion. Thank you.

DR. ABUKHDEIR: And then can we have Dr. Gruber introduce himself?

DR. GRUBER: Yes, my name is Steve Gruber, I'm a cancer epidemiologist and medical oncologist by training. I'm director of the Center for Precision Medicine at the City of Hope National Medical Center, and I have spent my career concentrating on clinical cancer genetics and population-based epidemiologic studies largely focused on solid tumors. My only conflict is that I was a founder of a company called Progen International, I have equity in that company, but we do not have any products in this domain.

DR. ABUKHDEIR: Okay, thank you to everyone. And now I'm going to turn it over to Dr. Ghosh.

DR. GHOSH: Hello. Thank you, Abde, and good morning, everyone. So as Abde indicated, the purpose of this brief presentation is to bring everyone who is with us today on the same page and to set the stage for the discussions across the different panels.

So with that, I'll start with the outline of our presentation. I will give a very brief introduction on liquid biopsy as screening and early detection test, followed by an overview of how FDA reviews in vitro diagnostic tests, and then Abde will walk you through the

lessons learned from the review of a single cancer screening test. Then we will talk about the important issues to compare single site versus multisite screening, and finally end with the concept or the idea of aggregated prevalence, again, to be on the same page on this idea and trigger discussions.

In this slide I attempted to summarize some of the key advantages and limitations of a liquid biopsy test, mainly in the context of screening and early detection. As you all know, liquid biopsy is a simple noninvasive alternative to tissue biopsies and these tests are able to capture the complexities of intratumoral heterogeneity, the impact of the tumor microenvironment, and also capture the signals from multiple different clones that the tissue biopsy tests cannot capture. And in diagnosed patients, liquid biopsy test is also able to capture acquired mutations during the course of therapy and as we all know, these tests are fast and they are associated with less risks and complications compared to tissue biopsy tests. One of the most important advantages of liquid biopsy tests is when there is not enough tissue available, they present a great alternative.

However, just like all other procedures, there are multiple limitations to liquid biopsy tests and one of the primary limitations is the shedding rate of the tumor-derived materials, including ctDNA. The tumor-derived materials also include the proteins and other metabolites, the exosomes, the intact CTCs and so on. And these shedding rates are variable across the different tumor types. They are also variable because of the size of the tumor and the location of the tumor relative to the vasculature. Also, there are major challenges due to the CHIP (clonal hematopoiesis of indeterminate potential) mutations or the mutations in aging but healthy individuals and these mutations actually present an additional confounding variable to the output of the test and so we typically see that the false positive rate of the liquid biopsy tests may be impacted due to CHIP and this is something assay developers have to keep in mind when developing their assays.

So having discussed some of the key advantages and limitations, I want to take a couple of minutes to go through the FDA review process of in vitro diagnostic devices. One of the main components of our review is the intended use. The intended use should include all these elements including what the device measures, what specimen types it needs, what is the intended population and indications for use of the device, all this information should be included in the intended use of your test because that dictates the review path and evidence we need to determine whether the test is substantially equivalent if it's a moderate risk Class II test to its predicate or whether the test is safe and effective for its intended use if it's a high-risk Class III test.

There are other components of our review that include adhering to manufacturing SOPs and software documentation. Key additional components of our review include the pre-analytical, analytical, and clinical validation of the device. For liquid biopsy tests, pre-analytical validation is important. So these variables should be standardized by all assay developers because they influence the quality and quantity of the analyte that is used as input for the assay, for example, the ctDNA.

We also expect that you establish the performance of the specimen collection device or the blood collection tube. Also, standardize and optimize the specimen collection, preparation, and processing methods. These are just some examples. The liquid biopsy community has made several efforts to standardize and harmonize the pre-analytical variables and we highly encourage the assay developers to adhere to those.

The other two important components of our review, are analytical validation and clinical validation. The analytical validation studies should be conducted for all steps of the test, right from sample preparation to the results generation, and they should be conducted using the intended use clinical specimens as far as possible. However, if it's a rare tumor or a rare variant, you may use nonclinical specimens; however, you have to establish the

functional equivalence between the clinical and nonclinical specimens when you do that. Here are some of the key tests that should be conducted to determine the performance of your assay and we will go over these during our discussion section.

The final component of our review, which is clinical validation, should be aligned with the proposed intended use. A clinical study should be designed that is, again, very much aligned with intended use and based on the study results, sensitivity, specificity, as well as positive and negative predictive values of the assay should be determined.

One important consideration of the clinical validation is the benefit-risk balance, for example, what is the nature and magnitude of the benefits and risks, and what is the level of uncertainties in these results. We believe that all these different aspects should be addressed in your application.

So with this, I turn it over to Abde to walk you through the lessons learned during the review of one of the cancer screening tests.

DR. ABUKHDEIR: Thank you, Soma.

So I'm going to run you through an example of an FDA-approved single cancer screening test for colorectal cancer, called Cologuard.

So Cologuard is intended for the qualitative detection of colorectal neoplasia associated with DNA markers and occult hemoglobin. A positive test indicates the presence of colorectal cancer or advanced adenoma and it should be followed up with a diagnostic colonoscopy. It's indicated for adults of both sexes and at average risk for colorectal cancer and it's not indicated for diagnostic colonoscopy or surveillance in high-risk individuals.

And who is it not indicated for? Well, the studies were done in average-risk individuals and it has not been clinical evaluated in patients with a history of colorectal cancer, adenoma or related cancers. It has not been evaluated for patients who have had a positive result from another colorectal cancer screening test. It has not been evaluated in

patients diagnosed with a condition that puts them at higher risk for colorectal cancer, including, but not limited to, inflammatory bowel disease, colitis, Crohn's disease, familial polyposis or having a family history of colorectal cancer.

So what is Cologuard? So Cologuard is a stool test. It looks at levels of hemoglobin, methylated NDRG4, methylated BMP3, mutations in KRAS, and it measures the total amount of beta-actin DNA. All of those numbers go into an algorithm and that algorithm gives you a binary output. If the test is positive, it indicates that colorectal cancer may be present. If the test is negative, it indicates that it's likely not present, colorectal cancer.

So the data that was presented to FDA was a prospective, cross-sectional, multicenter study in men and women ages 50 to 84 who were at average risk for the development of colorectal cancer. So about 13,000 patients were enrolled but only 10,000 went into the primary study. The rest of them were not used due to unusable data. Patients were collected at 90 sites across the U.S. and Canada.

So results from Cologuard and FIT were compared to optical colonoscopy and histopathological diagnosis of significant lesions. These significant lesions were broken down into one of six categories. So Category 1 is all stages of colorectal cancer. Categories 2 through 5 were advanced adenomas of different sizes, and Category 6 was no neoplastic finding.

So when evaluating the performance of Cologuard, so it had a sensitivity of 92.3%, so 65 patients were Cologuard positive and of those, 60 were diagnosed with colorectal cancer. The study goal or the -- the study goal was to have a lower bound of the 95% confidence interval of greater than or equal to 65% and it was at 84.5%, so it passed that criteria. For specificity, it was at 86.6% and the study goal was to have a lower bound of the 95% confidence interval of greater than or equal to 85% and it was at 86%, so it also met that.

Another analysis that was done was a comparison to the FIT test, so subjects without colorectal cancer, that was Categories 2 through 6, and in this case the specificity did not outperform FIT. However, when we review a test, we look at the totality of the data and in this case, Cologuard had a sensitivity for detecting colorectal cancer at 92.3% and a specificity for detecting colorectal cancer at 84.4%. FIT was 73.8 for sensitivity and 93.4 for specificity.

So what are the key takeaways? Well, Cologuard was approved on its ability to detect colorectal cancer, and as general takeaways for all diagnostic devices that come in for FDA review, we recommend that sponsors lockdown all their cutoffs and before any clinical or analytical validation studies are performed, all clinical and analytical studies should be robust and successful to successfully support their application. And finally, we recommend that sponsors prospectively conduct studies in the appropriate intended use population.

DR. GHOSH: So thank you, Abde, again. And so you just heard some of the performance requirements for establishing performance of a single cancer screening test, and I will not go through the details of this slide which compares the single-site versus multisite screening mostly in reference to benefits and risks, but one of the questions we would like the experts' opinion today is, how do you do this analysis? What evidence you would need to establish the performance of a multi-cancer screening test, and should the benefits and risks be assessed on a cancer-by-cancer basis or across all different cancers pooled together?

And so I wanted to throw in this idea of aggregated or pooled prevalence which is very eloquently described in the context of multi-cancer screening tests by Dr. David Ahlquist in an article in *Nature Precision Oncology* in 2018. This concept is not new, it was used in other diseases. However, he used it in the context of multi-cancer screening test

analysis and if you look at the graph here, for individual cancers, the prevalence is low and therefore you get a very low PPV. However, it's a very attractive way of pooling the cancers together. All cancers are cancers of a particular system, here the GI system, if you pool them together, you come up with a higher PPV.

So is the right approach to determine the benefits and risks of a multi-cancer screening test? If it is not the right approach, we would like the experts to provide feedback on what is the right approach. What information do we need to include in the labeling of these multi-cancer screening tests that physicians would need to explain the results to their patients?

So with that, let's dive into the discussion questions. One of the first questions I would like the panel to address is what are the rationales, advantages, and disadvantages of a multi-cancer test? And I would like them to answer this question in the context of the available reasonably effective screening approaches for some cancers that are available.

So can we start with Dr. Bettegowda, please?

DR. BETTEGOWDA: Sure. So I think the earlier presentations this morning have highlighted some of the potential advantages of a multi-cancer test, but in my mind it's an opportunity to transform the concept of cancer screening from organ-by-organ, site-by-site, to a whole system, patient, individual approach. And so having blood-based test that could potentially detect multiple sites of disease is advantageous in terms of ease of use, potentially higher ability for patients to adhere to these tests, and an opportunity to detect tumors for which we currently have no existing screening modalities which, unfortunately, is the majority of cancers that affect humans.

The evidence really, I feel, needs to come from prospective interventional trials, meaning the literature is really full of observational studies from retrospective cohorts in which there's been no application of the impact of the intervention, i.e., detection of a

blood test on the risk and benefit for an individual patient. And while these retrospective studies are important in building confidence in moving forward to a prospective study much like, I think, most of the drugs or diagnostics that are approved by the FDA, really should be an interventional study showing the risk and benefit analysis on the target population for the intended use of this multi-cancer test. And so I don't think there's really a very viable alternative to that. Until we know that risk-benefit ratio, it's difficult to understand the true impact of these multi-cancer tests.

In terms of the target populations, I think one can think about a number of different opportunities. I think the highest impact would be on the general individual adult population, asymptomatic, and it's the opportunity to potentially make the biggest impact in terms of lives saved and ultimately, if this is the end goal, there is going to be no substitute until we test this patient population. Going into higher-risk populations, while informative, would not be a substitute to eventually test average-risk individuals in a prospective interventional study. So I think that if that's the ultimate goal, we should use that as a target population and start gathering data about the risk and benefit ratio.

DR. GHOSH: I would also like the panel to take into account whether a stepwise approach would be good when you target the population, like, high-risk patients with higher incidence, and then average risk, or whether including average-risk patients in the first study. So I would like you to comment on whether the stepwise approach is a better option.

DR. BETTEGOWDA: I think a stepwise approach is a reasonable consideration, but I personally think that a high-risk population, number one, often these patients are already identified and are part of screening strategies that address the risks that that particular population is susceptible to. I think when we think about positive predictive value/negative predictive value of a screening test, those numbers are going to be vastly different in a high-

risk population versus a standard-risk population. And so I think the ultimate goal is to get these multi-cancer screening tests out in the average-risk population with a potential benefit to citizens who are going to be maximized.

And so I think that, in my mind, it's not necessarily something that needs to be staged, potentially, in parallel. Or perhaps even going straightforward with a target population of average-risk individuals, in my mind, makes a lot of sense because of the potential and hope that these types of modalities have for the general populace.

DR. GHOSH: Dr. Doubeni.

DR. DOUBENI: May I clarify? You want us to address all of the bullets plus high-risk --

DR. GHOSH: I was thinking we will go question by question, but you can address all of them together because --

DR. DOUBENI: I can do that.

DR. GHOSH: -- that's how Dr. Bettegowda addressed.

DR. DOUBENI: Okay, I can do it either way.

DR. GHOSH: Sure, thanks.

DR. DOUBENI: Yeah, so I believe that the potential for multi-cancer screening is enormous. There is an opportunity for us to, in one test, be able to detect multiple cancers, especially those for which we have standard of care, and that's mostly the experience we all have in practice, so having patients come back for multiple different screening strategies, and we know very well that once you do so, the adherence and your ability to screen patients decline exponentially. It's the same with medications, multiple doses of medications will reduce adherence and the same phenomenon that we observe with cancer screening. So that's really appealing. So the advantages are clear, it's noninvasive, one blood test, patients are accustomed to having blood tests done or the samples are

accessible, and so those are clear.

The potential disadvantages as meds, cancer screening is inherently appealing, the idea that we can detect cancer, and we know that most cancers detected earlier will have much better prognosis, is therefore appealing. Unfortunately, we also know that cancer screening comes with risks. In this case, it's a very simple blood test, so if it was a blood or some other sample, it's not that much a risk from the test itself.

What can be a problem, and obviously it's something that we need to consider in addressing the safety of these, is the false positive but also false negatives and particularly over-diagnosing incidental findings and seeing the contents of the presentation relative to get a universal or multi-cancer screen. For some of these cancers we do not know if early diagnosis is going to improve outcomes and in those instances, I don't see a very clear pathway in terms of how we make the clinical utility line up clearly. So I think that is one thing the FDA should consider. If I may sort of reflect back on my other roles, on U.S. Preventive Services Task Force, the task force considers the harms and the benefits of screening assay, foundational principle, and the reason for that is simple. There are lots of things we can find, but many times some things we find end up causing more harm than the good that we all hope it will provide. And so in those scenarios there's a fairly high potential that -- so if we're going to approach the cancer screening as a potential to detect things that indolent, would never have caused cancer in the first place or cause death and therefore lead to harm and so this is really important for us to consider. So in my view, safety first. Innovation is important and critical and needs to be advanced, but safety is really important in this.

So that's my view on perhaps the first two bullets. But on that second bullet, I do agree that -- I've sort of been thinking about it since I'm talking about this for a while and I've not got around the need to demonstrate benefit. If we have a standard of care in a

particular cancer type, colorectal cancer, breast cancer, lung cancer, cervical cancer, I think it's fairly straightforward that we can benchmark a new technology on the old technology.

So I think, to me, there's a bifurcation here. So with standard of care, so we know that these screenings work, but if we can compare the new technology to the old one and it works or performs comparably, there's no reason to doubt that the outcomes will be similar and that will be in terms for the spectrum of disease that is detected by these new screening technologies. So I think that is clear.

What I do believe in, and this is really grounded, is that for those that we do not have current evidence of benefit from screening, either because of issues with the type of cancer and other issues, pancreatic cancer, ovarian cancer, I think you still have to demonstrate that it improves outcomes. See, I don't think you have many other options but to take that step, knowing that it is very difficult and challenging.

I do actually hold a view that, you know, I'm sure many in the audience who are device manufacturers know this very well, I think it's almost impossible to demonstrate the value of these tests in a very low-risk population. When I say high risk, I'm not talking of people with familial cancer. If you're looking at colorectal cancer, it makes sense, most of the trials leverage this. Cancers increase, the risk of cancer increases with age, so you're better off testing this in some people who are at higher risk because of age. These are long genetic factors because those are things that are difficult to demonstrate without knowing the performance of this very well.

So I would think that, from a sort of FDA approval study process, it would be prudent to test the performance of these tests in a higher-risk population by virtue of -- by way of this risk factor such as age, that allows you to demonstrate the performance factors of the test. That said, if you plan to use it in a more broad population, asymptomatic population, then I think the next step still needs to be demonstrated.

DR. GHOSH: Thank you.

Dr. Hayes, please.

DR. HAYES: Thank you. I want to start out and go on record that I am in favor of screening for cancers. I'm also in favor of good clinical science, which is not an oxymoron, I'm afraid some people think it is. And so therefore, in addition to the intended use, I want to emphasize the intended outcome and the intended outcome is either to reduce morbidity or improve overall survival of patients at risk for cancer. The intended use is not to find a cancer. I think that's what you sort of touched on, too, is the risk of over-diagnosis is considerable here.

The second thing I want to point out for me is that in addition to developing biomarkers through the years, I've also been very involved in several guidelines bodies, ASCO and CCN, the Association of Clinical Chemists and so on and so forth, and in that role I've had a real opportunity and pleasure of getting to co-author some papers about how do we actually decide what to use and what to recommend, both in terms of analytical validity, which I'm not going to touch on, but also clinical utility, which is a word you all don't use, but I noticed Dr. Philip bounced around about it, she used a different word but it means the same thing.

DR. GHOSH: Maybe clinical benefit?

DR. HAYES: I can't remember what you said.

DR. GHOSH: We could try clinical benefit.

DR. HAYES: Yes, thank you. And a paper that I just published with Michaela Dinan, a very talented Ph.D. clinical economist at Duke, and Gary Lyman, who had been her mentor, and Rich Schilsky, I'm sure most of you know, is the chief medical officer of ASCO. And Dr. Dinan came up with two very simple but brilliant terms that all of us think about but haven't articulated and that is how do we in clinic actually use a biomarker? And one is the

standard of care is to do nothing, the biomarker would have us opt in to do something and if the standard of care is to something, the biomarker would have us opt out and do something that we would've otherwise done. And while that may seem intuitively obvious, it is, to my knowledge, never been articulated and it's really helped me think about how we make recommendations on what we do.

So for example, there are several very good screening tests that have been pretty well demonstrated to reduce mortality, screening mammography, screening colonoscopy, Pap smears. So in that case, a circulating biomarker that's negative would have you opt out of a procedure you might not want to go through. I happen to have colonic polyps and my mother died of colon cancer, so I get a colonoscopy every 5 years. Trust me, it's not my most favorite thing to do every 5 years, but I do it and you'd have to tell me that you've got a test that has a really high negative predictive value for me to opt out of undergoing colonoscopy.

The alternative is opt in. So for example, if we don't have -- by the way, lung cancer now with CT scanning. The alternative is to have a standard of care in which there is no good screening test, one could argue pancreas cancer, for example, and you would find something that would make you opt in to a pancreatectomy that would otherwise be totally unacceptable if the test were positive and that's where the positive predictive value becomes so important. I recognize I'm being pedantic here, but I think these are important issues as we go forward and as the FDA decides what to do. And these become amplified, in my opinion, when you're doing a liquid biopsy ctDNA test because of the multiplicity of cancers that you might pick up and also the multiplicity of tests.

Now, I've been very impressed with the advances in technology over the last 2 or 3 years, so it's not just ctDNA, it's methylated ctDNA that may give you tissue specific -- it's the ability to begin to say well, one little bitty tiny piece of DNA with a P53 mutation

probably isn't really a cancer on its way up, but maybe 5 or 10 or 15 or whatever. And again, there are really smart people working on this, but I don't think we're here yet. We're going to hear more about clinical trial designs later by people who know more about this than I do in terms of screening.

But I believe we have to be very, very careful, the stakes are very high. If we tell people they can opt out of a standard of care that's shown to improve survival, the consequences are terrible and if we decide to opt in the tests that don't really allow us to improve overall survival, for example, has every woman in this room had bilateral mastectomies when they were 20? I suspect not. That would actually reduce the odds of mortality by 95 to 98% due to breast cancer in this country, but a lot of women would have something that was totally unacceptable done in order to help a few and a few is several thousand, but nonetheless.

So again, these are really important considerations that I think that our colleagues in the FDA need to consider. So I've sort of taken and hit one of these three points with my own agenda, but I'm looking forward to what the rest of the speakers have to say.

DR. GHOSH: Sure, thank you.

Dr. Srivastava.

DR. SRIVASTAVA: Well, most of things have been said, but I'm going to take a different take on it. So in biology, you think about something like all tumors shed DNA so that fosters something that they equally share. Second is that all the circulating tumor DNA represents the clones that lead to aggressive tumors, so that's also something. The third thing is that the biology of all the diseases, which is still evolving, that all the two things I said may not be mature at the time of using this multi-test, multi-screening test. So the biology is still not there in terms of how they really perform, but advantages are numerous, for example, one thing you ought to look for in all these tests and the screening is that we

need to look for a noninvasive test and also the compliance.

So there are a number of tests and you talked about Cologuard. There, if you look into what David Ahlquist talked about, effective screening tests, that there is a sensitivity time, adjuvants time, accessibility. If you look into it, it's not that high for Cologuard. Why? It could be many reasons and I know one reason that I heard in China, that people just don't try to collect the stool, it's just a culture issue. So I think that having a noninvasive blood-based test would be very, very useful and likely to be more adhered to as compared to other tests.

Other advantages that we have is that many tumors, we don't have really established the screening tests. Look into pancreas, look into ovarian and so on. So emergence of this technology of tests brings some hope to patients who might benefit from this. All things said, also the disadvantages. So multi-panel or multi-analyzed screening test may also bring the issue of over-diagnosis. So you might detect chemo types, what we call indolent, that is, that are not likely to be lethal or kill people because of the concept. So that's one challenge that we need to deal with and especially if you have a multi-analyzed panel, think about those which are localized disease, so ovarian, pancreas, and so on. And there, unless you have a very, very high specificity you're likely to have more, you know, what I call false positives or false negatives as you ratchet up the sensitivity.

So multi-analyzed tests are very promising but they need to be tailored in a way that captures the clone that's likely to become cancer, it captures also the localized disease that may be part of that panel and also, as I said earlier, that it also leads to -- based on sound biology. So we know about -- so just randomly selecting samples on what we call serendipity and having them in part of the panel may do more harm than good. So the panel should be constructed based on solid biology, what we know about each organ site and therefore they can help us to take those low-prevalence diseases. So this is my take on

those three points.

DR. GHOSH: Thank you, Dr. Srivastava.

And I think we'll move to Dr. Gruber, who's online, if you could share your perspective, please, on these three questions.

DR. GRUBER: Thank you. The panel has already raised many really important points, so I'll just highlight some features for emphasis, the first of which is I do believe that there is really an opportunity for circulating tumor DNA and other blood-based assays to have a real impact on improving the long-term prospects for cancer screening in populations, especially average-risk populations. That is driven largely by the opportunity for screening for cancers where other techniques are either expensive, inconvenient, underutilized or, in fact, not yet currently effective in their current state. The prevalence of the way in which we identify tumors that will benefit the most is one of the reasons why I agree with the earlier statement that what we really need here are prospective studies in average-risk populations. And as we'll address a little bit later in the panel, it's interesting to see some of the promising data that look at the spectrum of what tumors are being considered, and I'd just like to raise the issue that the extent of those panels for multi-cancer panel screenings ranges from one to a few cancers to as many as 12 or more, and I think we need to be attentive to the test characteristics for each of the performance components of multi-risk -- multi-cancer panels. But indeed, for many of the multi-cancer panels, in my view, they really do hold promise for the future.

DR. GHOSH: Thank you. I think we heard very important considerations to address these questions, especially the panel's note on effective screening based on liquid biopsy tests, how it's not only sensitivity but it is also access and adherence, and we also heard how the results could be analyzed keeping in mind whether it's an opt-out or an opt-in test. I think these are very important takeaways from the discussion we had so far. We'll move

on to the next question now.

DR. ABUKHDEIR: Okay, so a blood-based cancer screening test will indicate that a cancer is present, but in the absence of a tumor of origin test, a physician is going to have to continue a clinical workup until a tumor is found. So the questions that I have for the panel are, is there a benefit to a test without a tumor of origin output, and then I'd like to have a discussion on what are ways to get at the tumor of origin.

So can we start -- let's go in order here and start with Chetan.

DR. BETTEGOWDA: Sure. I think this is an incredibly important consideration for a blood-based multi-cancer test and that's indentifying where the incipient cancer resides, and I think there's a lot of interesting research out there using bioinformatics or bespoke methods to try to identify based on the pattern of mutations or other analytes where the neoplasm resides. I think, as a clinician, these sorts of modalities have yet to be tested. And so if I were given a test result and it gives me a positive tumor DNA level and says well, based on the informatics, we think this lesion resides in the ovary, that, to me, would need to be tested separately and independently in an interventional prospective fashion, that being the actual algorithm, before I feel confident that I should start with a pelvic ultrasound or a pelvic CT. I think, currently, the standard is you're worried about cancer, which is a clinical phenomenon I see on a routine basis in my patient population and the test we use is a pan-CT or a PET-CT of the whole body.

In my mind, that still remains the gold standard to identify where a neoplasm that hasn't yet been identified is still best handled because it provides anatomical localization, it identifies possible follow-up studies and, as Dr. Hayes said, these tests are not used in isolation. When we have these discussions about tumor DNA it's almost as if you get a positive result and it automatically sets off a chain reaction of events when, in fact, clinicians are used to, every day, incorporating a multitude of data points to then formulate

a decision in conjunction with the patient about what the next option should be.

And so for me, if I had a patient today that came up with a positive blood test and we don't know where the tumor resides and this is not just from tumor DNA but based on anemia or other systemic clinical findings, I'd order a pan-CT. I think having other methodologies, while interesting and important, still need to be proven and should be proven in a prospective interventional fashion to make sure that we're not hurting patients by not scanning the whole body to identify. And I think even if we did identify potential organ of possibility, we still need anatomic localization within that organ system, and so I'm not sure that you can really get away from correlating these findings with the true radiographic findings.

DR. ABUKHDEIR: Thank you.

DR. DOUBENI: So I'm a primary care family physician, so I'm going to take this from the perspective of a family physician who has a patient in front of me that presents with a blood test result that says you have cancer in your body. So the odyssey that will start is that a patient would be worried. Most people are very concerned about -- is this coming from me or from -- I think it's somewhere.

DR. HAYES: It wasn't me for a change. I wasn't banging on it.

DR. DOUBENI: You were not banging on it. Dan does that every now and then. So the patients, it will increase worry enormously of the risk of cancer even if you demonstrated, for instance, that there is no cancer by any type of a test you may carry out. I think the enormous risk from having a blanket test that says you have cancer somewhere in your body is that the diagnostic test, the downstream test that I perform is invariably going to detect something and the incidental finding rates will increase. Even for lung cancer screening in which you're doing just low-dose spectral CT, that's where you find the most incidental findings. Colonoscopies find polyps, we believe all of those are important,

but we know that only a small number of those are important. Mammography finds things.

And so you can go down the list and I think that my advice is to take this off the table as an option because there's no way to assess the benefits and you go down a rabbit hole, and you most likely would demonstrate that your test will not be effective in improving health outcomes because you will have so many more incidental findings that will literally obliterate any potential benefit that you can demonstrate from doing that test.

Now, if you're doing this in a patient who may have cancer and it's for diagnostic confirmation, it's a different scenario. Say if it was an asymptomatic population, this is completely different than someone who is presenting with symptoms for which you don't have -- for instance, potentially has cancer, you have no idea where it is and you're doing this test, that's a very different scenario. But in a patient who is coming to primary care for a wellness visit wanting to do a cancer screening, I think this is not -- should not be on the table.

DR. ABUKHDEIR: Thank you.

Dan.

DR. HAYES: I could easily say I agree, but I'm an academic and no academic gives up a chance to talk more. So again, referring back to my previous comments, it depends on what you think you're trying to do. If the issue is to opt out, there would be terrific benefit to knowing that a simple blood test tells you you don't need to do any screening. I can't imagine we'll find that, but it's possible, I suppose, and that would take a multi-tissue of origin test to do so. We kind of have that in regards to a precedent, the presence or absence of a Y chromosome. So for example, I have never had any screening for breast or ovarian cancer and I suspect there are a lot of women in this room that have never had a PSA drawn because the negative predictive value for those is pretty high. That's sort of a silly example, but at least it kind of suggests that maybe we can do something like this.

The other, though, is in terms of what about positive predictive values and there are a number of harms here. One is that if you're positive without a specific tissue of origin, what are you going to do? Probably the first thing is some sort of general imaging and we already know that our -- the current generation of imaging, PET scans, MRI, CT scans, in an asymptomatic patient, the false positives will far exceed the true positives, especially for very, very small cancers which is in theory what we're trying to find here. This is why it was necessary to do the randomized trial of pulmonary CT scans versus not for lung cancer because prior to that, all the screening tests had been negative for survival and I was very supportive of those studies.

The second is -- and I think you hit on this, Dr. Doubeni, is the anxiety involved with a positive test and you can't find anything. A friend of mine once called that the Demosthenes syndrome and that is you've got this thing hanging over your head but you don't know what to do about it. And even in my own practice, and I hate anecdotes, but almost every patient I have in which I am doing a screening test and it's done a few days before she sees me and again, remember all I do is breast cancer, this patient's coming to my clinic almost shaking about what the results are. Ninety-nine percent of the time the results are negative, but it's just the anxiety of the test and the anxiety of the results and a false positive result, it really adds to that. And I'll be interested in the public comments later today and whether there are non-physicians and non-scientists who will give us some insight into that.

The final thing that could be good is risk re-categorization and that is, for example, we have fairly solid evidence in breast cancer that taking an antiestrogen will reduce your odds of getting breast cancer and yet the uptake of that is quite low. I don't take tamoxifen, for example, and not everyone with two X chromosomes does, and so we start looking at different factors of risk that would make you opt in to tamoxifen and an AI. And

if there was a way that tumor of origin might result in either a proven prevention strategy, either systemic therapy which I think is pretty unlikely, frankly, without an obvious tumor of origin.

Or lifestyle changes. But I'm going to emphasize here, does the negative predictive value tell us that people should not live healthy lifestyles? I don't think so, I don't think we'd tell people don't exercise, go ahead and smoke, eat whatever you want to eat, because I don't think the negative predictive value is good enough to do that. So it's really an opt in and in my opinion, the healthy lifestyles people should already be opting in anyway, and we don't need a test to tell us that. So I agree with you, I think they should be off the table.

DR. ABUKHDEIR: Thank you, Dan.

DR. SRIVASTAVA: And I also work in what's likely the academics, so I'm not going to be terms of, you know, the U.S. Preventive Task Force. So I'll just respond to the second one, what is the best way to find tumor of origin? So there's no best way right now, but there are two promising ways that people are using it. One is methylation, tissue methylation, which I consider to be specific to a given tissue type, and the second one is the mutational spectrum, which are tissue specific, and those two are being used in terms of locating the tumor of origin. There are many challenges in those because there are a lot of assumptions being made when you do the machine-learning language, and so right now we don't have any best way to find it, but there are two promising ways.

Is there a benefit to test the output without two? Yes, I think so if you use them with a reflex test. For example, if I'm using PSAs, which are being used regardless if they're just simply tasked with -- but if you use a PSA in anything that can improve the positive and negative predictive values and maybe perhaps this test can do that, that could be used without knowing the tumor of origin. Similarly, we are detecting using low -- CT for lung

cancer, a number of nodules, 8 mm, 2 mm, not knowing what actually they do, then they can be also used as a reflex test to see where those nodules are likely to become cancers and so on. So there are uses without knowing the tumor of origin, if they're using the right content. Thank you.

DR. ABUKHDEIR: Thank you, Sudhir.

DR. HAYES: Just by the way, a little known secret is that both benign and malignant breast diseases cause a rise in PSA. It's been pretty well documented. Just to add --

DR. ABUKHDEIR: Steve, would you like to comment?

DR. GRUBER: Sure, just to briefly add one obviously very thoughtful comment. I think as we are concentrating on the question about tumor of origin, at least the way I like to think about it both from a clinical and public health perspective is from a diagnostic approach of probabilities. When we are screening a group of individuals who are at average risk, what we're doing is we're looking at what's the pretest probability that this test will come back positive and how will the test itself alter that post-test probability in a way that leads to action. From a diagnostic workup of the tissue of origin, I think one of the things that we can look at most carefully is the extent of the technologies within a panel that directs us towards any specific tumor. And getting back to Dr. Hayes' point about the fact that yes, we rely on imaging studies and for tests that are concentrating specifically on things like colon cancer, the imaging study is very straightforward, that is a colonoscopy.

The broader the panel, the more attentive we have to be to what are all of those post-test probabilities and indeed, mutational signatures can make a difference there. I would say we also need to be attentive to what are the different components of each of the different circulating tumor DNA tests that are out there. In fact, we know that you can certainly use mutational profiles and genomics alone to offer really high-quality predictive probabilities of what is the tumor of origin.

But then as you're able to add on additional technologies, including epigenetic methylation studies or even studies of fragmentomics to be able to look at what things are predicted based on chromatin rearrangement, that can also offer additional insights. So I think we should just continue to be attentive and FDA should continue to be attentive to the predictive potential of each of the components that help drive those post-test probabilities.

DR. ABUKHDEIR: Thank you, Steve, and thank you to the panel. We're going to move on to the next question.

DR. GHOSH: For the next set of questions we are going to change gear and ask the panel to comment on how to address the technical challenges as liquid biopsy tests are being developed for multi-cancer screening and as we discussed earlier, there are variabilities in the shedding rates for the tumor-derived materials. And so, since we are discussing in the context of early detection and screening, we want to hear our experts' opinion on how good these plasma tests are for detection of early cancers. The next question is about what are the optimum number of biomarkers one would look for or how do you evaluate that your test has an optimum number of biomarkers to detect multi-cancer. We can have the panel comment on these two questions, and we can start with Dr. Gruber, who's online.

DR. GRUBER: Thank you. Assessing the optimal number of biomarkers that will help us overcome some of the technical challenges is something that each company working in this area and each laboratory that's investigating has really invested different approaches and different resources to try to address that. We've seen some approaches, at least in the early days of biomarkers, as many on the panel have already demonstrated, where simple pure protein assays haven't worked very well and they have led us down a pathway that has been challenging to develop into public health strategies and I think what's becoming clear

is that the more data we have, the more precise that we can get. Yet, we have to be parsimonious about using the number of technical markers that offer the highest information content for any particular panel.

So we need to think about this really to divide it into thinking about the number of tumors that you're looking for and the number of markers that are relevant for each of those tumors, and I think they each have to be addressed individually. Certainly, panels that use plasma-based studies to identify a smaller number of tumors can optimize and tailor their approaches to be directed just towards the one, two or small number of cancers that they're looking for and certainly, for each of the cancers that are being evaluated in the larger panels, certainly there will be a larger number of markers that need to be used. But I'd caution us just to be careful so that we are evaluating the individual contributions of each of the markers to best understand how that influences the test characteristics of each of the tests.

DR. GHOSH: So do you have any comment about how good these tests are for detecting early cancers, given that there could be limitations in the shedding rate into circulation?

DR. GRUBER: A great question and the only thing I can comment on are the data that has been published and presented both in the literature and at meetings like AACR, and I would say that there is really promising pulmonary data based on a number of the different platforms that have used either genomic profiles alone, genomic profiles in combination with epigenetic studies or some studies that have concentrated largely on methylation-based assays, but I'm not aware of any circulating tumor DNA approach that's relied on a single marker. So it's really largely focused on a combination, but in many respects, the sensitivities and specificities of the published data out there begin to look quite promising to me.

DR. GHOSH: Thank you.

Dr. Bettegowda, could you share your perspective on these two questions, please?

DR. BETTEGOWDA: Sure. I think it's incompletely understood, the relationship between shedding rates, tumor size, tumor of origin. I think that our group and many others have shown that there's a wide range of detectable DNA that is not clearly linked to stage, size, location, and so there are biological factors that I think still are not yet fully understood.

And I think the important point here is that there is a lot of data to suggest that early detection of not only localized cancers, but earlier detection of cancers that might be more advanced but not yet symptomatic still leads to improved patient outcomes and that's because our therapeutics work better when the disease is not as disseminated, is not as large, not as bulky. And so obviously, the highest chances of cure are there when we find cancers at their earlier stages.

But I think the potential benefit of a multi-cancer test is not just to detect Stage 1 cancers, but also to detect slightly more advanced cancers for which there are good therapeutics. And so I think that that has to be considered when taking into account the potential benefit of this type of an approach. I think the ability to detect early cancers varies based on the organ, the technology, the methodologies being used and I think it's difficult to be able to answer that question across multiple cancer types. We know that for certain cancers like those involving the central nervous system, blood-based tests tend to be less capable of detecting them, whether they're early or more advanced, and there are other cancers that tend to shed more rapidly, like colon or lung, that are more readily detected. Why that's the case, I think we still need to understand.

The question of how many biomarkers are needed is, again, I think a difficult one to answer, that needs to be tested in prospective fashion, empirically. I don't know that

there's a golden number of markers that can be combined that we know of from just theoretical studies, I think it still needs to be tested practically. And what are the analytes that need to be tested? I think it's, again, something that needs prospective validation in interventional studies.

I think all of these retrospective studies give us hints, maybe it's mutations, maybe it's proteins, maybe it's methylation, maybe it's fragmentation pattern, maybe it's everything combined, but I think it's a very difficult question to answer based on where we are with our scientific understanding of a multi-cancer test. And again, I think these markers are one part of clinical assessment for our patients and not necessarily a knee-jerk reaction. So I think that we sort of need to continue to take our composite pictures of the patients that we're taking care of, their individual risks, and combining them with these multi-cancer tests while determining the best next step for them.

DR. GHOSH: Thank you. And so since it is very difficult to say what the collection of biomarkers are to design these tests, I think it is a good segue to our next bullet, the third bullet on the slide, as to there are multiple biomarkers, but we don't yet know what the correct number and type of biomarkers are for a multi-cancer test, how do you really determine the thresholds of the limits of a such multiple analyte tests? I mean, do you go analyte by analyte, cancer by cancer? I think it's really complex. So what is the panel's take on that kind of a problem? How do you really validate such a multi-cancer, multi-analyte test? Do you have to look at the validation of each threshold separately? For example, if it is multiple biomarkers, how do you really test for the limit of detection of the test? What are your thoughts?

Dr. Bettegowda.

DR. BETTEGOWDA: Yeah. You're getting right to the crux of some of the most challenging aspects of this technology and I think that the thresholds, again, are going to

vary based on the technology being used. I think this is no different than the current biomarkers we have in medicine, not just in cancer, that depending on what analyte you're testing, you're going to have a different threshold.

I think ultimately the most important threshold is one of clinical utility, which other panelists have brought up earlier today, and I think that really is something that's a composite of all the results for all the different cancers and compared to the number of cancers that are detected versus the number of potential cancers that were missed in the positive and negative predictive value. So I think it is difficult to know but I think, ideally, one would have a composite score based on prospective studies that will allow understanding of what the next step would be based on the test results.

DR. GHOSH: Sure, Dr. Hayes, please. And then we can move to Dr. Doubeni.

Dr. Hayes.

DR. HAYES: So I was struck by this slide, that all three of these questions are Clinical Chemistry 101 and I have developed a circulating protein assay, CA 15-3, and I'm not a clinical chemist and I realize I needed to go talk to my clinical chemistry friends to figure out what's needed, and there are people in the Agency that I have spoken to who know more about this in their sleep than most of us do awake. Again, I'll say that it's amplified by the multi-tumor of origin issue. But fundamentally, I think you're held to the law of ROCs and that is if you increase sensitivity, you'll lose specificity and if you increase specificity, you'll lose sensitivity unless you have a perfect assay, which I've never seen.

And so I would not ask our colleagues in the FDA to make this a top-down mandate. I think the top-down mandate is that as you're doing it, you have to think about the clinical science and I would let the really smart folks in the labs and in the companies who, with bioinformatics, figure out these answers. Again, it's Clinical Chemistry 101, but I think they have to do it in a way that gives a true answer. And of course, we're going to hear more

about this this afternoon, but the hierarchy is either a prospective randomized clinical trial which will take hundreds of thousands of patients, or a cohort prospective study which will also take thousands of patients, or case controls of archived specimens which are highly suspect to a variety of biases both pre-analytical and analytical, and clinical.

So those three statements were not made randomly, it's a hierarchy of the best level of evidence to a not very good level of evidence. One could argue that you can do a lot of the clinical chemistry with case-control archived samples as long as the pre-analytical issues have been considered and then move that up into either cohort controls or prospective randomized clinical trials. You asked about stepwise earlier, I think that's a perfectly reasonable stepwise approach as long as the archived samples are similar and there have been several examples, for example, where people pulled archived blood tests, blood specimens, out of a freezer from patients who had had surgery for cancer and compared them to normal blood samples from normal subjects walking down the street or working in the lab who had not just recently had cancer and then decided that there was a difference between the two. Well, there's no question there would be a difference between the two, it may have nothing to do with the cancer. So you have to be very careful about how you do these, but I think it's possible.

DR. GHOSH: Thank you.

Dr. Srivastava.

DR. SRIVASTAVA: We have experience in organ-specific screening tests, earlier detection tests, where we have limited the numbers somewhere between 10 to 30 and one of the considerations that we always had in mind, how -- in the context of early detection screening, how to improve negative predictive value. So our statisticians, they usually try to bring many biomarkers and to see whether it's improving or not improving. So that's one of the considerations that we do. In terms of multi-analyzed tests, I think it's wide open. I

don't know how many markers you need for a so-called -- universal cancer screening test.

Then the question about more than one threshold, I guess -- remember I said earlier that some of the low-prevalent disease may not be detectable by this multi-analyzed test, so therefore it's important that -- and I also said that you may need to tailor that test for those kind of situations. So again, I think it is doable.

The challenge is this, that in terms of many markers, our statisticians, they tell us there's -- in statistics they call it multiplicity and that can create the false discovery rate and that, I get pointed out that's -- we test each marker at a time for the diagnostic performance before you can see it, then with a combination and that, since I'm not a statistician, a statistician can deal with it and that's something that needs to be considered when we have a number of analytes or a number of biomarkers in a given panel. So be careful about multiplicity, that may lead to false discovery rate.

DR. GHOSH: Thank you. Dr. Doubeni, your thoughts, please.

DR. DOUBENI: I was thinking about what Dr. Hayes said about chemistry, and that's what I was thinking about and I couldn't quite put those pieces together in my head.

DR. HAYES: That's because I got burned about 30 years ago by a clinical chemist who said --

DR. DOUBENI: Okay.

DR. HAYES: -- you need to come talk to me.

DR. DOUBENI: Perfect. Yeah, so maybe I would sort of make some broad comments about the last bullet. I think cancer is distinct, they share some local and genetic features in common but they tend to be distinct, so I would think you want to be able to detect them at the same levels, going back to earlier conversations. So be sure that you are -- your tests are performing adequately and the performance characteristics are adequate. So to me, it's likely going to be a cancer-by-cancer specification.

But again, if we go back to the first bullet, I think embedded in that bullet is the same construct that, by the end of the day any cancer screening test needs to be able to demonstrate the ability to improve health outcomes and if you look Cologuard, for instance, Cologuard is a stool-based test. One of the things that's different than other noninvasive tests is that it has a higher sensitivity for detecting adenomas and it also has a comparable sensitivity in detecting early cancers. You know, so it's sort of -- it's a double-edged sword. Earlier cancers that are detected are more likely to be indolent cancers, but at the same time it's clear that that is one of the mechanisms by which you improve health outcomes.

And so I would encourage us to think about how we assess the ability of these tests, not just to pick up cancers that are advanced because at that point in time the benefits diminish exponentially, even with the existing therapies, I think it's a much more arduous test to do, but really be able to detect cancers at earlier stage and if not, be able to detect the precancerous lesions in the conditions for which we know there's such detection as the high potential in improving outcomes or reducing risk of the disease. So I think I agree with the earlier comments, but -- so keep that clinical object in front of us because that's what will guide any of these decisions in terms of what are you -- you know, multiple biomarkers are likely needed because we can't really do discrimination with lung cancer and breast cancer by just a single biomarker. You'll probably need to have a different threshold for different cancers because if you use the same threshold, you'll get very high sensitivity perhaps in one, but very low sensitivity in the next cancer. But I do agree that the developers will, without a doubt, take this into consideration. But at the end of the day, it's about health outcomes, right, whether it produces benefits versus increases risk.

DR. ABUKHDEIR: Thank you, Chyke.

Dan, yeah, please.

DR. HAYES: Because Dr. Doubeni just hit me with -- you didn't say it, but there's

been a term that's been around for about 20 years and that is minimal clinically important benefits and really what that is, is how much are you willing to give up, yeah, how much are you willing to accept to give up? And go back to my opt-in -- actually, Michaela Dinan's opt-in/opt-out. So if you're opting in, you know, how much benefit do you need to outweigh the risk and the risks are physical, emotional, and financial. And if you're opting out, how much of those do you want to give up, how much benefit do you want to give up in order to avoid those?

And I think this term has come up in what I've written about several times and I really like it, minimal clinically important benefits, and that's going to be in the eyes of the beholder. Not everyone's going to agree with everybody else and I think you need to worry about that. The Agency, though, has dealt with this since time -- since you guys started in terms of, you know, how much do you regulate versus how much do you let out in terms of those judgments?

DR. ABUKHDEIR: Chetan.

DR. SRIVASTAVA: This is a discussion we always have in EDRN. So you called that benefit, we call it what kind of diagnostic performance one would need in terms of, for example, there's a generally acceptable, I don't know scientifically valid though, acceptable notion that in ovarian cancer one can accept up to 10% positive predictive value. That means you do 10 biopsies, you get one positive. And this is something we have been discussing with an EDRN, what kind of diagnostic performance we can have for each organ type. No agreement on this. So again, that's indirectly related to cross-cancers analysis. So this is a good point we need to consider. At first we need to consider the -- not the maximum, but the minimum performance diagnostic parameter that we would require from any multi-analyzed or single-analyzed test.

DR. ABUKHDEIR: Chetan, go ahead, you had something to say?

DR. BETTEGOWDA: I think just to follow up on the other comments from this very valuable discussion, I think Question Number 1 gets to something that's been brought up and that's the possibility that the likelihood that would basically detect indolent disease and disease that will never go on to harm a patient if we never detected it, but we do harm because we now have seen it and intervene on it and have interventions that were previously unnecessary.

And I think that really the only way, and I sound like a broken record, to understand this, I wish our technologies were that good that we're only going to detect indolent cancers because that means we're doing something spectacularly well. I think the data and the literature suggest, in our own laboratory and others, that we're not that good, that we're likely detecting aggressive malignancies that will need intervention but really, the true answer here is, again, prospective studies. I'm an empiricist; you try it, you make it interventional, meaning you have patients intervene or physicians intervene on the result and I think that's really the only way to truly assess risk versus benefit. And I think, you know, the opt-in/opt-out paradigm that Dr. Hayes presents, I think, is a very important one and I think one that we should caution against being widely practiced because I don't know of very many people who would be in favor of using this test, multi-cancer test, today as a substitute for well-established screening strategies. So I think it's important when the FDA, if they were to approve these, to ensure much like Cologuard, is not saying opt out of colonoscopies. It's an additional assay to help patients and clinicians understand their risk of potentially having an adenoma or a carcinoma of the colon or rectum. Much like that, I think a multi-cancer test is a potential tool for patients and clinicians to use to understand whether there's an incipient cancer.

And I think that would be important when there are sort of approvals made or hopefully approvals made, that we make sure the language is around it that it's a

supplement, an addition rather than a subtraction of existing methodologies. Perhaps in the future, one day we can get a test that's that good but I think, based on existing data, that would be very unlikely. But again, the real risk/reward can only be measured if we're allowed to test these in prospective studies that are actually interventional.

DR. GHOSH: Thank you, Dr. Bettegowda, but I just wanted to add that, just as you pointed out that Cologuard is a supplement to the existing screening tests, but the comparison or the analysis that was made to get there was with the existing methodologies. So that's a point that I think the next two sessions will also discuss that, what kind of comparison is needed to establish the performance of these multi-cancer tests. Should it be an overall cancer or should it be with the ones -- should it be keeping in mind the standard of care modalities that are existing? So that is going to be a recurring discussion in the three sessions, so I just wanted to point that out. Thank you.

DR. ABUKHDEIR: Okay, with that, we're going to go to the next session. We're going to have to go through this one a little quickly. So Dr. Ghosh had mentioned earlier some of the key analytical studies that we look at, at the FDA. So I'm going to touch on two of them here, limit of blank and accuracy. So limit of blank is the smallest concentration of measurement that you can reliably detect. So one thing that we know happens, since we're talking about ctDNA, there is a phenomenon called CHIP, right, as Dr. Ghosh defined earlier. So I'd like you to comment on how critical of an issue this is for these screening tests and what are some of the methods that can overcome this limitation. And if we have time, I'll get into the next questions about accuracy. So can we start with Dr. Gruber online?

DR. GRUBER: Sure. Hematopoiesis is very, very challenging and especially when we think about it's a most commonly recognized form when we see it in patients who have what is indeterminately either somatic or, in fact, potentially even germline changes in P53, it leads to a lot of consternation. I would say that there are some technical ways to address

the source of whether or not something is indeed clonal hematopoiesis, but they're not convenient and they're not easy to do and on the germline side, that can include things like looking for a second source of tissue including something like a skin biopsy.

And the challenge, of course, here is that clonal hematopoiesis is age related, so there are some statistical and modeling ways in which we can better understand what the likelihood is that that will be reflected as CHIP versus something that is somatically derived from shedding tumor DNA. But one of the ways to just simply be attentive to it is to know that this is a real and very, very challenging problem and it also leads us to make sure that we're not overly reliant on a single marker which is particularly vulnerable to the principles of clonal hematopoiesis.

DR. ABUKHDEIR: Okay, would anyone else like to comment?

DR. HAYES: Well, the only thing I would throw in, because it's -- I wouldn't make a top-down mandate, I would say these are concerns you need to figure out and let the scientists in the field figure them out because I think it's not just CHIP, although CHIP worries us all. One of the things is, for example, comparing blood samples from patients with cancer to blood samples from otherwise normal, unaffected folks, it's going to miss a lot of false positives and I think you need to make it clear that you need to have patients who have benign conditions that can potentially result in a release of mutated cancers.

I can almost guarantee you that as you begin to get people who have benign breast diseases and that sort of thing, there are going to be mutations in blood and the issue there is how do you use the bioinformatics to weed those out and only go after the ones that are necessary? So again, clinical validity, I think, and analytical validity are the first steps. Without those, you shouldn't go forward. But the clinical utility is once you find something, does it actually help you make someone live longer or better is the clinical utility that needs to be established.

DR. ABUKHDEIR: Okay, so I'm going to jump to the next question, we just have a few minutes left. So in single-input assays we look at accuracy, so we ask -- will often ask sponsors to compare to an orthogonal method. But it's a little bit more complicated here now because we have multiple analytes, we have multiple input types, and we're trying multi-cancer tests. So can you speak to some ways to look at accuracy, to look at an orthogonal method?

So how about Dr. Bettegowda?

DR. BETTEGOWDA: I think this is a tough question and one that the field is still determining. I think that certainly one of the important considerations here is certainly analyte specific and technology specific, but I think what's common to several of the burgeoning technologies is the use of next generation sequencing and I think one of the important considerations with that technology, really, is to ensure that there's very robust error correction and methodologies to ensure that the assay itself is not generating false positives to PCR sequencing artifacts. And I think that what we certainly think about CHIP, I think the potential false positives that are generated from the assay itself and limitations of next generation sequencing, I think, are pretty vast and I think that -- well, I don't have a great composite answer for this. I do think that it's going to be important to ensure that inherent error suppression methodologies are going to be taken into account for all of the different assays to try to minimize the rates of false positives, especially if the intended use is the average-risk patient, average-risk individual, where a false positive is going to have pretty significant consequences. And so I think that doing multiple demonstrations of that are going to be quite vital going forward.

DR. ABUKHDEIR: Dr. Hayes.

DR. SRIVASTAVA: So the last question, how do you determine analytical accuracy of these tests? So this is something we all are concerned about, so we work with N-I-S-T, NIST,

which is National Institute of Standards and Technology. So I'll give you an example. So we started developing what we call reference and reference is, for example, the first reference said that we develop one for methylation study with the same primer, same things were given to seven labs to use it as internal control and I guess this is the kind of approach that also we used for other analytical tests to adjust -- to measure the analytical accuracy. What was the first question?

DR. GHOSH: Sorry, limit of blank.

DR. SRIVASTAVA: Okay, limit of blank, false positive rate. One of the methods to overcome this, you know, this is something very challenging especially unless you do somewhere a randomized trial and so on. Even the prospective trials cannot really adequately address false positive rate.

DR. GHOSH: Thank you.

Actually, we will not go through these questions because, partially, they were -- the second question was partially addressed by Dr. Srivastava, I think you were getting into whether there is a necessity to develop reference materials that could provide full control of the process from beginning to finish and I think maybe that is a perspective we heard from Dr. Srivastava.

So about 5 to 7 minutes are left for concluding this panel, so we would like to open it to questions from the audience, if they have any questions for our experts here. Please introduce yourself before you question the panelists.

DR. HAYES: While he's walking to the microphone, for the second point, the FNIH is sponsoring a standards thing I'm sure you're aware of and absolutely keep your ear to the ground on that.

DR. GHOSH: Thank you. Yes.

DR. PUTCHA: Hi, my name is Girish Putcha, I'm with Freenome. This is sort of an

academic question, so I guess it might be appropriate, Dan, given your earlier comment. But the question I had, especially to the epidemiologists since they're the public health officials on this panel, if you think in general terms about in an asymptomatic average-risk population is there some sort of target that you even have in your mind for what the -- you know, the predictive values or the likelihood ratios really ought to be to try to say that there is some clinical utility benefit for actually screening in that population?

DR. DOUBENI: I won't give you numbers, but I think you're asking -- the question you're asking I will take back to the earlier point that if you have an existing test that has been shown to be effective -- because predictive value is one thing, but we don't know if it actually produces -- provides health outcomes, positive health outcomes, so different, complete different claim. And so that's a way I will look at it because I think just taking a number as a predictive value, positive or negative, or the sensitivity or specificity I think it's probably not, because if we look at FOBT, guaiac FOBT and look at the sensitivity of guaiac FOBT, it was still able to demonstrate, depending on the number you're looking at, 15 or 18% or something higher mortality reduction in colorectal cancer with a variable sensitivity, right? So I'm not sure they can pick a number that is a correct one. Then downstream from there, we've made all the extrapolations to every screening test we have for colorectal cancer.

DR. PUTCHA: That's very fair. I wasn't expecting you to give me a number, to be frank, but I'd be curious if there was even any thought, epidemiologically, on what kind of guidance there might be, but it's a very fair answer. Thank you.

Sharmistha Ghosh, NCI. I was just thinking about two things. One is for us to have this test, even if it has the perfect sensitivity and specificity, to have it in the clinic and for a patient going for a physical, getting this blank test for a pan-cancer analysis, should we be talking about the possibility of sampling error or how to address that?

And the other thing is the mosaicism that you have in normal cells, how much that can confound our understanding, and I know at this time a lot of the tests, they're based on advanced cancers. But you know, I'm from the Division of Cancer Prevention, so for us to take it to the earlier spectrum of cancer growth where you may have the most clinical benefit, how much biological understanding we need for this, which we don't have this at this time, but Dr. Srivastava can tell you that we have -- we're working on developing this precancer atlas that may lead to some targeted leads in the future but we don't have that at this time. So factoring in all these three things, then how would the panelists comment on this?

DR. HAYES: The sampling error, one -- because that's a really big issue when you're sampling specific sites in a human body, liver or lung, whatever. But the blood test, in theory, one of the beauties of liquid biopsy is it gets around sampling error because you're sampling the entire tumor burden or body, if assuming that every site is putting the ctDNA out at the same rate.

I think the second issue for sampling error, though, is timing and the other beauty of liquid biopsies, you can do it serially pretty easily as opposed to, say, serial colonoscopy which you can't do, you know, once a year. Or you can, but it's tough. So I think the bigger issue is, you know, to do things serially and I think, again, I wouldn't make that top-down, I would have the companies or the scientists decide how often you need to do this to pick it up and take us to the minimal clinically important differences that you're willing to give up or you're willing to accept based on serial sampling.

And I think a great precedent for that right now is PSA for prostate cancer. I mean, I think we all know that the randomized trials that have been using PSA versus not are minimally positive, if positive at all, and yet the folks involved in that disease are working very hard now on looking at other assays that will help out but also, accelerated PSA rises

versus slow PSA rises and that sort of thing to determine who should have observation and who should actually have something done.

And so I'm actually impressed with the advances that have been made in that field in the last 5 years and I think the opportunity to get around that sort of thing exists for this, as well, but I'll say it again, it's amplified because you're looking at multiple different analytes and multiple different cancers.

DR. GHOSH: All right. I mean, I would just want to quickly add that we are thinking of bringing back a patient at multiple time points before going to the second tier, which could be some sort of screening, if we have that.

DR. DOUBENI: Well, I was going to actually respond to your earlier point at a high level. For most cancers, the evolution of the test used in clinical practice tends to, I think, go in the way that we're discussing because, initially, it's very hard to understand how to optimize the use of the test in clinical practice or even the interval that should be used for screening, which I think is what you're getting at. Even for the stool-based tests, we know that stool does not shed constantly by tumors and so it's -- so to sum up, it is a probability there to detect blood in the stool or the markers at the time that they collect the samples, so that's an error in the timing.

So I am not so bothered by the fact that you have false positive rates, you will have false positive rates, you want to minimize them and with experience of how to use the test and with development of the test, I would trust that we'll get to a point where there's increased accuracy in detection of tumor and reduce false positive rates. So I think that is inherent in any technology. Even lung cancer screening, we're beginning to understand how best to interpret the images to reduce the false positive rates. So I think that's inherent in many of these technologies.

DR. GHOSH: If I could request the panel to provide some concluding remarks, 1

minute each, and we can start with Dr. Stephen Gruber who is online.

DR. GRUBER: Thank you. First of all, I think the discussion here has highlighted both the challenges and the opportunities of working in this space and there is an enormous potential benefit to be able to identify cancers in those who do not yet have good screening tests available. What's also, I think, increasingly evident is that there is no one-size-fits-all approach to how we're going to be able to develop those tests. These are complex assays and prospective studies that look at the clinical characteristics of each of these tests will, I think, ultimately reveal what are the approaches that lead to the highest yield with safe and efficacious approaches to introducing tests into the market which will lead to actionable approaches to early identification and reducing cancer burden in the population.

DR. GHOSH: Thank you.

Dr. Srivastava.

DR. SRIVASTAVA: So this is an extraordinary time for molecular diagnostics. The pace and the scope of universal screening is rapidly expanding and as long as we have proper study design and intended population for those tests and also, as I said, robustness of the assay, I think this will provide an alternative to early detection in screening, especially for those organ types where we do not have any established screening tests and in fact, there are many of them. Except for five of them, many of the cancers do not have any established screening tests. But this is really an excellent time for molecular diagnostics. And as Dr. Dan Hayes pointed out, that let's not be top-down but bottom-up, so let investigator, let scientist and others in the field come up with their guidelines and criteria and perhaps that would be a time for us to debate pros and cons of those. Thank you.

DR. GHOSH: Dr. Hayes.

DR. HAYES: I'm very enthusiastic about the application of liquid biopsies in cancer across the board. I'm also very enthusiastic about making sure we have high levels of analytic validity and high levels of clinical utility and those are the things, I think, the Agency needs to address.

DR. GHOSH: Thank you.

Dr. Doubeni.

DR. DOUBENI: I'm really excited, as a family doctor, to even dream of the possibility that a blood test can detect multiple cancers and I think this is a technology that can be very disruptive and I hope that those who are working on it can continue to work on it to get to the point where it gets to clinical use and benefit patients. We should not forget that the most important concentration in cancer screening is to produce benefits and reduce harms, and so we should keep that front and center as we look at developing these technologies and strive to get to the point where we can demonstrate that. False positive rate is important because I think it would minimize the value of a test if it's not -- if it produced false positives, and that also goes to the point that we need to be able to identify a potential tumor of origin for a positive test. Thank you.

DR. GHOSH: Thank you.

Dr. Bettegowda, please.

DR. BETTEGOWDA: So I just want to echo all the other panelists' comments that this truly is an incredible time in cancer diagnostics. I think that the multi-cancer screening test offers a paradigm shift and I laud the FDA for being open to consider such paradigm shifts where we take our attention away from a single organ, a single site of detection, to multiple sites of possible origin and developing a test that treats individuals as whole persons and not just individual body parts -- the breast, the colon, the cervix -- and to try to integrate it in a way that I think makes sense to all of us as human beings, because that's who we are, is

an integrated composition of all of these organs, and to be able to have assays that address the entire body as opposed to just one part of the body, I think, has a lot of intuitive and profound potential.

As all the panelists have highlighted, I think truly we need to prove the benefit outweighs the risk and that's fundamentally the truth for everything we do in medicine and likely everything we do in real life, and I think that benefit-to-risk ratio can truly be proven with prospective interventional trials and I look forward to having those trials conducted to allow these multi-cancer tests truly meet their power and reach clinical utility. Thanks very much for your attention.

DR. GHOSH: Thank you. And on behalf of FDA, Abde and I would like to thank all our panelists for their time and for the insightful and informative session. We will take a 10-minute break and then reconvene for the next session. Thank you very much.

(Applause.)

(Off the record at 10:52 a.m.)

(On the record at 11:05 a.m.)

DR. SEIDMAN: Okay, we're ready to start Session 2. Okay, so we'll start with introductions and disclosures first. I'm Jeff Seidman, pathologist and medical officer in the Division of Molecular Genetics and Pathology in CDRH here at FDA. I have no disclosures. Anand.

DR. PATHAK: Yes, I'm Anand Pathak, medical officer, Center for Devices and Division of Molecular Genetics and Pathology.

DR. JHA: Prakash Jha, medical officer, the same division.

DR. KRAMER: I'm Barry Kramer, medical oncologist and contractor and scientific advisor at the National Cancer Institute, and I'm a previous division director in the Division of Cancer Prevention at the National Cancer Institute.

DR. McANENY: Thank you. I'm Barbara McAneny, I'm immediate past president of the American Medical Association and also happen to be a medical oncologist in independent practice managing a multispecialty cancer-focused practice.

DR. BEAVER: And I'm Julia Beaver, I'm a medical oncologist at FDA on the drug side in CDER in the Office for Oncologic Drugs.

DR. SEIDMAN: And we have Geoff Oxnard on the phone. Dr. Oxnard.

DR. OXNARD: Hi, I'm Geoff Oxnard, I'm a medical oncologist at Dana Farber Cancer Institute and I developed targeted therapies over the past years. More recently now, I collaborate to develop genomic diagnostics for advancing -- more advanced cancer care and more recently have been working on cancer detection technologies. From a clinical perspective, I do ad hoc consulting for a number of diagnostic companies and I'm an investigator on GRAIL clinical studies.

DR. SEIDMAN: Okay. Well, continuing from this morning, what are the key issues? And I think we've certainly hit a lot of the key issues in Session 1 and there's certainly going to be overlap in Session 2 with them.

So we all know that there are recommended screening tests for defined populations for breast, colon, and cervical cancer and also for high-risk groups for certain other cancers such as lung cancer. It's important to elucidate how a multi-cancer screening test would affect or complement the recommended screening tests in terms of test performance characteristics and compliance.

The target of any test should be clearly and carefully defined. A test for multiple different cancers could be difficult to evaluate without very specific analytic and clinical targets. A historic example involving screening for ovarian cancer, recent evidence suggests that perhaps a better target would be the fallopian tube.

Assessing the benefits and risks of harms or harms are key to any evaluation of a

screening test, as we've heard several times this morning. Test performance characteristics are key to FDA's evaluation and it's also important to consider the impact on public health.

So we have a conundrum here. The performance of a test, as well as the benefits and harms, may differ by organ site and tumor type. Well, first, what evidence is needed for each specific claim? But if you look at a healthy screening population, a trial in a healthy screening population would have to be quite large to obtain enough cancer cases for adequate statistical power. A screening population, by definition, is generally healthy and asymptomatic and the prevalence of even common cancers in this population is generally going to be low.

So in the process of evaluating device applications, the FDA staff review the data submitted as part of the premarket approval application and determine whether or not the data support the claims made by the sponsor. The data should demonstrate that the probable benefits of use of the device outweigh its probable risks when used in accordance with the proposed intended use. And a balanced consideration of probable benefits and probable risks is an essential part of FDA's determination that there are reasonable assurances of safety and effectiveness for the device.

So what are the benefits? What proportion of the intended use population is expected to benefit from the test? What is the nature of and magnitude of the benefits? How should benefits be weighed when they differ by cancer site and cancer type? For example, the benefits of diagnosing a curable cancer are clearly of greater magnitude as compared to the benefits of diagnosing an advanced-stage cancer.

Every test has risks. Screening tests can have significant harms. The nature and magnitude of these harms need careful assessment and we're going to get into this in our discussion. What are the consequences of false positive and false negative tests? What is the nature of the risks with respect to likelihood and severity?

Before you develop a screening test, you have to know what you're looking for. You need to have a well-defined disorder with known prevalence, known natural history, known and available preventive or treatment measures known to be more effective at the screen-detected stage. And there are certainly some cancers, in fact, many low-prevalence cancers in which there may be insufficient data to justify screening.

Okay, so let's move on to the first discussion question. Dr. Pathak will lead that discussion.

DR. PATHAK: Okay, so we basically are talking about multi-cancer blood-based tests that can detect multiple cancers from, you know, one single test and one of the issues that has emerged and was discussed in the last session was tumor of origin. So I'd like your feedback on both of these questions. First of all, do we need a tumor of origin component and if you don't have a component within the test, can you use a different modality such as imaging? And the second question is if tumor of origin is not identified, what is the appropriate follow-up?

Can we start off with Dr. Kramer?

DR. KRAMER: Well, most of the questions that have been posed, I think the answer is we don't really know and I think we've been hearing that all morning. Ideally, it would be nice to have a tumor of origin component because that cuts out a lot of the secondary tests that would be needed. In this case, if no tumor of origin component is identified, then not only is the fear, the patient fear, amplified because they have no idea where the cancer is and they have a suspicion or they've been told that there is an enhanced likelihood that they have a cancer somewhere, it would have to be either a different modality or different liquid biopsy techniques to refine the initial liquid biopsy.

So far I agree with what was said earlier, that right now CT scan has been used. Total body CT scan is sort of the only thing out there, but that's been tried before and it's

very unlikely that a total body CT scan has a net benefit. Incidentaloma incidents would rise exponentially, most of them false positives, and beyond that, one issue that hasn't been talked about is the total body radiation because even if the CT scan is negative, then there has to be a decision made when do you do the liquid biopsy or screening test again and do you need a CT scan of the whole body again, and then you're getting into radiation doses that actually have been shown to increase the incidence of cancers.

DR. PATHAK: Well, Dr. Kramer, the use of imaging in this context would be in a population that was test positive from the blood test, so they would have a higher pretest probability of being positive on the CT scan. Would that impact how you look at things?

DR. KRAMER: If you had perfect information, the answer is yes, but all the information we get is imperfect because -- especially if you don't know the tumor of origin, then any positive finding could be a flag for subsequent follow-up biopsy and so forth. And so you have to be pretty specific, I think, if the -- part of the problem, it's a double-edged sword. The sharp edge is that if you've been told that you have a higher likelihood of cancer, then a CT scan may or may not show something, but most of what it shows is likely to be -- you may even miss a cancer. If the ctDNA is good enough, the imaging test may actually be completely negative in the organ of origin but show multiple other incidentalomas that have nothing to do with the test.

DR. PATHAK: Okay, now Dr. McAneny.

DR. McANENY: Thank you very much. This is actually the crux of the matter. As someone who's likely to be the recipient of patients who have a positive screening test with one of these, and particularly if it does not target the origin of which organ, then we're stuck with doing imaging. And I would remind people that I spend a lot of time when I see breast cancer patients and I get asked the question why didn't they find this last year, if this has been here for many years, starting as one cell going to two, going to four, etc., why

didn't she see it last year?

We're going to run into the same problem with this because our imaging abilities to try to detect something that is very, very small are limited. And so if I have someone who comes in with a positive test for an unknown primary tumor and I go to their insurance company and I request to do a PET-CT or an MRI of a suspected organ or some other imaging thing, I'm likely to get denied and then what do I tell that patient and how do we find it?

And even if I can talk the insurance companies into it, the next step is to say well, we didn't find anything. Do we do a mastectomy, do we do a Whipple procedure and cause great harm if you don't happen to find that tumor at that point in time? And then if I don't find anything and we decide to go for watchful waiting, am I going to be repeating these imagings? And I absolutely agree with you that the concern then becomes the total radiology dose, the radiation exposure that we're creating with patients. So I think that as we look at this, and as each of these tests are developed, part of that development needs to be the plan of how a positive test will be worked up for future interventions and secondly, what are the recommendations if the imaging interventions are negative and we don't see anything that's actionable at that point in time, and those questions need to be clearly addressed.

DR. PATHAK: Thank you, Dr. McAneny.

Dr. Beaver.

DR. BEAVER: Sure. So I agree with a lot of the points that have been raised. Without a tumor of origin component, the level of risk of the test certainly increases, so there is -- there's more risk that additional imaging will be needed, additional invasive procedures will be needed, and so to combat that -- and I know one of the other sections is talking about clinical trial design, but clearly this type of claim requires really careful

attention to the clinical trial and demonstration of benefit in that clinical trial to offset the potential risks that my co-panelists have mentioned.

DR. PATHAK: And finally, Dr. Oxnard.

DR. OXNARD: I want to echo a couple points I've heard. I think Dr. Kramer's point about differentiating the detected cancer from an incidental cancer is a really fascinating one, right? You really need that tumor of origin (TOO) prediction in order to evaluate accuracy or else you'll say the test was positive and you end up with another thyroid nodule or something. And so it helps you know that you're -- that the positive you're getting is a positive you were looking for. But I also want to echo Dr. Beaver's point about safety, effectively that any screening test has some true positives and some false positives and that's, you know, a relative ratio. For mammography, we know that we pursue positives until we've done the biopsy and found it's negative and then we say okay, we've had enough of diligence in our evaluation of this positive result because there are certainly plenty of false positive mammograms. The same kind of diligence we need to work on with these positive blood tests and the TOO allows you to say there's a certain amount of diligence to be done, you perform the diligent evaluation and then you say well, we tried, it was negative and now we'll move on to routine follow-up and without that tissue of origin prediction, I worry that there could be excess diligence leading to accumulated excess risk.

DR. PATHAK: So I'm going to be asking a question now, but what if the tumor of origin is not a hundred percent accurate, are there risks associated with that, Dr. Oxnard?

DR. OXNARD: I think the tissue of origin, you know, may not be a hundred percent, there's still some biologic signal to sort out here. It depends sort of how you call it, right? I thought it was going to be colon cancer, but it's actually rectal cancer. I thought it was going to be squamous lung cancer but it was actually squamous throat cancer. So there are some -- some of that is terminology of almost how you frame accuracy. But I agree that a

diagnostic that sends you off in one direction and misses you from looking in the right place would be more risky. I guess I would respond that we haven't totally sorted out how best to evaluate the accuracy of these tests. But you're right that you want the right amount of accuracy to send you down the right direction in a productive way, ideally without excess imaging of all patients which could lead to excess incidentals, so there's got to be a balance there.

DR. PATHAK: Okay, thank you so much. Now moving on to the next question. First of all, the question is how do you establish truth in terms of clinical status? And I think we may all agree that a histopathological diagnosis would be the ultimate truth. But a question we're concerned with is how much follow-up for the negatives on the blood test is needed to verify that there are true negatives?

Dr. Kramer.

DR. KRAMER: So again, it's an unknown but it, again, is a double-edged sword. First of all, let me focus on histopathologic diagnosis. If there's something to go after, I think, and if you've identified an abnormality, there are also some reasons that you had to go after it and sometimes you're going after multiple abnormalities and each with its own harm. We also have to remember that any histopathologic diagnosis is based on technology that we have used for a hundred years now and we don't -- and unfortunately, we know the natural history of virtually nothing that has been detected by most of our screening tests.

That's true even for our commonly accepted screening tests and here it's going to be even worse because we will dip more and more deeply into the iceberg of disease, the body of the iceberg or lesions that look like a cancer to the pathologist but may not have the same natural history at all as the tip of the iceberg that we have been able to observe over scores of hundreds of years.

How much follow-up is required for negatives? Again, we don't know. A prospective

trial, and by that I mean randomized trial with internal controls, is probably the only way to do that. But remember, if we keep following up with blood tests from negatives, then there's another biological process going on and that is more and more CHIPs are probably accumulating. Age and cancer are increasing and it may make the interpretation of a test even more difficult and that's why I think it's a big challenge to the FDA to judge single tests and extrapolate when you need the next test and how frequently.

DR. PATHAK: Okay, Dr. McAneny.

DR. McANENY: Well, I think, in the first part the histopathological diagnosis, until the validity of the tests have been really well proven, I don't think that I would intervene on a treatment plan without proof of biopsy, a standard biopsy rather than a liquid biopsy, that the initial diagnosis is there. That will be different first for if we use these tests to follow people who are in remission and we're trying to see if they are developing a recurrence, but for right now in the setting of screening, I think you have to establish a firm diagnosis.

The follow-up for negatives is particularly tricky because I think there will be a lot of patients out there who say "but my test was negative, so I don't need to go and get that unpleasant colonoscopy, I'm not going to do the low-dose CT scanning even though I'm an ex-smoker, and I'm certainly not going to go get that mammogram" and until we know that that is a safe recommendation, that's a problem.

The second problem that occurs to me with this question is that if the overall incidence in men is one in two and women is one in three and you have a negative now, years later, sooner or later, a lot of those people will develop a cancer and how does one manage then the anger and the frustration of that? How do you manage the fear of a false positive, but how do you manage the inappropriate reassurance of a false negative?

DR. PATHAK: Thank you for that perspective, much appreciated.

Dr. Beaver.

DR. BEAVER: So I think absolutely you would want a histopathologic diagnosis in this screening space in order to do some sort of larger intervention. In terms of following the true negatives, I think the key point is what was already discussed, which is to make sure patients understand that -- you know, and as we understand the performance characteristics of the test, patients understand that this is not they're never getting cancer, they don't need to do their other screening tests, I think that will be very important to convey, if we get to that stage, either in labeling or in messaging.

DR. PATHAK: Okay.

And Dr. Oxnard.

DR. OXNARD: To echo Dr. McAneny's point that -- you know, how are these negatives going to be handled, we should think generally about cell-free DNA diagnostics. These types and those we use in advanced cancer, they are high specificity tests with impaired sensitivity and I expect similarly that in cancer detection they will be rolled out as high specificity tests with impaired sensitivity and that's dissimilar from most of our screening methods, colonoscopy, PSA, you know, these are high sensitivity tests. Maybe, you know, PSA is too high sensitivity perhaps with impaired specificity and so these are definitely rule-in tests, the ones that we have, and not rule-out tests.

So none of the tests, I think, are even designed to reassure that there is no cancer present. It's merely to state that there's no cancer detected and certainly, there are not any of the existing tests, per their available performance characteristics, that are intended to replace standard screening, only to complement standard screening particularly given the variable adoption of standard screening. And so this is a way of finding a signal and finding the bad signals, right?

I think an important thought about the negatives is that they may not need that much follow-up, you know, there may not be 10 or 20 years of follow-up because these

tests are more detecting risky cancers and so what effectively you might be saying is, is there is no high-risk/high-morbid cancer developed maybe over a shorter period of follow-up rather than requiring the longer period of follow-up to rule out lower-stage cancers which might be less likely to shed and might be less likely the kinds of cancer detected with this

I do hope that our understanding of biology can reassure, to Dr. Kramer's point, which is I don't think these tests are going to find the sort of deep part of the iceberg and finding neoplastic processes that are not really cancer, because the hope is that certainly some of the data would say so far that those processes aren't shedding DNA to the degree where you can detect them robustly with this. These are really hoping to find the higher-risk shedding neoplasms.

DR. PATHAK: Okay, thank you so much. I think there are a number of important issues brought up with that question. Now, moving on. In a prospective cohort study where you follow for a certain amount of time, is the EHR an acceptable method as a data collection modality?

Dr. Kramer.

DR. KRAMER: Yeah, let me address the issue. There is suspicion, at least in theory, circulating tumor DNA is going to identify only the faster-growing tumors. Most screening tests, that's certainly not the case where you have to deal with length, biased sampling, that is most screening tests pick up the slower growing rather than the faster ones because the fastest grow up in between the screening tests. And although it may be less true for circulating tumor DNA, I suspect we don't know that for a fact and we ought to be prepared for some surprises that often occur in the field of cancer screening.

Now, for EHR, is it acceptable as a data collection method? Well, I would say that it may serve as a complement, but there are serious problems with EHR currently. Number

one, most electronic health records are not designed to be research tools, they're a little bit more in managerial tools, building tools, and so forth that haven't been fully explored and adapted as research tools.

But that said, even if they were, could they substitute for other forms of evidence? And I'm skeptical. Remember that there are two forms -- there are two issues that can give you the wrong answer. One is statistical variability and the other is bias, and EHR is almost a guaranteed way of incorporating hard-wired bias into the decision and codifying it. Why is that? Because physicians, if they're doing their job well, are adapting what they do and their recommendations and advice to patients based on the patient that's sitting in front of them. All of that is incorporated into the decisions that are recorded in the EHR and there's no -- at least I'm not aware of any known way to statistically fully correct for the hard-wired bias that, by design, gets into the EHR.

And as a matter of fact, there was a paper earlier this -- last year in the *JCO* that looked at the difference in estimates of outcomes of two interventions in oncology that were being compared by observational study versus randomized trials and there was no statistical significant correlation between the two outcomes and, as you might imagine, in most cases -- well, in about 40% of the cases the point estimate from the observational trial, observational study, didn't even fall within the 95% confidence intervals of the randomized controlled trial and as you might also guess, it was more frequently an overestimate on the positive side than the negative side. So I would say EHR is good as a follow-up to a proven test with a net benefit and as a complement, but we're not going to get a free lunch out of EHR.

DR. PATHAK: Thank you so much.

Dr. McAneny.

DR. McANENY: For sure, the answer is not yet. If anyone's ever tried to correct an

error in EMR, you will know what I'm talking about because we can't even get the medication list properly done because it varies from EMR to EMR. If something goes on there, you can't get rid of it. There's also a tendency to want to create documents, care plans and survivorship plans and other things like this in EMRs rather than going back to source documents and that can generate a lot of downstream errors.

And I agree that they're designed to be glorified billing machines that add a little clinical data, but I think it's possible to move this forward. But I think the way to do this is to look at the source documents, figure out how to extract that data, require that people who are doing tests like this also create them so they can be electronically embedded into the EMR and not go through the process of human error-induced, you know, clicking the right box to take the piece of paper that comes from a pathologist, is transcribed on a sticky note and then is put in the EMR in the right box or the wrong box.

So I think that once we figure out how to make that interoperability work and create from the EMRs data warehouses, cancer link type evidence, be able to pull that data, then I think it becomes a very powerful tool and I think if it's done right, it will allow us to use more real-world evidence, more ability to have people who do not have the wherewithal to travel for clinical trials, etc., to be able to pull that data forward and make it more valuable. So the answer is not yet.

DR. PATHAK: Okay.

Now Dr. Oxnard.

DR. OXNARD: I mean, all great points. I guess EHR is only as good as what you ask it to do, I mean, which is to say that if you ask too much of it, it can't pull it off. As a data collection method, merely as stated, it seems like this is kind of a de minimis thing that maybe we can ask the good EHRs to do and that's probably, you know, systems that use the unified EHR networks that are assembled, you know, but there can be variability where

folks can move from one EHR to another and then you would have a missingness problem. But I agreed with the point that if you ask it to do more than it's able to do, you will test its capabilities but, as stated, I think it's -- you know, I think in today's world we should be able to get data collection from EHRs if done right.

DR. PATHAK: And Dr. Beaver.

DR. BEAVER: So as you know, FDA is, of course, very interested in utilizing real-world data and thinking about how we can leverage some of the electronic health records. I think, in the case of this type of trial and given already the complicated nature of this and the -- perhaps the number of patients and also the fact that pathology reports are unstructured data, the radiology reports are unstructured data, I'm just trying to think of how that would play into this type of trial and I think, you know, right now I'm not sure it would be appropriate.

DR. PATHAK: Okay, thank you for the feedback on that question. Now moving on to the next question, it's kind of a multi-part question. Do you agree that PPV should be evaluated per cancer type? Is there any value in looking at overall PPV for the assay? Because the person coming in for the test is being simultaneously tested for multiple cancers. And is there any rationale for looking at overall device performance, like the performance metrics for diagnostic accuracy across multiple cancer types?

So Dr. Oxnard, could you kick this question off?

DR. OXNARD: This references some of the discussion from the earlier session, but I didn't get to see the exact reference that was being discussed. But, you know, if such statistical methods exist, this -- the idea of doing an overall PPV analysis is attractive. In the end, that is the value proposition, it's a detection of cancer in general, hopefully morbid cancer, and so -- and of course, you have more statistical power if you attempt an overall PPV analysis.

You know, as we found in some of the existing data from these devices, that as you start slicing the pie into smaller pieces you lose an ability to see a signal. I appreciate that this is different than the historical paradigm in development of screening tests, but I think now is the time to move towards a paradigm where you gain more statistical power and really, you more fully leverage the putative clinical usefulness or value for a patient by sort of collapsing into an overall PPV analysis, so I think that would be compelling.

DR. PATHAK: Thank you, Dr. Oxnard.

Dr. Kramer.

DR. KRAMER: So yes, I agree that the positive predictive value should be evaluated per cancer type, but we have to have the recognition that at least early data is going to inflate the positive predictive value, and why is that? Because the test may detect something that has nothing to do with the actual diagnosed cancer. Because imaging tests may show other incidentalomas, the diagnostic or the screening test may have missed the actual cancer and nevertheless you were fooled or misled into thinking that the positive predictive value was good. Is there any value? I'm looking at overall positive predictive value, yes, but with the same caveats since we don't know the natural history sufficiently for many of the things that we'll see.

And I want to say that looking at overall device performance across multiple cancer types in order to enhance the power is a very good idea. As a matter of fact, I'm just thinking out loud now, but if you have a test that is designed to detect the most common malignancies that account for cancer death, then you may actually be able to launch a trial whose primary endpoint is cancer death or even ideally life expectancy, overall mortality, since you are testing for the very cancers that dominate death in 30-40% of the population. That, again, requires a randomized trial and long-term follow-up, but I think there is a rationale for doing it.

DR. PATHAK: Thank you, Dr. Kramer.

Dr. McAneny.

DR. McANENY: Well, speaking as someone who deals with this, would deal with this in a very practical setting, I actually disagree. I think that it's very important that the positive predictive value should be evaluated per cancer type. If we have a test that just says it's overwhelming positive and better than all of the alternatives in all cancer types, I think that's a great marketing tool but I don't think it's a very good clinical tool because if I have a very common and easy-to-find test for things like breast cancer or the more common types of cancers, then that will increase this overall predictive value significantly. But the value of these tests that I think will be there for real clinical work is to find the test -- the tumors for which there is no good screening test. I would love to have a good screening test of ovarian cancer or pancreatic cancer or biliary cancer and there is not one. So overall predictive value is not particularly valuable to me as a practicing oncologist. It's a marketing tool. The individual cancer type may be helpful because in populations you have to look at different incidences there and it's not just the incidence of the tumor in general, but it's different incidences in different populations and that's where I think this could be exceedingly valuable if it's allowed to develop properly.

For looking at overall device performance, I think that is an important function of the FDA to make sure that if they're telling me that they are getting a particular segment of DNA that suggests a tumor, then I want to know that it's real and that it's repeatable and verifiable and I think that's exceedingly important, but I think that its importance will really vary among the cancer type and also the populations. We know, for example, that in black males I am much more aggressive about following up abnormal PSAs than I am in white males or older white males, in particular. In my own practice we see huge quantities of biliary cancer among my Navajo population. So we need to be able to not only have the

positive predictive value for cancer type, but population related to cancer type to make it a truly meaningful tool.

DR. PATHAK: Thank you so much, Dr. McAneny.

Dr. Kramer.

DR. KRAMER: I obviously wasn't very clear. My answer to the first question, do you agree that PPV should be evaluated per cancer type was absolutely yes.

DR. McANENY: Oh.

DR. KRAMER: No doubt about it. I just want to clarify that. The second question was, is there any value in looking at overall PPV, and I think that there's potential value to it and we have the same answer to a rationale for looking at overall device performance.

DR. PATHAK: Dr. Beaver.

DR. BEAVER: So I agree that definitely there's a value for PPV per cancer type, it should be evaluated per cancer type. The value is particularly great in those cancer types that don't have good screening tests right now and I think, in terms of value in looking at the overall PPV for the assay, I think there could be, I think particularly if it is shown to reduce overall mortality, as was mentioned, and then similarly looking at overall device performance across the multiple cancer types.

DR. PATHAK: Thank you so much, Dr. Beaver.

DR. OXNARD: Can I make a comment from -- if I may?

DR. PATHAK: Oh, sorry. Yes, Dr. Oxnard, what's your input on this?

DR. OXNARD: Thanks so much. I want to think about one other performance statistic, which is sensitivity, right? So I think sensitivity, it's very important to evaluate the cross-cancer types -- I mean, sorry, for each cancer type individually, right, because as you said, you know, we have great sensitivity per cancer but you're finding cancers that, I don't know, patients aren't interested in or worried about, you're right, that would falsely elevate

your sensitivity and it might play into a marketing campaign, as was said, right? So for sensitivity, it's particularly important to evaluate per cancer type to make sure there's sensitivity for those overlooked cancers, which is the theme that many folks said.

I just want to suggest that for PPV, it's relatively less important to do per cancer. It might be also important, but it may not be as important, right? PPV is more when you're the clinician getting the result back, you know, how often is that result, you know, successfully moving you towards a diagnosis? And I don't know, as a clinician who handles positive results, I don't think about my PPV for EGFR or ALK or KRAS, I don't really think of them individually. For krutinchen (ph.) genotyping techniques I think about, you know, when I get a result back, is the gestalt an actionable positive result. So I don't dispute that there's value in an analysis per cancer type, but I see greater relative value for analyzing sensitivity per cancer type and potentially greater relative value for analyzing PPV, a gestalt across cancer types but, you know, clearly there's some debate on that.

DR. PATHAK: Yeah, thank you so much, Dr. Oxnard.

Now moving on, this is sort of a complex question, but what are the challenges of validating a test accuracy for detecting multiple cancer types, including low-prevalence cancers? The size of even a large prospective cohort study will be unlikely to capture sufficient low-prevalence cancer for precise estimates of performance characteristics. How do we deal with this? Are there different study designs, different enrichment techniques that we could consider?

Dr. Kramer.

DR. KRAMER: Yeah, it's another immense challenge. Some of this goes back to what you heard from Dr. Srivastava a little bit early on. There are certain things you do for testing analytic validity and there, the best way to do that, that will meaningfully translate into whether or not you should launch a large prospective study are the characteristics of

the developmental stages in your in vitro testing. So there we don't often have this, but the ideal setting is a very, very large cohort in which you can do a nested case-control study where you actually know what the outcomes subsequently were and in a population that's not being actively screened for the target condition because if they are, that perturbs the performance of the test.

So ideally you would have a very, very large prior trial. We had this traditionally in the PLCO trial where we had samples collected and then we can look at cancers that subsequently developed, were not targets of the PLCO trial, and nevertheless provide us a signal ahead of time. That's the most definitive way to develop justification for a definitive clinical utility trial. That takes care often even of low-prevalence cancers if the cohort is large enough but otherwise, the precision is far too low to justify doing a large trial for those target cancers.

DR. PATHAK: So basically a case-cohort trial?

DR. KRAMER: Yeah. Yes, a nested case-control study in a cohort.

DR. PATHAK: Okay.

Dr. McAneny.

DR. McANENY: Well, I understand there will be a panel later discussing what the proper study process should be and they are far more expert than I on that. However, I have great concern that if we try to embark on a large prospective cohort study of any type, by the time we get any meaningful results, years and years will have gone by. I think we have a better process in place for this, we have multiple excellent academic institutions and physicians who are working on developing these tests and they may make them very specific, and the tests may be developed with a specific tumor type or a specific population in mind, and they will be able to come up with the appropriate -- not just the test for finding the DNA, but the test and other genetic material, but the algorithm by which they

use all these various biomarkers to be able to tell us the likelihood of a given cancer.

I think the most important role for the FDA here is to not be a bottleneck that requires that the science wait for the large prospective study, but to allow this to emerge from the bottom up and let the institutions come up with their particular tests and their evidence, and then the FDA's role is to figure out whether or not they can -- they are truly measuring what they say they're measuring and that their algorithm is verifiable.

DR. PATHAK: Dr. Beaver.

DR. BEAVER: So, in thinking about trying to capture low-prevalence cancers in a large cohort, I keep coming back to thinking if that's really the right approach or if instead the population that's chosen for the trial should be that which is a high-risk population. I think earlier it was mentioned, the potential to do sort of a staged-type design. You know, obviously if the population picked is a higher-risk population, you'll be getting the results, the events will be occurring more frequently and so then, you know, you could get -- gain information, perhaps, in that stage-type approach.

DR. PATHAK: Thank you so much, Dr. Beaver.

Dr. Oxnard.

DR. OXNARD: The only solution I potentially see here is through lumping diagnoses, right, and that there will be, of course, adequate numbers to make claims about individual high-prevalence cancers -- lung cancer, breast cancer -- but these other rarer cancers, they are important, right, they add up to a meaningful population of cancer patients. No individual test is focused on, you know, developing performance for X or Y rare cancer because the individual on a market is too small and the feasibility of studying these individuals is too small, and so the risk if we overlook them is that there is never a test that's able to be evaluated for them.

And so I hope that in cancers that more prevalent, we certainly can evaluate their

performance, certainly their sensitivity and specificity individually, but in cancers that are rare or there could be some ability to, in a pre-specified way, describe them as a community of rare cancers and therefore allow performance in this group of cancers in general.

DR. PATHAK: Thank you, Dr. Oxnard.

Now I'm going to hand it off -- Dr. Kramer has additional summary.

DR. KRAMER: So I think we're in agreement that you first start with the cheapest way to acquire preliminary information, but I want to emphasize preliminary, it doesn't take you farther, then, looking for a very large cohort. But case-control studies in individual medical centers or even consortia are limited because they are not prospective, but they give you an indication of whether it's worth it at all to take it to the next step. I agree with you, also, you don't build a whole new cohort which will take you a quarter century to get there. It's important to have bio-repositories for cohorts that are being collected anyway because you're asking a whole series of questions like the PLCO. When I was in the Cancer Institute, we wouldn't even allow for specimens to be sent out unless there was at least a preliminary indication from case-control studies that you describe that would justify taking a very, very and exceedingly precious commodity. And then if it's proven in a prospective cohort, I think then is the time to test the clinical utility.

DR. PATHAK: Dr. McAneny.

DR. McANENY: I agree with many of the things that you just said. I think that it's important to start with a population that has a higher risk. I can see the utility first being in that population of people that we know have high risk, Lynch syndrome, BRCA syndrome, etc., so that we can find out how good those tests are and then move it. But when you start talking about using a test, a screening test, I think that's going to have to be an iterative process and you start with something that can actually make that diagnosis accurately and then you start to look at large populations of people and that's, I think, where if we involve

the EMR to the proper area, if we start to use real-world evidence and databases that are being formed and data warehouses to then look and see not only did we successfully screen asymptomatic populations but we were able to screen them in such a way that we found things that we could intervene on to improve their life or prevent them from having morbidity from the tumor.

And then I hate to bring it up, but I think it's also important, though not the purview of the FDA, to look at whether the economics of not only do you do all the workup, the question we'll have to ask at that point is do you do a lot of evaluation of everyone with a positive test and does that break the bank in terms of getting enough imaging and repeat imaging and repeat testing, and are we successfully preventing people from dying of various cancers. It will be a long-term iterative process and I still think that our academic institutions are well positioned and our laboratory companies are well positioned to start this process up and make -- and give them the freedom to do the evaluations that are needed before we tamp things down with a lot of regulatory interventions.

DR. PATHAK: Thank you, Dr. McAneny.

Now I'm going to pass the torch to Dr. Jha, who will continue the next couple questions.

DR. JHA: So let's see if I can be as disruptive as NGS technology is. So my questions are, basically, is comparison to recommended screening method in the same patients are needed and what are the added value of the new tests? So let me begin by suggesting that, you know, all the recommended screening tests are not perfect. I mean, compliance is an issue. They are also, for example, LDCT, which is for 30 pack-year smokers, is missing adenocarcinoma, which is most common in nonsmokers.

This is a multi-cancer test and has added value, for example, diagnosis of cancers which are missed, diagnosis of other cancers which are not recommended for screening,

and can result in the change in the management for the benefit of the patient. Given all this and given what we have heard about the technological challenges with NGS, do we have to compare this test with the recommended screening methods in the same patient?

Dr. Kramer.

DR. KRAMER: I'm not sure I fully understood the nature of the question, but if there's -- I hope this addresses your question, you know, if it doesn't, if there's an existing proven screening test that has been shown to decrease cause-specific mortality in a definitive study, then you've created an extremely high bar for anything else that comes along including circulating tumor DNA. So you have to ask the question does it add to what you already have, is there incremental value, and we would design the trial accordingly unless you really want to replace the existing technology with circulating tumor DNA. If you have no accepted test, then of course there's nothing to compare it to and the bar may not be quite so high. I don't know if I answered your question, but there will be different study designs depending on whether you want to do it in conjunction with an existing test.

That said, single-arm trials are not definitive in showing the net benefit because if everybody gets everything, you can almost bet that the more tests you add, the sensitivity will increase. Sensitivity is a mix of good sensitivity that is identifying life-threatening cancers for which you may be able to intervene and do the patient good and over-diagnosis, tumors that never needed to be diagnosed in the first place, a single-arm trial is incapable of separating those two forms of increased sensitivity.

DR. McANENY: It is a difficult question, but I think it has several answers. One is if the new test is more effective than the existing screening tests, then I think it would deserve to be done instead of them. If it is not as effective, then you're going to have to figure out the position of this, do you use it only for those people who refuse to have the colonoscopy or refuse to have a mammogram, etc., and try to do it as something that might

be more acceptable to the population? The third instance would be if it is significantly cheaper than doing all of the other interventions. If doing the blood test is easier than doing a mammogram at whichever interval people feel is appropriate given their genetic makeup, etc., then I think that that has utility.

I think when we look at colon cancer, for example, and the Cologuard example that you gave before, a lot of the value in clinical practice of Cologuard is for those people who absolutely refuse and will not under any circumstances have their colonoscopy done even though you argue with them incessantly. And if that's the case, then you do the Cologuard and it's positive, you at least have another argument in the process of convincing a patient that they need to do what needs to be done, so you have to figure out where to put the new blood test in that process.

DR. BEAVER: So I think if you're trying to get at real clinical benefit in any of those scenarios, it will always be helpful to have a comparison arm, to have a comparator either to show additive value, certainly for a replacement-type value design, I think those are all -- those are all necessary. It's interesting to think about one of the other potential values, of course, being what was just mentioned, and that you get more people to undergo the screening and I think the flip side is what we talked about earlier, you don't want to necessarily reassure people for a negative test before you know that that's really a negative.

DR. JHA: Dr. Oxnard.

DR. OXNARD: Yeah, it's an important dilemma and CT screening is, of course, the best example because, while we now have data showing its value and certainly as a lung cancer doc I'm a proponent, adoption has been low and the reasons for that are varied, and some debate is it infrastructure, some debate is it stigma. And so how do we account for screening tests which are shown to have efficacy but are not adopted?

To some extent I think about tumor genotyping in advanced cancer as a good example. Liquid biopsies aren't claiming to be better than tumor genotyping, they're just more available, more feasible, and therefore they serve as a complement because it's so often the case that tumor genotyping is incomplete, unavailable, and therefore all you have is this blood test.

And so similarly, if you consider settings where screening is recommended but not followed through, or in discussions between a patient and a primary care provider it's decided that the screening test, you know, is felt to be too onerous by the patient and they choose not to proceed, there still may be supplemental value of a blood test. And so it would be attractive to think about more effectiveness approaches where a blood-based cancer screening is added to usual practices and those practices include some individuals who are enthusiastic about cancer screening and other individuals or communities where some cancer screening tests have been adopted less. And my wife's a primary care physician and I can presume that getting a blood test which you can do in clinic, on the way out, has an attractive feasibility assuming it's safe and assuming it's effective. It certainly has attractive feasibility as compared to referring an individual for a screening test where the follow-through is, to some extent, out of your hands. And so I think doing a strict comparison to all existing screening methods might not fully leverage the attractive feasibility and accessibility of a blood test.

Not all of those comparator methods have been FDA approved, right? Certainly there are some screening methods that we adopt more enthusiastically and others where there's more debate and that plays out, I think, in the real world, but not all individuals embark upon those screening tests. But still, those individuals might benefit from a blood test added.

DR. JHA: Let's move to the next question. So we know from existing screening tests,

for example, in mammography, that there's a significant amount of pre-cancer. For example, in mammography we pick DCIS, LCIS, atypical ductal hyperplasia. If you do colonoscopy, we pick up advanced adenoma even though it's recommended serrated polyps. Likewise, in LDCT we pick quite a lot of benign lesions which we have to follow. The question is will these precancerous lesions be detected by a multi-cancer test and if not, what is the significance, does it matter that we are not picking it up?

DR. KRAMER: So I think it's likely that circulating tumor DNA will identify more preneoplastic lesions because they're going to be -- as far as we know, they are clones that are forming early on in the pathogenesis of cancer and to the extent that they can be shed into the bloodstream, I think we will pick them up. At least for some cases we will when there's access to the circulating blood. What's the impact of not identifying these lesions? I don't know, there might be a strong benefit in not identifying the lesions, but -- in many cases, but in others there may be a benefit to identifying them if most of them are going to go on to progress to a lethal cancer and you can catch them in time. Again, this is another one of those famous double-edged swords in the field of screening and again, only a direct comparison can let you know what the net benefits and harms are in order to answer whether there's a net benefit or a net harm to identify all of those lesions.

DR. McANENY: I agree with the comment that we don't know yet. The test may identify some of these and it varies from tumor type to tumor type and whether or not the intervention that you can do, if you take out the advanced adenoma, you're likely preventing a colon cancer. If DCIS is not so clear, whether that is destined to become an invasive malignancy or not. We know that with prostate cancer that there are a lot of people who will live their entire lifespan with low-grade prostate cancers that should never be intervened upon because you'll do more harm than good.

So it really is too broad a question. It has to be narrowed down to the specific tumor

types and it leads us back to the question of if we have something that gives a positive blood test and says you likely have a cancer somewhere, the other impact of not identifying that is that it will drive patients crazy, they will have significant amounts of anxiety because my test said that there was something positive someplace, you found that benign thyroid adenoma but was that really what you were looking for? What else did we miss, what else is going to be there? We will create a large contingent of the worried well who are going to be in physicians' offices and demanding interventions to look for possible phantoms or possible things that are better left unmanaged.

DR. BEAVER: So I think these certainly could be identified by a multi-cancer test. What we know now is that it's more difficult to detect lower-stage disease in general, so it may be less of an issue. But as we are able to bring the limit of detection down and the tests get better, certainly they may be identified and I think, as was mentioned before, we don't know the impact of not identifying these lesions and I think that is what hopefully these types of clinical trials that will be supporting these tests could demonstrate.

DR. JHA: So since we are short of time, let me jump into our last --

DR. OXNARD: Please, I'm sorry, can I speak quickly? My apologies, I need to reference the data.

DR. JHA: Sure.

DR. OXNARD: We do have data from ESMO most recently, at least from our experience with the GRAIL assay, which is that the test does not detect these, that ctDNA is shed by advanced cancers, more advanced cases, by Stage 3 and 2 cancers; some, by Stage 1 cancers, less, and among Stage 1 cancers it's shed less by adenocarcinomas and more by squamous lung cancers, shed less by ER positive and more by HER2 -- or by triple negative.

And so it is definitely detection of shed is related to aggressiveness and these tests are not built to identify these pre-cancers which makes you wonder so what are they doing,

they're attempting to find cancers that need treatment and in a world where our treatments are getting better to detect those cancers when asymptomatic, to get them to treatments. But this is different than colonoscopy, which is really trying to find advanced adenomas before they become cancer. These blood tests are more likely finding cancers that are already invasive and are shedding DNA based upon our data.

DR. JHA: Thank you, Dr. Oxnard.

So my last question is, since it's a multi-cancer test and the screening test is organ specific, is there a -- does the term average-risk population apply in these kind of tests? Is it moot now? And further than that, what constitutes the intended use population? Should we include high-risk patients, average-risk patients, low-risk patients, symptomatic patients, all pooling into one and then conducting a trial?

Dr. Oxnard, you can go first.

DR. OXNARD: My feeling is that the greater your pretest probability, the better your PPV is going to be. Even if you push your specificity above 99%, you know, that low risk of false positives will be amplified, of course, if your pretest probability is too low. And so I think it makes perfect sense to offer a test like this to the kinds of folks you show on the page. I certainly wouldn't exclude them, I think they're exactly the kind of individuals most likely to benefit and indeed, individuals with risk might be more likely to be interested in a test like this.

And being able to account for potential false positives in individuals with these kinds of nonmalignant conditions is going to be very important and you want to know the potential false positive rate because -- so yes, the intended population is those at risk and I would certainly favor those with greater risk merely statistically because that increases the performance. So of course you don't want to zoom in on the highest highest-risk zebras, the -- you know, rare populations with really high risk aren't really meaningful on a

population scale and so folks with risk factors, I think, are an intuitive population to offer a test to and study.

DR. KRAMER: So we already have a spectrum of risks among people that we screen that ranges from nothing more than "how old are you, if you've turned 50 now it's time to have the test," otherwise it's the general population, to people who are less crudely stratified like "do you smoke and if so, how much," that's a higher-risk population, to cirrhosis, which is even higher and much of our efforts that are in liver cancer screening, I assume, are going to be focused initially at least on people with underlying cirrhotic disease or at least some prognosis, all the way to the genetically marked people that have the very, very highest risk, and there is no doubt that the PPV is going to be better and better and better across that spectrum. The one thing we need to be wary of is directly extrapolating the findings from the highest-risk populations to average populations, they have to be tested separately. And the other thing we have to keep in mind is there's a tension between patient health and public health. It works just like blood pressure, if you restrict your blood pressure treatment, antihypertensive treatment, to the people who have blood pressures of 190/100, then you're going to get the biggest magnitude of effect but you're going to affect the smallest number of people and you will prevent a relatively small proportion of hypertensive-related strokes and myocardial infarctions. And so we always have to figure out ahead of time what is the ideal population, what's the risk within that population, but never forget, don't extrapolate directly from highest risk to low risk.

DR. McANENY: I think those points are very well taken and I would certainly start with this population and anyone who is deemed to be higher risk because they have the most to gain and the least to lose by this particular intervention. I think, then, you're right in generalizing it to the population at large becomes actually a social issue and an economic issue, as well, that if you're going to do -- particularly if this turns out to be an expensive

test, to do this for the entire population every 6 months, every year, who knows how often it will be advantageous, it could be a significant problem.

The other point to consider here is that as we get more precise in our recommendations for individual patients based on family history, known genomic alterations, social risk factors, smoking use, exposure to pollutants, etc., that we may be able to, with data science and using the type of real-world evidence that's becoming more prevalent, to be able to pinpoint for an individual patient their optimal screening process and if we can get to that, then we can use these tests in a more targeted manner. No pun intended.

DR. BEAVER: Yeah, I agree, I think I certainly wouldn't exclude these patients. I think it makes sense to include them and in addition, think about some of the genetic risk, patients with various genetic risks for inclusion, realizing also the points about the population basis, that extending this more broadly will allow for that. But I think certainly initially it may make more sense to go for a high-risk population where the risk-benefit calculus may be more well balanced.

DR. SEIDMAN: Okay, let's move on to the next question. So we've touched on test performance characteristics multiple times this morning, but one cannot calculate performance characteristics of a test such as sensitivity, specificity, PPV and others, without having an unambiguous definition of a successful test or a true positive. Should only curable and treatable cancers, those for which there is a likely clinical benefit, be considered a true positive for the test? In any large trial, advanced-stage cancers are going to be identified. What stages and histopathologic types count as true positives or a success of the test?

You know, if a Stage 4 lung cancer, for example, is identified, is that a successful test? Arguably, there might be some benefit with respect to end-of-life planning for such a

patient; however, the benefit is clearly more limited as compared to identifying a curable cancer. What about Stage 2 and 3 cancers, do we weigh these differently? Should a Stage 1 cancer be given more weight with respect to benefits than an advanced-stage cancer?

Back to precancers. Should these be considered test positives or negatives or do we have to define them site by site? How should these be weighed in terms of benefits?

So maybe we'll start with Geoff Oxnard.

DR. OXNARD: As a lung cancer doc, I want to remind the audience that we are, in fact, curing some Stage 4 lung cancers these days with immunotherapy, it's absolutely remarkable. We're not doing it reliably, but we're getting better and I have certainly handfuls of these patients in my clinic today and so certainly, our therapeutic approach to advanced cancer is rapidly evolving. Not only that, you know, I just met a patient who presented too sick for any treatment and no question that this patient, if they had presented with adequate performance tests with no symptoms, for example, even their advanced cancer could've been treated better and, I hope, could have had a better outcome. So I would not -- I don't think we can throw up our hands at advanced cancer, we certainly can do a better job treating both more advanced cancers and less advanced cancers in a world where our treatments are getting better. It does, you know, require getting them to treatment. And weighting them is a fascinating idea, I think that could be considered in a pre-specified fashion. I'm not sure how exactly to do that calculus.

But to your final point about carcinomas in situ, you know, I think that that could be pre-specified and certainly -- and was pre-specified in the initial example presented at the beginning of the day. I think that, to some extent, we would love a test that can find invasive cancers but not find precancers which could be a higher risk for over-diagnosis. And so if we can find a test that has the sweet spot where it finds cancers that need treatment but doesn't find the excess of incidental over-diagnosis cancers that may or may

not need treatment, that could be sort of a interesting compromise and so -- but your flexibility on this is a great point, which is that this could be pre-specified with any test. My only strong feeling is that a Stage 4 cancer is very much a treatable cancer these days and I hope would be considered, you know, provided it's found at the time when it gets treatment.

DR. SEIDMAN: Dr. Kramer.

DR. KRAMER: So I will just point out that there's a limitation to the argument that we ought to be screening because there's more and more effective therapy. The irony of more and more effective therapy is that screening in some cases may become less important. We don't screen for testicular cancer, it's a simple exam, but it has attendant harms, but at almost every stage we can cure it. That's one extreme. So we feel like there is little additional benefit to be given by a screening test even though over the years we've evolved into highly effective therapy, so it's almost like a U-shaped curve. At the other extreme, cervical cancer, cancer is no longer even the target at all. There we have picked -- basically, we want to eliminate the pre-invasive lesions in order to prevent subsequent sorts -- can it be a disease-by-disease biology and test-by-test decision? Carcinomas in situ, the second question is are DCIS and HSIL and advanced adenomas test positives or negatives? Well, to definitively answer that, you need to know the natural history of everything that you are detecting or whether you are increasing over-diagnosis and again, sadly, we often don't know what the natural history is of detecting more and more and earlier and earlier cancers. Ideally, and we almost never have perfect knowledge on this, the target condition is an early stage lesion that is not over-diagnosed and in which, when we intervene we change the outcome, so therefore not simply curable cancers because, by definition, over-diagnosis is curing people that never needed to be cured in the first place.

DR. SEIDMAN: Dr. McAneny.

DR. McANENY: I think that this shows that we've moved away from the traditional staging process because I agree with what Dr. Oxnard said, that I would look at as a successful test is finding me a patient where I can intervene and they still have adequate performance data, Stage 4 or Stage 1, where I can either achieve a benefit -- if that benefit is cure, so much the better. If that benefit is managing that patient so that they get some process, some time of good quality life, that would be adequate. If you can take, say, a breast cancer patient and they're Stage 2 but they are Stage 2 with one positive node or they're Stage 2 with 20 positive nodes, those are entirely different patients and I'd much rather find them at the one-node positive stage. So for the others, it depends on the in situ cancers, it depends on their likelihood of becoming malignant, which we're still not very good at predicting in many situations.

DR. SEIDMAN: Dr. Beaver.

DR. BEAVER: The definition of a successful test is easier if you have a single disease that you're -- a single cancer that you're looking for. Much harder, I think, across cancers. But in general, I think the test would be successful if you can demonstrate it changes the natural history of the specific cancer or cancers, meaning that that detection is actually manifesting benefit for the patient and the natural history is the same.

DR. SEIDMAN: Dr. Kramer.

DR. BEAVER: You're finding it earlier.

DR. KRAMER: Let me go back to something that Dan Hayes said, which is a critical point, and that is the ultimate goal of all of this is to either improve life expectancy or to help people live better quality of life. One can't simply say that treatment has been improving for advanced stages of disease, therefore detecting it is a benefit, because most people in this audience realize that a lot of our therapeutic armamentarium has some downsides, and so it's not an automatic no-brainer that someone who was diagnosed when

asymptomatic, even if it is -- you know, it's beyond local disease, is going to have a better overall quality of life, that has to be tested definitively.

DR. SEIDMAN: Okay, I think we need to move on to the next question. So what are the metrics to determine benefit? What are the appropriate trial endpoints? There are multiple options here, there's stage shift, which we're down-staging, which is believed to reflect earlier detection; patient preference and patient-reported outcomes; compliance with recommended screening; non-inferiority to recommended screening; cancer-specific mortality, recognizing that things like mortality will take, you know, quite a bit longer to get than some of these other outcomes.

So if we could have Dr. Kramer.

DR. KRAMER: So first of all, again, repeating what Dan said, it's the health outcomes that are most important. Much of what is on that slide are intermediate endpoints, some of which have not really been validated as substituting for health outcomes. And it depends on how you define stage shift, it's certainly not enough to have simply an increase in early stage. Much more important is to see an increase in early stage and a subsequent meaningful decrease in late stage and that's my definition of true stage shift. There are cases where -- and health outcomes and most importantly, if you don't have a screening test, then I think you should focus on cancer cause-specific mortality.

But if you have a proven screening test, you can do something that we are doing in the TMIST trial and that is looking for a reduction of clinically important interval cancers and a reduction of late-stage disease when you have a similar enough technology which, in this case, is X-ray based, so something that has been proven to show that a reduction in late-stage disease can lead to a difference in mortality. So health outcomes, if you don't have a proven beneficial test in hand and if you do and the new technology is similar enough, you can start to rely on some reliable surrogate endpoints like reduction in late-stage disease.

DR. SEIDMAN: Okay. We're running out of time, I think, Dr. McAneny, if you could be brief.

DR. McANENY: I'll be brief. Down-staging is valuable if and only if it increases the ability to have that patient be cured. Having a test that finds people when they are more treatable than when they have very advanced symptomatic disease may be preferable. I agree, we don't treat all asymptomatic people with advanced disease. Having patient preference to get the test may be useful. Patient-reported outcomes are going to be variable because the increase in anxiety and concern about a positive test or insecurity with a negative one is going to persist. Compliance, I think, is very important, but I think it has to be non-inferior to the rest. And for the cancers where there are no screening modalities, we are desperate for some ability to find those patients and I would take as much lower level of sensitivity to have some degree of help in screening patients who I know are of high risk for uncommon cancers like familial pancreatic cancer, etc.

DR. SEIDMAN: Dr. Beaver.

DR. BEAVER: I agree with the last two speakers, so I don't have really much to add.

DR. SEIDMAN: Dr. Oxnard. Any comments, Dr. --

DR. OXNARD: I'll just focus on the sensitivity in other cancers, that, you know, it seems like even a test with compared sensitivity, an imperfect sensitivity still could be adopted if it's feasible and if the specificity is adequately high and the PPV is high. And so I think what level of sensitivity is needed, you know, I would be fine with a sensitivity of 50% if it's picking up cancers that otherwise wouldn't be found.

DR. SEIDMAN: Okay, I think we need to approach the end here and ask if there's anybody in the audience who would like to make comments. If so, please introduce yourself. No, can we have somebody else at the --

(Off microphone comment.)

DR. SEIDMAN: Yeah, go ahead.

MS. LANG: Hi, I'm Kathryn Lang from Guardant Health. It's a real-world evidence question. We addressed it. I think I could make a hundred comments about the panel, it's been amazing. But the real-world evidence issue, I'm an oncologist and epidemiologist and have been a real-world evidence researcher for 10 years, I was actually clinical lead on the IBRANCE study that got through CDER. That was hard.

Actually, I moved to diagnostics from pharma because actually this is where it should work, right? You are saying there is a ground truth against which our tests need to be examined. That ground truth is findable in real-world evidence. It's not as hard as actually trying to define what is an outcome, like a progression event in real-world evidence. We fundamentally need this sort of work to happen and people cannot go ahead and just create hundreds of thousands of cohort studies across the world, so we have to find this balance and I feel like I'm not hearing that when I hear real-world evidence, the EMR will never work. Actually, it's easier to do diagnostic research out of an EMR because you have a ground truth that you believe in. I'm not trying to infer.

DR. KRAMER: So I just heard what I was about to say and that is I don't think anyone on the panel said that EMR will never work. We were talking about the current state of art with EMR, which I think, number one, is administrative in nature in most places and number two, at least currently has built-in hard-wired bias, which is a good thing because physicians are making decisions that are in the best interest of their patients and that goes into the -- into the EMR. That's not to say we'll never get there.

MS. LANG: So just a follow-up. There is a lot of inherent bias in all CT design as well, and to pit one against the other to say that they need to correlate perfectly is not to see them in the same light. In fact, the IBRANCE study was based on very, very few patients across two different cohorts that held the signal and therefore, despite there being a lack of

size, we were able to show the signal and actually the obstruction technique, Immy Opineffe (ph.) has led the way here and she's here, she's at the FDA. I feel to get to where we want to go here, we have to embrace it and we have to accept that today's the day that we say we will overcome some of this EHR confidence issues.

DR. SEIDMAN: Okay, well --

DR. McANENY: I'd just like to comment really quickly on that.

DR. SEIDMAN: Well, if you could wrap it onto your concluding comments, we're running out of time.

DR. McANENY: Okay.

DR. SEIDMAN: And each panelist is going to have one last -- okay.

DR. McANENY: So in conclusion, and to wrap that in, I absolutely think that it's important to have real-world evidence because we cannot all be triathletes with a cancer, who have the economic ability to go to a tertiary center to have a clinical trial done, so I absolutely think that is going to be the future and should be. For this particular issue, what I would say in conclusion is the expansion of the science is remarkable and I think the role of the FDA and the other regulatory entities like CLIA and other -- and CMS, that have the ability to either accentuate this and help us roll this out or to act as a bottleneck and prevent things from occurring.

We have remarkable institutions that are able to develop this and to bring it here for appropriate regulatory, so my plea would be, and the plea from the AMA would be, to please allow these to develop, regulate them for safety, that they are actually doing what they tell us that they are doing and that they are not causing undue harm, which will be an iterative process that we have to look for, which is especially amenable to real-world evidence because we won't know what those harms are, good or bad or workups or not, until we have time to look at these. Thank you.

DR. SEIDMAN: Thank you. Thank you, Dr. McAneny.

Dr. Oxnard, if we could have your concluding comments in 1 minute.

DR. OXNARD: I would just emphasize that I think these tests are different than our existing screening paradigms, not just in their multi-cancerness, but that they are less likely, I think, to find precancers. This is not, you know, finding polyps at colonoscopy but rather finding potentially treatable invasive cancers that may be overlooked by standard screening or may be neglected by those who aren't proceeding with screening and they are potentially highly scalable. Certainly, I agree that the clinical benefit is not guaranteed. The clinical benefit, though, of finding these cancers is through the effort of figuring out how to interpret the results, which is not certain but is being sorted out and in doing so, we get treatable -- get cancers that need treatment to an advancing array of advancing therapies which potentially could then lead to downstream clinical benefit to cancer patients.

DR. SEIDMAN: Thank you, Dr. Oxnard.

Dr. Kramer, concluding comments?

DR. KRAMER: Sixty seconds.

DR. SEIDMAN: Yes.

DR. KRAMER: I think we've been talking about, all morning, the issue of signal versus noise. Every approach to a study has -- every one has noise and some have signal. An observational study tends to have inherently more noise than signal and that's why we need to be very careful if we depend exclusively on collected data with no internal comparators. And I think it bears reemphasizing that the signal we're looking for are health outcomes and in the field of screening, most screening tests -- and this might turn out to be the same with circulating tumor DNA -- are a close call and so we have to be prepared for study designs that are enabled to detect relatively small signals. Screening is a tough call sometimes.

DR. SEIDMAN: Thank you, Dr. Kramer.

Dr. Beaver.

DR. BEAVER: Sure. So I think to go to the real-world evidence comment and question, to kind of wrap that in, I think as with anything, FDA is really excited to hear your proposals and from the drug side, we always say, for the real-world evidence discussion, bring us the proposal, bring us the -- bring us the responses to why, you know, our concerns are, you know, incorrect or maybe misunderstood and we will think about different trial designs and ways to incorporate.

And I'm sure my colleagues at CDRH feel similarly, they'd be happy to look at that type of design that you bring. In terms of these screening tests in general, thinking about the multi-organ -- multi-cancer test versus a single cancer test, I think there's more inherent risks with the multi-cancer test and though there's also potentially greater benefit, particularly for the areas of cancer where we don't have good screening tests right now, and so I think all of that will need to be -- will need to be borne out in clinical trials demonstrating appropriate outcomes.

DR. SEIDMAN: Thank you, everyone. I believe that will end Session 2 and we're going to --

(Off microphone comment.)

DR. SEIDMAN: Yes, we're going to break for lunch and --

DR. ABUKHDEIR: Yes, we're going to break for lunch right now. Could you be back here before 1:35? We'll start promptly at 1:35 for the public comment session. And if I can ask the panelists to just sit tight for a second, I'll come over to you right now.

DR. SEIDMAN: Thanks to the panelists very much.

(Applause.)

(Whereupon, at 12:37 p.m. a lunch recess was taken.)

AFTERNOON SESSION

(1:35 p.m.)

DR. GALLAGHER: I'd like to welcome everyone back from lunch break. We're going to get started with the third session, the public comment session. My name is Pam Gallagher, I'm a scientific reviewer in the Division of Molecular Genetics and Pathology. And so we'll be hearing some public comments, the people that are speaking have pre-registered to give public comments and we kindly ask that the speakers limit their comments to approximately 5 minutes. So the first presentation we have is from Theo deVos from Epigenomics.

DR. deVOS: Thank you. It's been a really interesting morning, I really appreciate hearing all of these things. They don't sound like they've changed very much since 2007 when we first started interacting with the Agency and addressing this as a problem. As a bit of background -- is that the right one? So there is actually an FDA -- a PMA-approved cancer screening test based on cell-free DNA and that's the Epi proColon test, and that a project that I'm providing the context for, for the comments I'm about to make.

So I was involved in discovering, developing, going through regulatory and ultimately getting this test approved and now hoping to get coverage for it. It's a cell-free DNA and I would suggest actually that -- you know, we've been talking a lot about ctDNA and I would suggest broadening the concept and the language because the language matters to actually be inclusive of all nucleic acids and circulating rather than specifically tumor because there are tests and marker IDs out there that will probably fall under the same purview that are not necessarily tumor derived but could be sentinel markers. So that was my -- my first general comment is to broaden the language to cover all of that because the methodology and assessment would be similar.

Did I hit the wrong button? There we go.

So the approach that we took, and I think that's relevant to this, for a single cancer test based on a single marker was essentially a validation using what's called the probe design, which I think is an excellent design for making the assessments of sensitivity and specificity. We had the advantage, it was colorectal cancer, so there was a reference method of colonoscopy, as imperfect as that might be. There was predicate devices that you could compare to -- even though we're not a 510(k), you still want to look at what exists out there -- and there is a clear clinical workup for it. So it seems like this is the simplest, most realistic, easiest system to go through an approval process on but it has a lot of complexities and I think if the simple one is complex, the complex one, as we heard this morning, is going to be even more complex.

So this retrospective analysis of a blinded, prospectively collected sample set, it's the right approach, I think, to at least limit some of the cost and expense of going through, and what we discovered is that it holds very true for the actual performance when you go into subsequent trial designs and even now, as we're doing our post-approval and looking in the real world, this type of a design gives you an excellent assessment about sensitivity and specificity. It also allows you to estimate PPV and NPV or DLRs, if that's your approach.

And I had a second general comment on NPV. There was a lot of discussion about the importance of NPV this morning, but if you're dealing with a low-prevalent disease, even just random sampling of the population is going to give you an excellent NPV and so understanding that a 99.5% NPV is normal and that how are you going to measure, you know, what's better than that is kind of a tricky business.

So we used this design to establish sensitivity and specificity parameters and then we were asked to do a second methodology to assess comparative performance and I think the combination, then, of a modified case control that, as closely approximates prospective as possible without actually having to go prospective, is a really, really good approach to

augmenting your dataset. So in this case, the way we got to a proxy is we used screen-detected cancers as our cancer cases and then prospective samples as our prospective cases and what that allowed us to do is get a hundred cancers to get a good assessment. They were screen detected, so when you looked at the stage distribution it was nearly the same as in the prospective setting, not quite, and so that's a good sort of augmentation. Then I'll skip to -- I could talk forever. I'll skip to a couple of things that were challenging for us. One of them is, and I think this --

(Off microphone comment.)

DR. de VOS: Okay, all right. Thanks. If you want to talk, I'm happy to be here.

DR. GALLAGHER: Okay, so our second speaker will be John Sninsky, who's an independent consultant.

DR. SNINSKY: Thanks, Pam. Good afternoon, I'm John Sninsky and I'm speaking on behalf of my colleagues at CellMax Life, whom I consult for. I wanted to first thank the Agency for the opportunity to comment on the topic of this workshop, as well as applaud the Agency on their past leadership efforts. My comments will focus on colorectal cancer, CRC, which both has shared and distinct features relative to other cancers.

I wanted to first emphasize that the primary objective of CRC screening is removal of precancerous adenomas. As highlighted in the ACS 2018 multi-society task force, the main effector of CRC prevention is polypectomy. Widespread polypectomy has been a major driver of declining CRC incidence and mortality. Increased screening would further reduce CRC mortality; however, a large fraction of those who should have colonoscopies do not follow guidelines.

We have a new understanding of natural history of cancer. In brief, we know that first, rather than stage molecular sweeps, CRC may reflect a big bang of genetic alterations upon which selective pressures of the micro-environment act. Secondly, thanks to the

pioneering studies of Judah Folkman years ago, we now know that neoplastic cells recruit angiogenetic and lymphangiogentic vessels when they're less than a millimeter in size and therefore cells gain the capacity to disseminate much earlier than anticipated. And third, there's ongoing extracellular communication that conditions premetastatic niches earlier than we expected, so it's important to provide early detection.

Biomarker assays using somatic variants and cell-free DNA have insufficient positive predictive value to be maximally effective as a screening tool. However, biomarkers can sample various biological compartments and yield orthogonal information. The more orthologous the sample compartment, the more likely the biomarker information will be confirmatory and complementary rather than redundant. Promising orthologous content compartments include somatic variants and methylation patterns, circulating epithelial cells, proteins, as well as demographic factors such as patient age and gender.

Building multimodal tests is complex. For example, some biomarkers may be informative for different stages of disease. Challenges of multimodal tests include integration of biological, clinical, and statistical domain expertise for mode and feature selection, as well as feature engineering and model development while avoiding overfitting.

This slide reminds us of the significant confounding already mentioned today of CHIP, particularly in older patients. We must remain vigilant in the interpretation of these detected somatic variants, but we should also explore whether this chronic inflammatory condition may play a role in ongoing cancer development.

As depicted in the graph on this slide, we recommend integrating multimodal markers into multi-category probabilistic risk model scores instead of simple binary classification. In general, two approaches are frequently used in diagnostics. First, classification wherein the presence or absence of a disease condition is reported using metrics such as sensitivity and specificity and second, risk probability wherein advanced

modeling permits the stage of disease progression to be estimated using regression along with confidence intervals. This is similar in principle to the rich information obtained from colonoscopies which extends their prevalence value to inform subsequent surveillance and interval colonoscopies.

Margaret Pepe and Drew Watson pioneered the use of predictiveness curves that superimposed actionable thresholds on risk probability curves. Innovative study designs with appropriate power depend on the prevalence of each of the required clinically actionable outcomes with enrichment of lower prevalence outcomes which will benefit from agency submitted discussion to ensure safe, effective and timely patient access.

Machine learning will prove a powerful tool for cancer screening but requires fully transparent, interpretable models rather than black boxes. We recommend that simulated data much like OVA special controls that the Agency developed in the past be used to augment empiric data and we should plan on a fit-for-purpose staged introduction of machine-learning models as noted on this slide. We thank the Agency for their recent dynamic learning draft guidance.

In summary, my colleagues at CellMax and I encourage the FDA and our industry to:

- ensure that precursors to cancer as well as cancer are detected at high predictive values;
- integrate the new understanding of the natural history of cancer to develop screening assays;
- require multimodal assays to attain the requisite predictive values for screening;
- use predictiveness curves indicative of disease stage and future risk with clinically actionable results;
- understand that machine learning in cancer screening represents high stakes decisions as noted by Cynthia Rudin at Duke that requires special attention.

And then lastly:

- to consider study designs for the intended use indication informed by prevalence of clinically relevant outcomes rather than solely average cancer risk populations.

Thank you very much.

DR. GALLAGHER: Okay, so our third speaker will be Girish Putcha from Freenome.

DR. PUTCHA: Good afternoon and thank you for this opportunity to make some comments, at least some of which will echo those from this morning. So I am Girish Putcha, the chief medical officer and director of the clinical laboratory at Freenome. We were founded in 2014 with a mission to develop tools to prevent, diagnose, and treat diseases with an initial focus on colorectal cancer.

So first, there are over a hundred kinds of cancers but the USPSTF has an A or B recommendation in asymptomatic individuals for only four, A for cervical and colorectal and B for breast and lung. Indeed, the USPSTF specifically recommends against screening in asymptomatic individuals for certain cancers including, for example, ovarian and pancreatic, noting that there is no net benefit or that the harms outweigh the benefits.

You may have also noticed that the populations that are recommended for screening, even among the four for which screening is recommended, can be meaningfully different. So what exactly would be the intended use population for a "multi-cancer" test, let alone a "pan-cancer" one? And is there truly clinical utility for such a test even assuming that the test had the required performance characteristics?

Next, I submit to you that there are actually different performance requirements in terms of sensitivity and specificity based on the cancer that we're actually screening for because even the diagnostic let alone the therapeutic pathway for different cancers is meaningfully different. In other words, there is no one-size-fits-all requirement for sensitivity and specificity for a "multi-cancer" test. For example, a false positive result for a

CRC screening test results in a colonoscopy with its associated risks and cost, but I think we can all agree that these risks are meaningfully different than the risks of major abdominal surgery required for an ovarian cancer diagnosis. Stated simply, the benefit-risk calculus can be very different for different cancers and therefore I suggest to you that the optimal performance characteristics are, as well, with some requiring greater sensitivity at clinically appropriate specificity and others requiring very high specificity at a clinically acceptable sensitivity.

I'd also encourage us, a community, to think very carefully about how these novel tests fit into the existing patient care pathways. If, after taking one of these tests a patient still has to get another standard of care screening test let alone a non-standard of care one like PET-CT, does this really help or are we just adding risks and costs without providing incremental benefit and value?

Next, I'd like to switch gears and point out that in CRC just as in other cancers, multiple biological pathways contribute to tumor pathophysiology. Similarly, while tumor-derived signals are readily detectable in late-stage cancers, they can be at the limits of detection in early stage disease let alone in precancerous lesions, yielding essentially stochastic results.

Given the dual challenges of biological and temporal heterogeneity, we at Freenome believe that ctDNA or even cfDNA alone will be insufficient to solve the early detection problem and if sought to complement purely tumor-derived signals with those from other sources including the immune system, we -- therefore the Agency -- when creating guidances or standards, to not focus on only the technology such as NGS or the specific analyte such as cell-free DNA or circulating tumor DNA and instead develop approaches that will be broadly applicable.

Finally, I genuinely believe that the Agency could help us all improve how we

develop and validate cancer screening tests. Many of us know all too well the challenges, starting with the low prevalence and incidence of disease. For example, in DeeP-C nearly 13,000 patients were enrolled to get to approximately 10,000 evaluable samples to get to 65 cancers, which is clearly not sufficient for R&D purposes. This leads many of us to use post-diagnosis samples and case-control designs. These approaches obviously engender a variety of concerns from generalizability to the intended use population to the lack of the full disease spectrum to the use of controls who may only be negative by history or by self-report.

But I sincerely believe the Agency and others can help us with solutions to at least some of these problems. These could include shared public resources like the availability of samples from groups like the Early Detection Research Network, though sample volumes and matrices can be limited, to balancing pre- and postmarketing commitments in a way that safeguards patients but also rewards and protects innovators to developing best practice guidelines for the community with groups like the CLSI and others. Thank you very much.

(Applause.)

DR. GALLAGHER: Okay, so our next speaker will be Eric Fung from GRAIL.

DR. FUNG: Thank you for the opportunity to speak today at FDA's public workshop. I'm Dr. Eric Fung, Vice President for Clinical Development at GRAIL. We greatly appreciate FDA's efforts to better understand, with public input, the optimal regulatory approach for assessing circulating tumor DNA-based cancer tests being developed for the early detection of multiple types of cancer.

GRAIL is a healthcare company focused on alleviating the global burden of cancer by developing pioneering technology to detect and identify multiple deadly cancer types early. Why did we at GRAIL undertake this public health challenge? Cancer is projected to

become the world's leading killer and despite many advances in cancer care, we continue to diagnose cancer when it is too late because most deadly cancers have no effective early detection tools. Cancers detected when they are localized or more amenable to curative treatments and across all cancers, have a high -- of cancer-specific survival of approximately 90% compared to 20% for cancers found when regional or distant spread has occurred. However, cancers diagnosed when distant metastasis has occurred represent nearly half of cancer deaths within 5 years among persons aged 50 to 79.

Of the 1.8 million diagnosed cancers in the United States in 2020, we estimate that about 700,000 could be found by conventional screening, but over 1.1 million will go undetected until symptoms arise or they are found incidentally, most of them caught too late.

The five cancer types with screening recommendations represent less than 40% of the cancers in individuals over 50. This means that any individual who participates in an appropriate guideline-recommended screening regimen is not being screened for other potentially deadly cancer types. In fact, an individual seeking one of these single cancer screening tests is 2 to 80 times more likely to be diagnosed with a completely different cancer.

For all of these reasons, we believe that a multi-cancer early detection approach using a single blood test rather than multiple individual tests for each type of cancer can help address this enormous public health challenge. We think this is now scientifically possible due to a confluence of technological advances in genomic sequencing and artificial intelligence.

However, in order to achieve a significant benefit at the patient and population level, it is important that any multi-cancer tests have the following characteristics:

- The test should avoid over-diagnosis and be able to identify most types of cancer

and the majority of deadly cancer.

- It must optimize overall cancer detection while maintaining a very low false positive rate, resulting in a high positive predictive value.
- It should be able to localize a cancer to specific organs in order to efficiently direct a diagnostic workup.
- It should be supported by robust analytical and clinical validation at population scale, be intended for an elevated-risk population such as those over the age of 50 who have over 10 times higher incidence of cancer.
- The test should be simple and easy to use to maximize adherence.

And finally,

- It should serve as a complement, not a replacement, for conventional guideline-recommended screening.

In assessing the benefit-risk ratio, we recognize that a false positive result could lead to complications related to diagnostic follow-up procedures and/or to psychological harms which highlight the importance of a very low false positive rate. A false negative result could lead to patients foregoing guideline-recommended screening, and therefore it is critical that patients be provided with robust instructions that they should continue with guideline-recommended screening and that a negative result does not rule out the presence of cancer.

These risks, however, must be viewed in light of the potential benefits for individuals and at a population scale. When evaluating the benefits, we believe that assessment of the performance characteristics of multi-cancer tests should be placed in the context of the likelihood that a positive test result is indicative of cancer, that is the positive predictive value. The positive predictive value will be driven primarily by the specificity of the test.

Cell-free nucleic acid assays, if designed appropriately, have the ability to detect a

broad set of cancer types even if not every cancer type itself is detected at apparently high sensitivity. Therefore, from a public health viewpoint, the ability to detect all -- to report all detectable cancer types with high positive predictive value rather than relying on pre-specified per cancer type sensitivity thresholds maximizes benefit without introducing additional harms.

In the United States, nearly 11 million people are living with cancer and approximately 2,000 loved ones die each day. We believe that early detection technology is now available to reduce this cancer burden and we appreciate FDA's efforts to ensure that this paradigm shift in science, when done right, can translate into a major public health benefit.

Thank you again for the opportunity to present GRAIL's comments during this public workshop.

(Applause.)

DR. GALLAGHER: So next I'd like to introduce Anne Marie O'Broin Lennon from Johns Hopkins University.

DR. LENNON: I'd like to thank the Agency for the opportunity to speak, and my name is Anne Marie Lennon, I'm a Professor of Medicine, Surgery, Radiology and Oncology and the interim director in the Division of Gastroenterology and Hepatology at Johns Hopkins. I am here because I'm a physician, I've been a physician for over 20 years, and I'm actively involved in patient care. I run the multidisciplinary pancreatic cyst clinic at Johns Hopkins which sees over a thousand patients every year and I have a particular interest because of my involvement with patients in early cancer diagnosis, particularly for rare, devastating cancer such as pancreatic cancer where the overall survival is only 9%.

I want to thank the Agency for organizing this workshop which addresses a very important unmet need, that is the lack of a screening test for the majority of cancers. As

mentioned by our previous speaker, 1.8 million individuals will be diagnosed with cancer this year in the United States, of whom it is predicted that 600,000 will ultimately die from their disease. It is predicted over the next 10 years that cancer will become the number one cause of death in the United States. Patient survival is directly correlated with stage, that is how early can cancer be identified. We know that screening, such as colon cancer, is effective and has been proven to decrease cancer mortality by finding cancers earlier. However, currently, routine screening tests exist for a small number of cancers such as breast, cervical, lung, and colon cancer and no screening test exists for the majority of cancers.

Despite the recognition that screening tools which detect cancers early could dramatically improve patient outcomes, efforts to develop traditional image-based single ordinate screening modalities for diseases such as pancreatic or ovarian cancer have failed. The advent of blood-based multi-cancer screening tests that combine a very low false positive rate with the ability to find not one but many cancers with a single test is poised to finally solve this longstanding problem.

Because these tests are fundamentally different in how to detect cancers relative to any familiar precedent, making them available to patients will require a fundamental rethink of how benefit and risk are evaluated and managed by both assay developers and regulators.

The Agency has already shown its ability to perform a paradigm shift in regulation; for example, in cancer drug development where tissue agnostic drugs have been approved for cancer development based on the presence of molecular biomarkers. These novel therapies have already helped many patients with rare cancers who lack other treatment options. The Agency has already implemented the necessary regulatory innovation in these -- in the therapeutic contents and I applaud the Agency's willingness to look at similar

opportunities for multi-cancer screen tests.

It is also important for developers and physicians not to be blinded by the promise of novel tests and to ensure that we understand both the benefit and the risk before introducing them into routine practice. It is particularly important to understand the potential risks associated with over-diagnosis, subsequent testing, and the impact of introducing these tests on these current screening modalities. To do this responsibly, prospective interventional clinical trials will be required.

The potential to be able to screen for multiple cancers and the impact for our patients of diagnosing cancers such as pancreatic cancer earlier is huge. I want to thank the Agency for helping make these tests a reality.

(Applause.)

DR. GALLAGHER: Our next speaker will be Stacey Adam from FNIH.

(Off microphone comment.)

DR. GALLAGHER: Okay. Dana Connors from FNIH.

MR. CONNORS: Thanks for the opportunity to provide comment today. I'm replacing Stacey Adam. I'm Dana Connors, I work with the Biomarkers Consortium's cancer steering committee at the Foundation for the National Institutes of Health.

The FNIH is considering assembly of a collaborative community in the area of liquid biopsy. Our organization has already begun a coordination effort between one of our FNIH Biomarkers Consortium projects, the ctDNA Quality Control Materials Project, which is seeking to develop a set of nationally recognized standards to enable the publication or production of suitable quality control materials for widespread use in ctDNA testing to provide confidence and interpretation of ctDNA biomarker assay results, as well as other consortia in the liquid biopsy field. These include BloodPac, which developed a genomic data commons, of which I'm sure you're all aware, to aggregate liquid biopsy work, now

open for contribution, and they also developed minimum technical data elements for pre-analytical variables and patient context. Cancer ID is a consortium of 40 partners across 14 European countries working to evaluate liquid biopsy technologies and guide standardization of pre-analytical requirements, protocols, and results sharing, and will be replaced by the European Liquid Biopsy Society, or ELBS, which expects to provide a hub for liquid biopsy research in Europe with the key goal to translate liquid biopsy assays into clinical practice for the benefit of patients. We're also working with the International Society of Liquid Biopsy, or ISLB, whose main aim is to introduce recommendations to develop reliable and sustainable diagnostics and prognostic tools using liquid biopsies which will benefit patient health management and their wellness.

Friends of Cancer Research, you probably all know, has developed a framework for standardizing clinical ctDNA testing with their ctDNA pilot project and is soliciting retrospective and prospective studies to validate the hypothesis that ctDNA is reflective of outcomes in their CT monitor project. Friends is also building a framework for best practices to measure and report on TNB.

JMAC has joined our group, which is a Japanese industrial consortium that was established in 2007 to support biotechnology with international standardization activities. MDIC, the Medical Device Innovation Consortium, is developing somatic reference material samples in solid tissues and finally, the National Institute for Biological Standards and Control, or NIBSC, from the UK, is the WHO international laboratory for biological standards, which is producing the WHO first international standards for ctDNA, which is intended for the calibration of kits, assays, and secondary standards.

This coordination effort comprised of these organizations and foundations recognizes the importance of working towards the global use of liquid biopsy in oncology practice to support clinical decision making and regulatory considerations and seek to

promote it in their communities. The group recognizes that a preponderance of technologies and pursuits in the field of liquid biopsy has the potential to confound the field and obscure important progress and thus, the current participating organizations have joined together to share in their scope of work, discuss lessons learned, and to disseminate the tools and data they have developed as a coordinated effort. As an alliance, the partners have found strong value in the exercise of information exchange through in-person meetings and teleconferences and the dissemination of their collective efforts. The groups believe in educational outreach to the scientific community to publicize the formation of the collaboration and to make resources more broadly available is a value.

This coordination is open to additional groups, for your information, who wish to join and is working towards becoming a collaborative community that the FDA could join as a participating member. Thank you.

(Applause.)

DR. GALLAGHER: Thank you.

So our next presenter will be Jonathan Cohen, who's part of the Small Biotech Business Coalition.

MR. COHEN: Good afternoon. My name is Jonathan Cohen, I chair a diagnostics working group of an organization called the Small Biotechnology Business Coalition. We are advocating for -- or rather against over-regulation to the point that it would prevent small innovative test developers from being able to develop, validate, and ultimately commercialize their tests from throughout the United States.

We are concerned some of the perspectives offered by some panelists today could result in a de facto ban on the introduction of pan-cancer tests, multi-cancer tests in our lifetime. A number of people have suggested that the Agency insist that the benefits of tests be proven, definitively proven to outweigh harms before any test can be introduced

and the question I have this afternoon is, is that possible? Let's take PSA, for example, it's been well over what, about 40 years since PSA has been utilized here in the United States and there's still a debate, has that been proven, has PSA been proven the benefits of that test to outweigh the harms? We could have a panel of expert urologists, I suspect, here today and there would be a split of authority. There are studies from Europe that say that it does and studies from the United States that say it doesn't and vice versa. So is that an achievable goal?

Similarly, it's been argued that only randomized controlled trials are going to be sufficient before any test can be introduced. How much time and money does it take to undertake such a test? It's decades, hundreds of millions of dollars. Well, screening tests are not drugs, they don't generate -- you're not charging tens of thousands of dollars a month for the test, they're a few hundred dollars. When the economics of such a test are looked upon by a company and its investors, the answer will simply be don't do that, do something else.

Lastly, I just want to point out there's a very different paradigm. We in the West do very little screening, as has been discussed. The task force has given A's or B's to about four, four organ types, but that is not the case in the East. In Japan, Korea, Taiwan, mainland China, even parts of India and the former Soviet Union, people routinely go to what are called health check centers and spend the better part of the day being screened. All they do is screen, they screen with blood, they screen with imaging, and that approach is growing in popularity, it is not retracting. So the question I ask is do they know something in the East that we don't know in the West? Are we perhaps exaggerating the harms? And I leave you with that this afternoon. Thank you.

(Applause.)

DR. GALLAGHER: I'd like to introduce our next speaker, Bekrum Shawn from

AstraZeneca.

(No response.)

DR. GALLAGHER: Okay. So it looks like we're running a bit ahead of time, so if any -- oh, yes. I'm sorry about that. Kathryn Lang from Guardant is our final speaker of the session.

DR. LANG: I would have been terribly disappointed to be forgotten, so thank you. Thank you, my name is Kathryn Lang, I'm here representing Guardant Health, I'm vice president of Outcomes and Evidence.

Guardant Health is -- it's an 8-year-old company now that started to -- it's a leader in molecular diagnostics. We have a laboratory developed test called G360 which is focused on advanced cancer. We are therefore highly attuned to this world of multi-cancer testing and we recently, in October of last year, launched a 10,000-subject prospective clinical trial to evaluate our early diagnostic and colorectal cancer screening.

Patients come first, that's the reason we're here, it's the reason that Guardant does every single thing that it does and we wanted to reflect a little bit on what we've had to do and what the Agency is trying to do in general in the context of an organization that, in the advanced cancer space, is doing multi-cancer testing but in the early detection space is in the single organ testing. We believe very much in screening, that it is the "so what" of the test that is critically important and that we need to think about that "so what" from the perspective of the patient.

So three comments and observations to make from what we've heard today. In the first instance, the performance characteristics that are being discussed, sensitivity versus specificity, how do our patients recognize that? When you undergo a test as a patient, you expect something and we make all of these claims about PPV and NPV, but I think we have to think about what that means to a patient when they undergo testing, what do they

expect for it to find. Ultimately, that also draws back to clinical acceptability, but we do need to take that patient perspective into account and so is specificity really what a patient is expecting when they go in for multi-cancer screening, I think that's very important.

Also, from a performance characteristics perspective, we have to compare against as relevant as possible standard of care with the attendant understanding that bringing molecular diagnostics into screening, the paradigm is changing and we do not compare like for like when we compare 25-year-olds' tests against molecular diagnostics today and the kind of things that we need to look at in these tests perhaps need to be actually much more disruptive.

I come from hematology and in hematology for years we called acute myeloid leukemia with normal cytogenetics, we just called it AML and we didn't think much of it. Then we added in a molecular characteristic which then actually draw the complete new understanding of treatments and paradigms. We have to think in the same way when we're doing screening, what are we actually going to change in the disease characteristics and understanding? So simply to compare the old test against the new is not to do justice to what molecular diagnostics are bringing to the field.

I think the second thing is to think about the downstream clinical utility of these tests. What is the diagnostic odyssey that we send the patient on upon a positive signal? And I don't just speak about false positives, I'm talking about true positives, too. We have spent a very long time on picking, and admittedly in the non-malignant world, but in picking the relative harm done by CT pulmonary angiography when we are picking up way too small peripheral blood clots and then giving people a lot of anticoagulation on top. We know that the benefit-harm there of simply having a more sensitive test did not actually bare out into patient outcomes. We have to think very long and hard before we put something into the clinical realm of pan-cancer screening to make sure that there is going to be downstream

clinical utility that benefits patients. For multi-cancer screening it is also, I think, fair to say that finding something earlier is a bit reductive and we do need to think about what those clinical trial designs are that actually prove that earlier is better and relevant.

And then the last thing I would say is that what a lot of these tests do, and we've heard it said a few times, I think Geoff Oxnard said it best, is that it is a compliance issue, too. The best test is the test that gets done and the test that doesn't get done has a sensitivity of 0%. And I think, with our FDA colleagues, we need to think how do we recognize that, is it a third performance characteristic that we should be taking into account? I think the chance here is that we become incredibly disruptive and we really think about what are we trying to do in the 21st century with screening.

And those three points are what we've taken away from discussions today, but I would very much like to thank the FDA for working as partners in this and Guardant looks forward to a future where we can bring ctDNA to a very, very wide audience. Thank you.

(Applause.)

DR. GALLAGHER: Okay, so I'd like to thank all the speakers from the public comment session and that will conclude the session, thank you.

(Pause.)

DR. PATHAK: Well, welcome to Session 4, Clinical Study Design Considerations. I'm Anand Pathak, a medical officer from FDA's Center for Devices. Co-moderating are Dr. Wei Wang, a mathematical statistician from CDRH, and Dr. Erik Bloomquist, team lead in statistics at CDER, and if we could have our panelists give brief introductions. I'd appreciate if you could start with Dr. Gene Pennello.

DR. PENNELLO: Hi. Hello, everyone. I'm Gene Pennello, I'm at the Center for Devices and Radiological Health. I've been a team leader in the statistical evaluation of diagnostic devices for many years, since 2005.

DR. PATHAK: Dr. Berry.

DR. BERRY: Don Berry, I'm a statistician, MD Andersen Cancer Center, and they're not supposed to know I'm here because they prohibit travel, so I'm flying on my own nickel. In the disclosure sense, I was on the GRAIL SAB, scientific advisory board, when it existed. They said they wanted a skeptic on the board and they got a very vocal skeptic. I also am co-owner of Berry Consultants, that designs clinical trials in lots of areas including coronavirus in Europe and Canada and New Zealand and Australia, but not the United States.

DR. PATHAK: Dr. Doubeni.

DR. DOUBENI: Hi, hello again. I'm Chyke Doubeni, a family physician, clinical epidemiology at the Mayo Clinic and I'm also the director of the Center for Health Equity and Community Engagement Research at the Mayo Clinic across enterprise and a member of the U.S. Preventive Services Task Force. As I said before, I'm here on my own, I don't speak for the task force. Anything I say and do here is mine. Thank you.

DR. PATHAK: Okay. And also on the phone, Dr. Colin Begg, could you introduce yourself?

DR. BEGG: Yes. I'm Colin Begg, I'm head of the Department of Epidemiology and Biostatistics at Memorial Sloan Kettering Cancer Center. Let me say that like Don, I was prohibited from traveling and I actually obeyed the rule, so I'm still here in New York. And I've been interested in screening for many years.

DR. PATHAK: Wonderful.

Now, Dr. Skates, if you could introduce yourself.

DR. SKATES: Okay, this is Steven Skates. I'm a biostatistician at Massachusetts General Hospital and Harvard Medical School. My research career has been on early detection, primarily of ovarian cancer. We've developed a longitudinal algorithm and

implemented that in multiple clinical trials and that's been -- the software for that's been licensed to Abcodia, so that's a conflict of interest announcement. I'm also a consultant to GRAIL and I guess that's it. And branching out to other early detection cancers, as well, so --

DR. PATHAK: Okay, thank you.

And Dr. Gatsonis, if you could introduce yourself.

DR. GATSONIS: Sure. Hello, I'm Constantine Gatsonis, I'm a professor of biostatistics at Brown and a group statistician for the ECOG-ACRIN Cancer Research Group. I've worked in diagnostic imaging for decades and all the way from screening to patient management and surveillance.

DR. PATHAK: Okay, thank you so much.

Now, why don't we get started with Slide Number 69. Diagnostic evaluation of devices includes analytical and clinical validation. In terms of analytical performance, one looks at: does the test measure the analyte it's intended to measure correctly and reliably? And for cancer screening, we also look at diagnostic accuracy, does the test result correlate with the target condition of interest in a clinically significant way? And finally, there's clinical benefit, does the test support clinical decisions for patient management such as effective treatment or preventive strategies? So, all three elements are taken into consideration.

Current recommended and effective cancer screening tests include colonoscopy, sigmoidoscopy, and other stool tests for colorectal cancer. Some have been shown to reduce the risk of dying from colorectal cancer. There's low-dose helical CT for lung cancer for individuals in the high-risk group and that's been shown to reduce lung cancer mortality among heavy smokers. There's mammography for breast cancer, which has also been shown to reduce death in women, and there is the Pap test and an HPV test, and these tests lead both to early detection and prevention of cervical cancer.

Here's a brief outline of today's discussion. We're going to discuss multi-cancer tests with and without a tumor of origin component. For multi-cancer tests with a tumor of origin component, we'll devote the time mainly to diagnostic accuracy. The first question is whether the evaluation should be per cancer type or can be done actually across multiple cancer types pooled together.

And the second question is how do you evaluate cancer types with recommended screening tests?

And conversely, the following question is how do you evaluate cancer types without recommended screening tests?

And finally, there are additional discussion questions such as how do you design a supplemental study to supplement the prospective cohort studies, ground truth identification and so on, and this will be followed by touching on the clinical benefit aspect, as well.

And finally, we'll end with some brief discussion on multi-cancer tests without tumor of origin.

So now on Slide 72, the primary question we wanted to start off with is should the primary analysis for diagnostic accuracy be based on per cancer type analysis or is there any value in pooling multiple cancer types together for a test that actually has the ability to detect multiple cancers at one time?

So if we could start with Dr. Doubeni.

DR. DOUBENI: So looking at this question, I was reflecting to myself sort of the question of what is the question, right? So the question what you do in analysis will depend on the question that is asked and if the question asked is the ability of the test to detect a particular cancer or spectrum of disease or it's stages in a cancer, then that has to be done per cancer type.

But if the purpose of the study is to determine whether a particular screening test or assay is capable of identifying multiple cancers, then the analysis would need to look at all those cancers pooled and this is akin in some ways to many of the cardiovascular study outcomes where they pool outcomes together for a more powered analysis versus one that looks at, say, stroke or heart attacks or MIs. I mean, in the context of what we're talking about, I will presume that the interest here, it would be on looking at per cancer unless you have a very well-grounded idea about what the standard is and I would think that the interest would be per cancer but again, it depends on the question of interest.

DR. PATHAK: Okay, Dr. Berry.

DR. BERRY: So as we heard this morning, it's critical to analyze by cancer type. I mean, we can't have the knowledge that a patient has cancer, a person has cancer or the knowledge, the suspicion that a person has cancer, not knowing where it is. It's bad enough that we might learn that a patient has breast cancer and we don't know where it is in the breast and so the resort is to do double mastectomy. It's even worse if you don't know where -- what organ it is. So that's critical, I think. It's a sine qua non, as far as I'm concerned. It would be unethical to do otherwise.

It may be reasonable to pool cancer types in a trial, let's say you've got a pan-tumor assay and you do the right trial, which is a randomized trial. For the people that were just talking in the session, yes, it's going to be expensive and if you can't afford it, don't do it, do something else. If you can't afford to build a rocket to the moon, don't go to the moon.

Let's see, where was I? Pooling cancer type. So it's not unreasonable to build a randomized trial where the -- at least theoretically not unreasonable, where half of the patients are not told what the assay says and the other half are, ethics are a little bit queasy there, and cancer-specific survival is the endpoint. Not unreasonable, it would be a huge study, very expensive. I think it's -- the companies ought to, you know, lower their hurdle a

bit and not try something like that, but it's not unreasonable.

DR. PATHAK: Dr. Pennello.

DR. PENNELLO: Yes, I have some comments now from a statistical perspective and, you know, since I'm from the FDA, my comments may or may not reflect the entire FDA perspective or it may, but -- so just to keep that in mind. But the question is on diagnostic accuracy and -- but you can't, you know, disentangle diagnostic accuracy from what it -- you know, what diagnostic accuracy performance confers in terms of clinical benefit, and the clinical benefit depends on the benefit-risk tradeoff evaluation within these cancer types, so I don't -- I'm not sure how you can evaluate the diagnostic accuracy across cancer types because it won't be very interpretable in terms of the benefit-risk with individual cancers.

DR. PATHAK: Thank you, Dr. Pennello.

DR. PENNELLO: I would say one thing, though, that if the ctDNA biomarkers were tissue agnostic in some sense so that you would think a priori that the sensitivity and specificity would be the same across all cancers, then you might be able to pool them but that would have to have -- you'd have to have a good biological understanding of that first.

DR. PATHAK: Thank you for that perspective, Dr. Pennello.

Dr. Begg, can we get your perspective?

DR. BEGG: Yes. So let me preface my remarks regarding this question with some overarching comments here and it's partly based on some of the opinions that we heard this morning. It seems to me that ctDNA testing has the promise of a paradigm shift in screening because screening for individual cancers by current technologies is burdensome. People generally accept blood tests, they're very convenient to do, and the public and doctors generally like screening even if experts don't. So there's tremendous promise here. But many of the screening disadvantages are still present, the false positive numbers will be relatively large even if the specificity of the test is high and the harms of screening due to

false positives remain and they vary by cancer site. In addition, what's special about this technology is that you have the additional problem of identifying the primary site and you can't separate it from the problem. So in answer to the question that was posed, since this is a global test, I think pooled accuracy measurements are relevant to give us an overall sense of the quality of the test, but site-by-site evaluation of accuracy, harms, and benefits are crucial, also.

DR. PATHAK: Okay, thank you so much, Dr. Begg.

Dr. Skates, can we get your perspective?

DR. SKATES: Yes. It's very similar to some of the speakers before, which is each cancer type is going to have a very different harms -- benefits-to-harms ratio and therefore the characteristics or diagnostic accuracy that would be acceptable for one cancer is going to be very different for another cancer. So I can't see how -- you will have to do a per cancer type analysis and judge for each cancer type whether or not the harm -- benefits outweigh the harms and if five out of eight cancer types, the benefits outweigh the harms, then maybe that's -- those are the cancers that would go forward in the multi-cancer test and three would be eliminated because those three have harms outweighing the benefits.

And if you averaged over them and saw that the five together with the three on average, the harms -- the benefits somehow outweigh the harms, I think that would be doing a disservice to the patients because you can identify which of those three out of the eight cancers end up producing more harms than benefit. So I really see a minimal role for overall pooling across cancer types for diagnostic accuracy, at least at a clinical level, and as one of the speakers in the morning said, it might be a marketing tool but I think that's about it.

DR. PATHAK: Thank you so much, Dr. Skates.

And finally, Dr. Gatsonis, can we get your input?

DR. GATSONIS: Yeah, thank you. I'm going to step back for a moment and just remark that in this whole proceeding today there's a lot more questions than answers and this is understandable because the topic really goes well beyond the current paradigm we have for diagnostic test evaluation. It's really new territory and we should embrace the possibilities but do it in a way that -- cognizant of the fact that the only burden we have today, really, is how you evaluate a single test or maybe a couple of tests for a single condition, so going, you know, well beyond that on rhetoric rather than on data is going to be problematic.

So back to the question, then, of whether the analysis should be based on all cancers or by cancer, I think as we approach evaluation of cancer, we should tailor it to the intended use. There's a different kind of questioning of the test if it's for a screening test, let's say, and not for surveillance or for managing patients. So we have to separate the intended use and then develop an approach that is amenable to this. For instance, there are therapies that could be relevant across a variety of cancers because they attack a particular mutation. It could be that one of these pan-cancer tests can help in that process.

But by and large, I agree with what everybody has said so far, that if it's a screening test and if we're thinking of these modalities as primarily for screening, which I don't think that's a primary use, I think that surveillance, for instance, is an important other use. But if we're thinking about them as screening, then we have to try to understand how they perform for each kind of cancer.

And I'll stop with one more thing that we do need to address here, and I'm sure the FDA is very well aware of this, the realization of multiple inferences, in other words, if we start developing estimates of performance of any kind for every kind of cancer, for a single test and so on, we have to be well aware that there's a huge issue of multiple inferences and we need to do inferences that are adjusted for multiple comparisons.

DR. PATHAK: Thank you so much, Dr. Gatsonis.

DR. BLOOMQUIST: I'd like to follow up briefly with Dr. Gatsonis --

DR. SKATES: Yeah, I'd like to follow up on that, too.

DR. BLOOMQUIST: Oh, go ahead.

DR. SKATES: Yeah, Gene, go ahead.

DR. BLOOMQUIST: Whoever's on the phone that wanted to follow up, I think it was Steve, go ahead and proceed.

DR. SKATES: Yeah, all right. Okay. I'm hesitant to go down the multiple comparison path if there is a defined hypothesis as to which cancers we're going to pick up with this test. So if you say I've got a screening test and it's going to detect any cancer out there and there's hundreds -- and, you know, there's over a hundred cancers, I think it's that sort of a priori setting of not knowing which cancers you're exactly going to pick up with a test that you're -- where multiple comparison adjustments need to be made.

But if you've got eight specific cancers and you've got a lot of a priori data that those are the cancers that this test detects, I would be reluctant to make the additional hurdle of a multiple comparison adjustment in that. I think just like we've got a per cancer analysis of the diagnostic accuracy, I really think each of these tests stand on their own two feet and don't need to be adjusted for the fact that there is other cancers that are being evaluated at the same time.

DR. GATSONIS: If you're making -- yeah, if you're making comparisons to other standards, some of them will win, some of them will lose, for instance, so that's one of the issues.

DR. SKATES: Right, sure.

DR. GATSONIS: That's where the issue with multiple comparisons is and then I would encourage people to just -- instead of detecting six cancers to detect 60 and then they will

win on one.

DR. BLOOMQUIST: I'd like to follow up briefly with -- this is --

(Crosstalk.)

DR. SKATES: It's like making multiple comparisons for all of a cancer center's clinical trials. Oh, you did, you know, a hundred clinical trials this year, we'll adjust the p-value for all of them because --

DR. GASTONIS: That's not what I'm talking about. This is a very -- no, no, no, don't trivialize, don't trivialize the setting. The setting here is --

(Crosstalk.)

DR. SKATES: I'm not trivializing, I'm saying it's a very -- it's not a clear-cut issue.

DR. BLOOMQUIST: This is Erik Bloomquist. I agree. I mean, the multiple comparison, I think, is going to be a tricky issue to handle whether that's pre-specified up front or while it goes into the consideration of the individual test results. I just wanted to follow up on -- for those that thought there would be a use for a pooled analysis, so similar in therapeutic trials, sometimes we do a primary hypothesis of -- in the ITT, so there would be a top line result, there would be a gatekeeper, so we set a performance goal for sensitivity and specificity overall and then we could move down to each of the individual cancers.

For those that envision sort of a use for overall, would you envision some type of formal setup like that or would it just be let's get the results as is and then interpret them? And I would start with Colin Begg, because I believe you said that you thought there could be a use for overall pooled analysis.

DR. BEGG: I think my opinion here is motivated by the fact that when you search for circulating tumor DNA and you find it, in essence, any cancer -- any primary site is a possibility. So at some level the -- it is a global test and, you know, the first step would be

establishing the overall false positive weight and the numbers of cancers that -- the number of people who would be disadvantaged by having false positives. Overall, it's an inherent part of the evaluation. But I would emphasize the more important point, which is, every other speaker said that you can't dissociate the challenge here with the potentially very different costs and benefits and harms with respect to each of the potential sites of organs.

DR. BLOOMQUIST: Would any of the other panelists like to comment on how the pooled test could be used in conjunction with the individuals or is that too specific?

DR. BERRY: So when I suggested my theoretical pooled analysis based on the assay, it recognized that nobody's going to do this and what you've got to do, really, is focus on the individual and I want to stress something in that regard. When Colin said -- was talking about the harms, he mentioned false positive. When the GRAIL speaker talked about the harms, he said false positives, false negatives. The harm, it's written all over ctDNA, the harm to worry about is over-diagnosis and with the pooling, let's say you did a pooled analysis and you conclude that the assay is beneficial and it reduces mortality, it also is associated with a great deal of over-diagnosis. Trading off these things, nobody knows how to do that. Nobody even knows how to estimate what the over-diagnosis is and breast cancer numbers, estimates range from zero to a hundred percent.

The issue of efficacy, we don't really understand efficacy very well, even in breast cancer and mammography in the context, something that Barry Kramer was saying, that if there's no treatment, screening is worthless. If there's perfect treatment, screening is worthless. And there's a U-shape somewhere in between, maybe it's beneficial. In breast cancer today we've seen a 40% reduction in breast cancer mortality in U.S. and UK due almost certainly mostly to treatment. What role does screening play? It was originally developed when there was no benefit. Does it still show a benefit? And the task force struggled with this issue.

So it's really over-diagnosis and when you've got lots of things -- back to the multiplicities issue, when you've got lots of things you're looking at, you're doing differential harm in these various organ types. I'm a skeptic about the process, I love ctDNA in the treatment area, I think that's where we should be developing. I don't think it has a legitimate role today in screening.

DR. PATHAK: Thank you for the multiple perspectives on this question. Also, it looks like there's a rationale for per cancer type analysis and some people also think that there's some rationale for looking at multiple cancers all together.

So I think I'll transition to Dr. Wang's questions now.

DR. WANG: Okay, here is our hypothetical multi-cancer test based upon a single blood draw and this test use biomarkers on algorithms to detect the presence of a circulating tumor DNA in the bloodstream. And the device output indicates cancer detected or not detected. If detected, device will only give the most prevalent site of cancer which can indicate the tumor of origin. And the device will be limited to a predefined set of specific cancers of interest and the test is intended to be used on asymptomatic population, aged from 50 to 75.

And the predefined set of specific cancers contains, cancer types, with recommended screening test including breast cancer, colorectal cancer, lung cancer, and cervical cancers and also contain some kinds of types without a recommended screening test. For example, like a kidney cancer, pancreatic cancer, ovary cancer, etc.

I think since we still have some concern whether we should per cancer type analysis or if we should pool all cancers together but based on what Don Berry just said, since the benefit-risk analysis whether the benefit can be really achieved by this multi-cancer test, we should at least look at which cancer we are talking about. So based on the diagnostic accuracy analysis, the first question I want to ask the panelists to -- the panel members to

answer is how validate the diagnosis accuracy for cancer types with the recommended screening test, for example, like breast cancer, lung cancer, and etc?

Maybe we can start from Dr. Doubeni.

DR. DOUBENI: I'll give it a try. So I think we're talking about two scenarios. One is cancers that have an existing standard of care or a standard of screening and for those cancers, diagnostic tests or new ctDNA tests optimally could be compared to a gold standard. So take the case, for instance, of colonoscopy, assume the colonoscopy is perfect and can perform without error of measurement, then a test can be compared to -- in its ability to diagnose any one of those cancers to colonoscopy, and I think that would be a fairly robust way to test it.

But I think the challenge will come and I don't know that is the case for sort of breast cancer and again, we all know that some cancers that are too small to be detected will be missed by any of these tests, so it's not perfect, but I think that's a reasonable way to approach it. For the cancers that don't currently have a standard of screening, a screening standard --

DR. WANG: That's our next question.

DR. DOUBENI: Is that the next question?

DR. BLOOMQUIST: I think you can just sort of weave them all together, the --

DR. DOUBENI: Okay.

DR. BLOOMQUIST: -- "with screening" and "without screening," sort of what the clinical design would look like for each case.

DR. DOUBENI: Yeah. I think we're really challenged for those cancers because -- to be able to determine that your test is accurate, you have to have a gold standard for an asymptomatic population. Does that mean that you subject everybody to diagnostic testing for ovarian cancer to determine if you have ovarian cancer and then confirm that diagnosis

where the tissue is the pathology and I think that's very, very challenging. The other approach, obviously, is to follow a cohort over time, after you apply a diagnostic test to determine if cancer occurred or not. And that, obviously, would be valid but again, you'll still miss the cancers that were not fatal of someone that's from a competing condition. So I think it's a very challenging thing to do for the cancers for which there's no existing standard.

DR. PATHAK: So, Dr. Doubeni, going back to the cancers with existing screening, you said comparative gold standard. Now, what type of study design would you envision for such a comparison?

DR. DOUBENI: I think it can just simply be a cross-sectional study. If you have -- you created, like, similar to what I believe certain -- folks did or even Cologuard did, do a screening of, say, 10,000 people, whatever the number is, to determine if they have colorectal cancer, adenomas, and other lesions and you apply a test to see if it performs comparably. That, to me, is a good way to determine if it's accurate or not.

DR. PATHAK: Thank you.

DR. BLOOMQUIST: And would there be any long-term -- Dr. Doubeni, would there be any long-term follow-up, maybe additional screening tests done 1 year later or 2 years later or would it just be at a single time point?

DR. DOUBENI: I think, though, you can do a prospective study, but from all we've heard, I think the challenge is going to be the cost of long-term follow-up. Again, colonoscopy is not perfect but I think it's good enough as a medical standard to apply in this case for comparison and I think that's a reasonable comparison, maybe.

DR. WANG: And the question here for this -- we already have recommended screening test, what's the type of analysis to evaluate this multi-cancer test, actually. What do you think?

DR. DOUBENI: I beg your pardon?

DR. WANG: Like, actually, these kind of cancer types already have some recommended screening test. With those information available, how can we evaluate this new test, actually.

DR. PATHAK: I think the question is how do we know there's added benefit of this test over the standard of care.

DR. DOUBENI: I think you're asking -- let me maybe rephrase that question slightly. So if you subject a -- set up participants to the existing standard and you subject them to the new standard, assuming that a test is designed to do the same thing, right, so there are different standards, there are different ways to look at this. One is different technologies, a little bit different, more difficult to tell if the test is going to do exactly the same thing. For instance, if you're comparing a stool-based test to a stool-based test, you're detecting it from the same source of tissue, it's a little bit different.

Now, in this instance, I would recommend to you that you look at a distribution of lesions that are detected by the standard to which you're comparing, because if a new test is only detecting advanced cancers that are detected by the original or the standard, then they're not comparable and so that may be deceptive to think that the test is accurate because there are many ways you can measure the accuracy of this test.

As we know very well, when we look at colorectal cancer screening, we look at sensitivity and specificity with respect to cancer detection, advanced adenomas, and so those are all different ways of looking at it and you have to declare which is your outcome measure that you examine. But what I think is important, because we've had this conversation before, is to look at a distributional stage in the patients in whom the test is being applied and also make sure that there's concordance between the two tests, all right, it's not just sensitivity and specificity, but concordance between the two tests, that a

patient who is detected with cancer is also detected by a new test, as well.

DR. WANG: Okay.

Okay, Dr. Berry.

DR. BERRY: So I served on a panel in the early days of HPV testing and what they did was they -- what the company did was they did HPV and Pap and then, you know, analyzed the marginal abilities and -- back then HPV was actually better than Pap, as it is today. But they didn't know what to do in the case of positive/negative and there were like 40 MDs in the audience and I asked them, I said what would you do if you had a positive Pap but a negative HPV and they didn't know. So the FDA sent them back to do their homework and eventually, it got approved. Presumably, that's what they would do.

I just want to comment on this question that's up there, how to validate diagnostic accuracy, that's appropriate for a diagnostic test. The appropriate thing for a screening test, and I hearken back to this morning, somebody wrote a -- defining a positive or successful test and listed a bunch of things none of which was relevant for what I thought would be a positive or a successful test, namely that you show that it has clinical benefit that outweighs the harms and that means a randomized trial. I'm sorry, people, it means a randomized trial.

DR. PATHAK: For the tests that do have, let's say the four that we suggest, that do have, let's say, a recommended screening test, could you see sort of a sensitivity/specificity for those or just any -- every cancer needs the clinical benefit?

DR. BERRY: So it's a little bit like Barry Kramer said today when he was talking about the late-stage disease and I asked him about it, I said, "Barry, you and I agree on everything except I wonder about late-stage disease," and I said, "For me, it's necessary but not sufficient." He agreed. So it turns out we agree on everything, I think. And it's the same here, it's necessary. If you're going to develop something you want to have good

sensitivity, good specificity; you want to not only, in the opt-in/opt-out you want to not only address the possibility that adding it so you have two tests is better than just one or subtracting the other. So necessary, but not sufficient. What is sufficient is to show that you have a mortality benefit that outweighs the negative of over-diagnosis, etc.

DR. DOUBENI: Erik, can I respond to that?

DR. WANG: Yeah, go ahead.

DR. BLOOMQUIST: Dr. Doubeni, did you have a response to that?

UNIDENTIFIED SPEAKER: How about everyone else giving their opinions and then coming back for responses?

DR. BLOOMQUIST: Let's just -- well, I think we'll let Dr. Doubeni respond and then we'll try to quickly move to those on the phone.

DR. DOUBENI: So I serve on a group that's been accused of being very conservative, so I have to be careful about what I say, but I think the -- we have to look at sort of the way we've handled -- addressed this test and we're going through some of this exercise internally. When you have a test that has been shown to be effective and the example here is fecal guaiac FOBT. And so preface it, Canadians don't approve or recommend colonoscopy because there isn't a randomized trial and a -- and so on a basis of that, then came fecal meta-chemical tests which has a superior sensitivity and that's based on the modeling studies were deemed as being adequate as a screening test to use.

We now have a trial in the field that will hopefully show that it's effective in comparison to colonoscopy. I'm not sure what right answer we'll get because it's a comparative effectiveness study. We don't have a trial of colonoscopy at this point, we have one of sigmoidoscopy. So I say all of this and also, of course, we don't have one of stool DNA FIT because we recognize that it is not necessary when you have tests in which -- a condition for which you have a well-defined and proven mortality benefit to require that

level of study rigor. I don't think it's necessary, I think if we did so we'll sort of make it impossible for us to develop new technology. I think we just have to look at the way the test is used, its application relative to what we use currently, the performance characteristics of the test, make sure that it's comparable to what we currently use. I don't think you need a randomized trial for a test that is doing exactly the same thing with the same mechanism.

DR. BLOOMQUIST: Okay, thank you. Let's --

DR. WANG: Dr. Pennello --

DR. BLOOMQUIST: Dr. Pennello, did you have any comment?

DR. WANG: -- any comments?

DR. PENNELLO: How to validate diagnostic accuracy depends on the intended use and I agree with all the comments that are being made. You know, if the recommended screening test has been used and been shown to have clinical benefit and your intended use is to replace that, I don't know if that's very likely with a CT biomarker but if it is, then you could use the diagnostic performance of the screening test as sort of a starting place for how well the performance of the ctDNA test is. If the intended use is to add value to the recommended screening test as a rule-in or as a rule-out, then to me the important subsets to look at are screening test negatives and screening test positives in which you -- the test would disagree and then see what the performance is there because that's where you would change the clinical action.

DR. WANG: Okay, so next to the panelists on the phone, Dr. Begg, can you express your opinion about this question?

DR. BEGG: Yes, so let me say that the -- I think we're all struggling and having a hard time figuring out what the correct study design would be and I certainly don't have a firm opinion on that, so I'll just give some general comments. But I think the first comment I

would make is that I kind of view these ctDNA tests as different from individual screening tests. You have to -- instead of the new pan-cancer tests in a holistic kind of way, so it's entirely possible that you'll develop a pan-cancer test that is good for some sites and not good for other sites. That doesn't necessarily rule in or rule out the test as being appropriate for use. So while comparisons of the properties of the test at detecting cancers on a site-by-site basis are undoubtedly going to be interesting and informative, I don't think they'll be definitive in this context.

Having said all that, one is going to want to do some kind of statistical or quantitative testing and it seems to me that if you're dealing with a site with an existing test like breast cancer or colon cancer and so forth, trying to reach some kind of equivalence of diagnostic accuracy would seem to be a useful starting point, whereas for other sites without an existing test, just establishing that you can detect some kind of signal would be a positive thing. But once again, it's the holistic evaluation of the test that one is really going to have to grapple with.

And I just want to make another comment, though, here that if -- we can't separate the test from the classification of the primary site and I suspect that -- and the evaluation of the algorithm, the predictive algorithm for primary sites which has to accomplish -- accompany the test is going to be a challenge in and of itself and I suspect that rare cancer sites will have relatively low probabilities and that's something we'll have to be concerned about.

Finally, I just want to talk about the notion of a randomized screening trial. I mean, I agree with Don, that that's ultimately the gold standard and all this but I suspect we're going to have to make decisions about these tests before we wait another 10 years or so for such a study to actually be accomplished and I think there is a place for studies in the short term of the screening accuracy of a new test and important details of that include the

appropriate selection of cases and controls and so forth, but I think we'll be coming to that later in the discussion.

DR. WANG: Okay, Dr. Skates.

DR. SKATES: Great. I want to pick up on this distinction between tests where there's already a standard of care testing because I think that makes very different considerations from when you don't have one. Where you've already shown clinical utility and most of the utility is being disease-specific reduction in mortality, then a comparison of the same patient with two tests as, I believe, they did with the Cologuard test compared to the FIT test. It makes, to me, makes a compelling argument that you don't need to show, for the Cologuard test, a reduction in morality if it's already being shown with something that's even better than the FIT test, for example, the guaiac stool test that Dr. Doubeni mentioned.

So then you have to -- but you can't stop there because you have to make the argument that the harms are also commensurate. So there might've been improvement as there was in this case with an increase in sensitivity, so it picked up a substantial number, substantially more colon cancers, than the FIT test did, but at the same time it referred about three times as many to colonoscopy.

And so there's a potential harm there and you have to make the argument that that increase in the sensitivity outweighs the harm in the -- whatever morbidity that would occur with three times the number of colonoscopies, and I think that was somewhat implicit in the panel hearing that occurred then, and I was a member of that and my thinking has evolved since then, which is a more explicit need to trade off those benefits and harms. And so I don't think you need a randomized trial with an endpoint of disease-specific mortality any time you come up with a new test if there is a test already out there that has shown clinical utility in a randomized trial. I do think where -- I do think you need a

randomized clinical trial where clinical utility hasn't been shown because, as Dr. Kramer said, we're usually on the margins of, you know, clinical utility, it's 20 or 30% reduction in mortality and you need big trials to show that. So it all goes back to this tradeoff between clinical benefit and harm and I think that should be explicit in the comparison. We can do that, I believe, with a paired comparison on the same set of patients in a cohort study.

In terms of if you don't have a gold standard, I guess colonoscopy is considered a gold standard, but there are other situations where you might -- where the current standard of care might not be considered a gold standard and there the answer to that, in my opinion, is to follow up within, say, a year of everyone and whatever clinical diagnosis they have then, you will be able to identify cancers that were missed and cancers that were -- and then false positives showing those had no cancer, they would be considered false positives at the end, if they didn't have cancer at the end of that year.

And it might not be a year, it might be 2 years or something but you follow that cohort up within a defined time and that time might be reasonable to consider as to when you do the next test. So if you're doing an annual test then follow up after a year, if it's every 2 years then maybe follow up after 2 years and that will give you -- without a gold standard, that will still give you your ability to estimate your accuracies, your sensitivity, specificity, PPVs and NPVs. That's it.

DR. WANG: Okay, thanks, Dr. Skates.

And the lastly is for opinion from Dr. Gatsonis.

DR. GATSONIS: Yes, thank you. I turned a bit with the -- this formulation validates because if you're talking about diagnostic accuracy in order to validate it, essentially you need to show that the assessment you have is accurate, so to speak, it's correct. So it's not against outcomes necessarily, it's against the truth, that's how you validate diagnostic accuracy. I assume we will go to the question of how does the test affect outcome. There

we will talk about the need for randomized studies and possibility for randomized studies and so on. But I think for this particular question we need to stay with how do you assess the diagnostic accuracy, yeah, and typically for tests for which -- in situations for which there is a comparator, the paired design, in other words, where you do both the test and the comparator on the same patient is the design that has proved to be most efficient provided you can have an adequate reference and the reference then can be defined in the number -- in any number of ways and there are plenty of examples.

So if I have a panel that has four, five, six kinds of cancers and for every one of them there is a comparator and essentially what I'm talking about is a paired design in which you do the ctDNA test and you also do all the comparator tests for the particular organs and cancers. So that would be the design that I assume will get the answer about the diagnostic accuracy. The question as to whether -- to what you do if there are no comparators, then essentially you need to just estimate the diagnostic accuracy of each -- of ctDNA for these particular cancers and with an adequate reference standard and then report that in the literature. I think the answer for this, in other words, is fairly straightforward, it's how you would validate the diagnostic accuracy if you do it one cancer at a time.

What Colin is talking about, the pooled accuracy, that's a different -- that's a different kind of question, but a paired design is the way to go when it comes to validating diagnostic accuracy. I assume we will get to the question of how that does affect outcomes because paired design does not allow you, really, to evaluate the impact of the test on outcomes because you have both kinds of information. For that one you may need a randomized design if you don't have already several kinds of randomized studies with a comparator.

DR. PATHAK: So this is a question for Dr. Doubeni. We heard earlier that these blood-based tests may be able to pick up, you know, later-stage cancers and some early

stage cancers that are probably not, you know, precancerous lesions such as advanced adenomas. What is your take on a blood test that, you know, if it does not pick up advanced adenomas but has adequate performance on picking up colorectal cancer?

DR. DOUBENI: So the tests, Dr. Pathak gave it sort of -- we have a principle, if the test performs very differently than the current test we have, then you're back to the place where you need a more rigorous design to determine the benefit, does that make sense? So in other words, if a test is only picking up advanced cancers because we know that a current test picked up a spectrum of disease from early to late cancers and if all it does is advanced cancers, then it's not the same approach that has been used for this new device, in which case you need a randomized trial.

DR. PATHAK: Okay, thank you.

DR. BLOOMQUIST: Can the speakers on the phone mute just a moment?

(Pause.)

DR. BLOOMQUIST: Can you hear us on the phone?

DR. GATSONIS: Yes, yes, and we can hear also the noise.

DR. SKATES: Yeah, someone has hung up somehow. Is that Colin? Colin, are you on?

DR. BLOOMQUIST: You know, I'll try to talk a little bit while this dial tone is going on. I think what we've heard this morning as most people are thinking of a cohort study, let's say a large cohort study for the diagnostic accuracy portion. I think the clinical benefit part might involve some randomization, but at least for the diagnostic accuracy, I'm thinking of primarily the cohort. Just wanted to -- you know, on some of the logistics of this, for sample size should we try to power so that we have a minimum size, let's say for cancers of interest? And then also, you know -- and even in a lot of therapeutic studies, lost to follow-up is a big concern, so you can get people initially for that initial test, that initial screen,

what do you do for, let's say if 10% drop out at year one and another 20% drop out at year two, how do you handle, let's say lost of follow-up for these cancers that don't have a gold standard. And then third --

DR. BERRY: So could I address --

DR. BLOOMQUIST: Sure.

DR. BERRY: -- just the cap on the individual tumor types? So what you want to do is a Bayesian hierarchal model and you want, too, you know, random effects where you're borrowing across to the extent to which you're seeing benefit. For example, the -- in drugs, as you may know, they now are doing pan-tumor approvals for MSI high and, you know, therapy and many of the tumor types were not even represented but they get carried over because you're seeing a consistent benefit, it's like induction. You know, George Gamow wrote a book called *One, Two, Three, Infinity*. This is up to like a dozen tumor types and you're seeing the same effect so you just buy in to the biology and just say it's proof for everybody with MSI high. So that sort of thing you would want to do. In terms of the missingness, Erik, this could take until tomorrow or the next day or something and it's a ubiquitous problem and to quote Tom Flemming, "The only way to solve it is to not have it."

DR. BLOOMQUIST: Would you see any -- go ahead on the phone.

DR. GATSONIS: Yeah, I mean, in terms of the drug project and so on, if you're talking about assessments of diagnostic accuracy, there are methods to adjust for verification bias. So I think you could take some of these and adapt them for this situation. I don't think that would be too complicated. So the missingness part is not so much an issue and if you have a cohort, then you follow it prospectively, the - you have at least established a standard way by which you will assess a reference standard, which is the most important part of this. From there on, the rest of the --

(Phone cut out.)

DR. WANG: Online, can you hear us?

DR. BLOOMQUIST: You're missing. Well, while we get them back, Don, so I mean, these cohort studies would be very large, let's say in the five to six digit range. Oh.

DR. GATSONIS: No, no, no. We're not talking about -- we're talking about accuracy studies.

DR. BLOOMQUIST: Oh, you're back on the phone.

(Crosstalk.)

DR. GATSONIS: -- that's an outcome.

DR. SKATES: I mean, the Cologuard test was about 10,000 patients and done once, so that's not a -- you know, that's a feasible size --

(Crosstalk.)

DR. BLOOMQUIST: You know, I think I was speaking for the ones that don't have, let's say, like a recommended screening test like Cologuard, we'll say pancreatic cancer, and there were some suggestions of following these people up for 1 to 2 years. I think it's pretty standard in clinical trials for therapeutics that you're going to get people that withdraw consent, they don't want to have additional screening, let's say. And if this becomes a sizeable problem, well, how should we handle possible data here? I mean, these people could possibly have pancreatic cancer or they could not and just for that part of the --

(Crosstalk.)

DR. BLOOMQUIST: Just one at a time. Dr. Begg, please. Were you speaking, Dr. Begg, or was that Dr. Skates?

DR. SKATES: It was Dr. Skates.

DR. BLOOMQUIST: Oh, Dr. Skates, please go ahead.

DR. SKATES: So I guess I'm just picking up on what Constantine was going to say,

that the assumption that you might need to make is that these people drop out --

DR. ABUKHDEIR: Steve, can I interrupt you for just a sec? Can we ask everyone to mute their phones, please, just for a brief moment? We want to figure out where the noise is coming from.

DR. SKATES: I'm pretty sure, we haven't heard from Colin. I'm guessing it's from him, by all means.

(Laughter.)

DR. ABUKHDEIR: Colin, can you mute your phone? Is Colin there?

DR. GATSONIS: Well, he can try disconnecting his phone and then he will connect back.

(Pause.)

DR. BLOOMQUIST: While we're waiting, Dr. Berry, yeah, these are large cohorts. Could you imagine some type of Bayesian way to reassess the sample size as you're going along, you know, let's say 50,000 patients we assess, how many patients we're getting, and we try to look to see the additional cohorts or would you just recommend a fixed sample size, collect, and then reassess at that point?

DR. BERRY: No. So I'm all for -- somebody mentioned innovative trial designs earlier today. Definitely, you want to be adaptive and adjusting according to some prospective criteria, you set down prospectively what you're going to do and then if you get the answer to some questions then you move on to other questions like different cohorts, different tumor types.

DR. BLOOMQUIST: And then -- go ahead, Colin.

DR. BEGG: I have a comment on this one, if I can get in. Can you hear me?

DR. BLOOMQUIST: Yeah, we can hear you. Go ahead, Colin.

DR. BEGG: So I'm sort of repeating myself a little bit here in that I view the

evaluation of these pan-cancer tests as holistic, but it seems to me if you set up a cohort or if you did a randomized trial for that matter and the actual numbers of cancers, the number of cancer deaths downstream will be much greater when you're looking at pan-cancer than they would be if you're looking at individual cancers. So in fact, we may be able to do more with less in this context than what we might otherwise and you're going to be trading off the precision of the individual comparisons with regard to individual cancer sites, but if a test is to be used at all, it has to be used for all cancer types and so this sort of holistic overall question is sort of paramount to me.

DR. BLOOMQUIST: So, Colin, would you agree with a Bayesian hierarchical model to do what you said?

DR. BEGG: Sorry, say that again?

DR. BLOOMQUIST: Would you agree with a Bayesian hierarchical model to do what you said?

DR. BEGG: You have to think about that, but you know, I -- and certainly, that is another way of extracting more information out of the data you collect, but I think fundamentally -- my fundamental point is if you -- if you're following a cohort or a trial and it's pan-cancer, you're going to see more events, many more events than you would for an individual cancer.

DR. BLOOMQUIST: Yeah, so I agree. I'm thinking, because there's a lot of feedback in the room what we'll need to do is mute the individuals on the phone and then we'll ask, I think, our -- we'll ask Gene and then when we get to the individual speakers, we'll try to bring you back on the line to avoid some of the feedback, so --

DR. SKATES: Erik, can I suggest we all hang up and call back in? That might get rid of it.

DR. GATSONIS: I just did that and it did not work.

DR. SKATES: Right, but it's the person who's not hanging up that is the issue.

DR. BLOOMQUIST: So I think, you know, as -- what we're going to do is -- can you mute the speakers on the phone, Wei? Okay. So I apologize for the feedback there. And then once we move to the speakers on the phone, we'll have to endure it, but I think their comments are very valid, too, so we'll bring them back on. But Gene, while we're, you know, I think in silence here, do you have any comments on sample size considerations and sort of adaptive things for the sample size as we go on, basically follow-up issues, some of the logistical issues with these large cohort studies?

DR. PENNELLO: Well, actually I would like to hear from the panelists a little more about, you know, if you have to do a randomized trial is there a feasible, you know, efficient randomized trial because at FDA we're interested in -- and I think you've heard today that, you know, if these trials or other studies are infeasible or take forever to do, we'll never be able to evaluate these tests. Are there better study designs for which we can actually evaluate these tests?

DR. BLOOMQUIST: I mean, but for -- I understand that the cohort is not ideal and I think we're going to move a little bit later to some of these, let's say the clinical benefit questions, but could you see a cohort study being useful? Just a follow up.

DR. PENNELLO: So, you know, there's the case cohort design which you could do, you know, a prospective/retrospective study design that seems to me would be pretty efficient here because you're looking at multiple cancers and you could compare them against the same sub-cohort assuming that you had, you know, a full cohort in which you can evaluate all the cancers.

DR. BERRY: So Gene, say what you mean by the case cohort, what is the cohort, what is the case? What is the comparison? So you've got an assay and are you talking, for example, the multi- or the pan-tumor, what are you talking?

DR. PENNELLO: Well, I'm thinking a multi-cancer test.

DR. BERRY: Okay, multi-cancer test, so you identified, let's say a dozen tumor types that you're looking at and what is the control, what is the comparison? How do you know that you're doing better than these, than the assay?

DR. DOUBENI: If I may, I think the challenge here is that for us to determine the accuracy, diagnostic accuracy of any test, you have to have a gold standard, right? If we agree on that basic principle, then obviously the challenge of any of these approaches that we're looking at becomes quite apparent and so whatever we do is going to be imperfect. And I say this in part because we know that --

DR. BLOOMQUIST: So let me modify --

DR. DOUBENI: Yeah.

DR. BLOOMQUIST: -- and say none of the therapies, none of the tumor types that have screening tests now, so these are all, you know, pancreatic cancer, etc., and we've got, let's say eight of them.

DR. DOUBENI: Right.

DR. BLOOMQUIST: And we're going to do a case cohort to show that there's a mortality benefit. So --

DR. DOUBENI: I think you are --

DR. BLOOMQUIST: -- how to do that.

DR. DOUBENI: Yeah, never mind.

DR. BLOOMQUIST: How to do that.

And let's wait until the mortality, I think that's sort of the last part, is the actual mortality, actual time to late-stage disease. We're trying to focus on just the cohort here, so -- but I think --

DR. BERRY: Well, what is the endpoint?

DR. BLOOMQUIST: You know, I think the endpoint for the ones that do have a screening test would be the comparison to the screening test and then for the ones that do not, let's say, have a recommended screening test, it would be several years follow-up, let's say 2 to 3 years and then it would be those that we would diagnose with pancreatic cancer versus how the tests are done. I mean, I think that's the general sense of how this works and the idea is -- I understand that, you know, you wouldn't actually get the clinical part with that, but the idea was if we were doing a cohort study, some of the logistics here, so --

DR. PENNELLO: In a case cohort, it would be diagnostic accuracy, but I did want to -- for each cancer, but then it would be efficient because you would compare each cancer type with the sub-cohort. I want to mention, and I'm not sure how many folks are aware, but there's statistical literature now on measures of net benefit which compares the expected utility of a test compared against the expected utility of standard of care and it's a way of -- it's a way of evaluating the potential clinical utility of a test and it's a function of sensitivity, specificity, and agreed-upon risk thresholds for ruling in, for further workup.

Individuals are ruling them out with a rule-out risk threshold and this started with Vickers and Elkin's decision curve analysis and Stewart Baker's relative utility curves, etc, but I think -- and now there's -- and Margaret Pepe's written on this, as well, but I think it might be a way to use the clinical validation study together with agreed-upon risk thresholds to look at the net benefit from a decision/utility perspective, but not to say that that would replace the randomized trial, but that might be one way to look at it.

DR. DOUBENI: May I return to the diagnostic accuracy? What I would recommend is that, you know, for EDRN, that's cohorts with well-characterized and annotated participants, you have to able to know the status and be able to use a marker to identify whether that patient has cancer or not. In this case you're looking at nested case-control design, some of the design that allows you to look, be able to assess an assay and compare

it to the clinical status with patients. There's still some misclassification, but I think that's the best way. For pancreatic cancer, for instance, you have -- for instance, we could do an MRI to determine whether the patient has pancreatic cancer.

DR. WANG: Yeah, I think that's an interesting question if you really want to know, like our second bullet here, how to identify, since we focus on the clinical efficacy or diagnostic accuracy for the multi-cancer test, how to identify the clinical -- at baseline for diagnostic accuracy estimation, like whether we can apply somewhere using bias correction measure or use some follow-up information ascertainment, so I just wanted the panelist --

DR. BLOOMQUIST: And Don, that's actually --

DR. BERRY: Wait.

DR. BLOOMQUIST: Go ahead. I want to give the people on the phone just a chance, but go ahead.

DR. BERRY: I want to go back to Erik's -- and what I think Erik is trying to do is to get a -- some sort of resolution to this "no, you can't do anything" and if you did, let's say, let me call it the real test, the real trial with randomization, I think it's possible. I think you could build an adaptive trial that's looking at the data as you're moving along and building, you know, an automaton that's going to run the trial, you're controlling Type 1 error, etc. But then you look at the high risk and when I'm saying high risk, I mean really high risk.

Take the ones that you saw in the list today, add to it even higher risk of genetic risk factors and so that the events are -- even BRCA which, of course, is associated with a number of cancers, pancreatic, ovarian, in addition breast, prostate, and these are people, of course, that are taking measures, sometimes mastectomy, oophorectomy, etc., but we statisticians are great at handling such things. And then you run the trial, you try to have something that's going to get an answer; even a weak answer to the right question is more important than the strong answer to the wrong question and you can do this with, you

know, some small, relatively small -- not that the small biotech is going to be able to handle it, but maybe the NIH could handle it or the NCI, and so it is possible to do to get some information. To Barry's point earlier, you're then defining a niche, you know, the tip of the iceberg in terms of risk and what you find for the high risk is not necessarily going to be there for the average risk but it's establishing the concept, and once you've established the concept then there could be more enthusiasm for doing the average risk which is going to take, of course, you know, government support.

DR. PATHAK: Dr. Berry, thank you for that perspective. Are there any other ways to enrich the number of cases, like any other enrichment strategies that you could think of?

DR. BERRY: Well, those that have been proposed, we talked about that earlier and we had a speaker who that's a favorite subject of hers, real-world evidence and all of the biases associated with that, I mean, maybe you could do some modeling and understand what the biases are associated with that when you're comparing to the smaller randomized part and so recognize that there are differences, but get some information as to what the difference is so that you can bring that into the -- you know, all of the data. So yeah, I think it's possible. It would take some willingness on the part of the FDA to be innovative and, you know, to push the periphery of what they usually accept.

DR. BLOOMQUIST: And can you bring the speakers on the phone back on, Wei?

Steve, Dr. Skates, are you still there?

DR. SKATES: I am.

DR. BLOOMQUIST: Okay, so --

DR. SKATES: Can you hear me?

DR. BLOOMQUIST: Yeah, we can hear you. So I think -- I don't know if you've been following along here, we've been --

DR. SKATES: I have.

DR. BLOOMQUIST: We've been discussing some of this cohort and then I think sort of still the endpoint questions. In terms of enrichment, are there ways -- let's say you would do a large cohort, even a randomized study, I think the number of cases could be small, ways to enrich the study to, let's say, include enough cases for adequate estimation.

DR. SKATES: I find the whole discussion of enrichment to be a little puzzling because I think there's going to be minimal gains from it. I did a study on high-risk women at -- for ovarian cancer and these were women who had multiple ovarian or breast cancers in first or second degree blood relatives and we found that the incidence -- because we opened it up to women who were 38 years old and above, that the incidence of ovarian cancer in that age group really didn't change too much from the women who -- the women who were post-menopausal normal risk population primarily because the doctors who were referring their patients to the screening were on the lower end of the spectrum of disease and the women who were at the high end were going immediately to prophylactic hysterectomies.

So I don't see enrichment as a solution here to sample size. What I see is the big issue or the big change from doing a 200,000-women trial, which is what we needed for ovarian cancer mortality benefit, is to change the endpoint and whether or not it's acceptable to show a reduction in late-stage disease and the absolute incidence of late-stage disease as a clinical utility endpoint and that would change the length of the trial from 20 years down to maybe 5 years or perhaps even fewer. That's where the big change is going to be. Enriching, because you're looking at smokers, 30 pack-years, in low-dose CT defines a very narrow range of the population and it's going to miss 90% of the cancers. So, you know -- and it will have a purer sample size calculation, but you then have a population that you -- a target population that's a small fraction of those that -- where most of the cancers occur. So I hesitate to go down the enrichment path that far and I think the big change that could potentially be here is changing the endpoint.

DR. PATHAK: Dr. Skates, what about age-based enrichment? What about studying --

DR. SKATES: Yeah, so -- yes, absolutely. I think since for most cancers age is such a strong risk factor that having an age cutoff like 50, above 50 or above 40, maybe -- maybe -- and you might want the age cutoff changing with the types of cancers that you target, but definitely an age cutoff --

(Crosstalk.)

DR. BERRY: Or above 60.

DR. SKATES: Well, you could try that, too. It all depends what fraction of the cases you're going to be picking up in that population and the higher you get, the fewer -- the smaller the fraction of cancer. So there's going to be a tradeoff between putting it too high and getting a greater enrichment and putting it too low and getting a larger sample size so, you know, the above 50, I think, is a reasonable first start but it would be worthwhile to look at above 60, above 40 depending on the cancers.

DR. PATHAK: Thank you for that input.

DR. BLOOMQUIST: You know, is there any -- you know, we're running a little bit short on time and I still would like a few moments. Is there any questions from the audience for our statistical panel? Sure, go ahead.

DR. BERRY: While he's coming up, just to mention one other enrichment, at MD Andersen I'm running a trial with -- in metastatic disease looking at profiling, genetic profiling with Foundation Medicine and Tempus and we find that patients aren't willing to be randomized to get the test, so we've established -- we offer randomization; if they refuse randomization, we put them in the trial anyway and give them their preference. So here you would have people who would want to have the test and others who would not want to have the test and so you would follow them and what has been found in the literature of, you know, a few dozen such trials is there's no difference between the two groups, the

preference group versus the randomization group, unless you're asking a question like do you -- do you like the treatment that you got?

DR. BLOOMQUIST: Barry Kramer, I remember you from earlier, so go ahead and ask your question.

DR. KRAMER: Yeah, perhaps this is to Chyke, but we talk about two very similar tests which may allow us a different study design, the jump between FOBT and FIT was very small and you can sort of piece things together by operating characteristics, I think. The jump between FIT and mSEPT9 is much larger, one is blood and it's totally different and that, I think, becomes then a real stretch to try to draw on the characteristics. And then finally, I'll point out that ctDNA has been labeled here at these meetings as a total paradigm shift and so I think that that is yet different again, and so we can't rely on short jumps and extrapolations, in my opinion.

DR. DOUBENI: Barry, I agree with you and let me try to restate it. I think the questions went back and forth and maybe the lines were mixed. The ctDNA, in my view, if we're able to comparably detect the same lesions as are detected by, say FIT or colonoscopy in the same spectrum, then I think it's a reasonable comparison. But I think what I said, also, was that if you look at ctDNA as a different technology, which is look at stool tests and stool tests, the same --

DR. KRAMER: Yeah.

DR. DOUBENI: The same approach to diagnosis --

DR. KRAMER: Yeah.

DR. DOUBENI: -- is fairly comparable and it can make that leap.

DR. KRAMER: Okay.

DR. DOUBENI: Looking at ctDNA, you may be looking at a different technology, apply a different approach.

DR. KRAMER: Yeah. And I would just point out that two tests --

DR. DOUBENI: Yeah.

DR. KRAMER: -- with the same sensitivity and specificity may be drawn from a very helpful test and a very harmful test since these are close, anyway.

DR. DOUBENI: Concordance.

DR. KRAMER: Yeah.

DR. DOUBENI: Concordance. Detect in the same patients.

DR. KRAMER: Yes. Well, yeah, but sensitivity, you always have to ask sensitivity for what.

DR. DOUBENI: Um-hum.

DR. KRAMER: Is it sensitivity for the exact same spectrum of disease --

DR. DOUBENI: Correct.

DR. KRAMER: -- or sensitivity for a different spectrum of biologic disease? One last thing, just a brief comment, and that is I hear that we often can't do trials because it requires scores of thousands of people, but we have to be careful that sometimes the alternative is to take very low-level evidence and then start putting into practice where it's not scores of thousands, it's a hundred million people or scores of millions and then you almost can't do a randomized trial because it's out there and equipoise has been lost.

(Crosstalk.)

DR. BERRY: That why we have regulators, Barry.

DR. BLOOMQUIST: One more question and then we'll have to conclude. So please go ahead.

UNIDENTIFIED SPEAKER: So one comment. I know there's a premise that's been running through some part of this day that I just -- I would like to challenge because we have evidence to the contrary and that is that precancerous lesion detection is not

detectable in blood. We have direct evidence against that, so I'm just going to make that as a comment. The question I had for the panel is that I feel like we're -- I just may be a little lost, but I feel like we're conflating clinical validity with clinical utility, okay, so if we're talking about clinical validity and diagnostic accuracy, you're saying does this test detect cancer, just as an example, right? So what I'm struggling with as opposed to clinical utility, which is like if you use this test does it actually improve outcome -- you know, does it have an outcome benefit, that's a very different thing and I think we need to separate those.

But coming back to the clinical validity question, then, what I was struggling with is to understand, so let's just take the four screened cancers, right? So in a cohort study where we do that, you know, in DeeP-C everybody got a colonoscopy, right, so you had truth for everyone. So how do we do that when you add breast, does that mean that everybody is going to, what, get a breast biopsy?

That's probably not going to happen, so are you going to take just the positives and go to breast biopsy and the negatives you wait a year or two to see if they declare themselves and they're declared negative? And then do you do that again for, you know, lung? Do you do the same -- do you follow the same process in lung, do you follow the same process in cervical? That's what I'm struggling to understand, how that cohort design actually works even for the four screened cancers, right, because how do we get truth on everybody?

DR. BLOOMQUIST: Do we have any comments, quick response? We're running short on time, I apologize, so -- but just on -- let's say if it was just limited to the four screening cancers alone, how would we get truth on everybody here?

DR. BERRY: So that's definitely an important issue. I view it as sort of subsidiary to the bigger question we're talking about. I think we could handle those, it's going to take a lot of writing the protocol and the protocol may be, you know, big.

DR. BLOOMQUIST: Gene.

DR. PENNELLO: To me, you would confirm the negatives with the length of follow-up equal to the screening interval of that standard of care modality.

DR. SKATES Yes, that's what I said before, that we do a follow-up on everyone, clinical outcome, and then it's the pathology on those who undergo surgery is the gold standard. So I don't think you can do much better than that in a screening situation.

DR. BLOOMQUIST: Thank you. So I just want to -- there were some technical difficulties, but I think we did make some good progress on, at least, possible study designs in the area of the cohort area. I'd like to leave each of the panelists maybe 1 minute for just any general piece of advice. Whatever trial you're going to run is going to be very large, it's going to be in the maybe four to five/six digit range, just maybe any advice on running very large trial planning and sort of thoughts going into it.

DR. DOUBENI: I won't talk about trial, but I think the thing again I'll go back to is that there's a window of net benefit in any clinical -- in any screening test and that needs to be considered in any design and study that is done. Again, I think the struggle is going to be separate windows with standard of care and those without it, and at the end of the day I think you end up having to look at these cancers one at a time and that should be the way things should start and then over time you can pool them together. Ultimately, you need randomized controlled trials to demonstrate benefits. I would encourage folks to look at the recommendation, the task force recommendation on pancreatic cancer screening, which was a D but we did offer up some ideas about studies in populations which are similar to the ones we're discussing here and that could be the target of investigations to do.

DR. BLOOMQUIST: Don, any last comment on maybe, you know, planning or sort of planning for, let's say, a large, even a randomized trial or a cohort trial, things people

should consider going into it?

DR. BERRY: So it takes money and it takes the government and all we need is for -- to give up on the wall between us and Mexico and we could do it.

DR. BLOOMQUIST: Gene, any comments on sort of planning or advice on, let's say, you know, this would be a very large trial.

DR. PENNELLO: Well, one thing I know about randomized trials when you're trying to evaluate diagnostic tests together with treatments is you could randomize patients to whether to use the test or not or you could change the point of randomization to what to do with a test result and this is -- and, you know, so if you're diagnostic test positive you can randomize those patients and there's some literature to suggest that's a much more efficient way to do randomized clinical utility trials by Patrick Bossuyt and others.

DR. BERRY: So what it's going to take, Erik, in my view, is greater understanding of and publication of results and I've seen the Hopkins stuff, I was a co-author on the GRAIL stuff. I mean, for example, these are identifying cancers in people we already knew who had cancers. And that's, you know, really iffy. It's like several steps away from where we have to be. What we need is real sexy results and, you know, showing that you're better than mammography.

And by the way, there -- Sue Love, for example, is very enthusiastic about and has been looking for a blood test that will replace mammograms. And so if ctDNA does that, you know, it could be very much appreciated by and used by women. If it turns out that there is something that GRAIL or others do studies, Hopkins, do studies and show that they can identify tumor types really well with their assay and that generates some enthusiasm in, you know, for me --

DR. BLOOMQUIST: Sure.

DR. BERRY: -- which is not an easy thing to do, as you gather, and others like me,

they're skeptics, that is what it would take. And then the government would get behind it and it's going to take a village and we have more than a village here, at least so far until corona does us in.

DR. BLOOMQUIST: So just to -- maybe any -- each of the speakers on the phone, you have maybe, let's say a minute, just any general advice on running these large trials. Let's start with Dr. Skates.

DR. SKATES: So I wanted to throw out a design that might be of interest and might reduce the sample size, which is to essentially screen everyone, but screen Arm A for half the cancers and screen Arm B for the other half of the cancers and then use the retrospective results to look at positive cases in the control arm versus those that were intervened in on the opposite arm for each of the cancers, and that way you could cut down on the sample size considerably by having a blood draw on the two random -- on both randomized arms.

The only downside there is that you screen for half the -- half of the cancer sites in one arm and half the other cancer sites in the other arm. I don't see that as a big downside and the sample size would be significantly reduced in that method. You have more information because you got a blood draw on everyone.

DR. BLOOMQUIST: We need to -- and so that's very intriguing and so what you would do is have cancer mortality for the first half, cancer mortality for the second half, and you -- you know, use both to make both comparisons, right? One would be the control for the other. So it's intriguing.

DR. SKATES: That's right, yeah. So I want to throw that out as a way, and I'd still like to push on reduction in late-stage disease to be considered as something that would be of clinical utility, absolute reduction, and that would reduce -- and the FDA might consider that sufficient and then you might want to ask them to do a postmarket follow-up for mortality

but at least get them on the market in 5 years rather than waiting 20 years, which I don't see as a commercially feasible way, and the only way to get the 20-year study done is with government involvement. And I'd like to see more options than that, you know, given that we've had one government study on screening, that's the PLCO, over the past 20 years or even 30 years. So, you know, I'd like to push that change in endpoint and how you triage it before they can get on the market from mortality reduction to late-stage specific incidences --

DR. BERRY: I don't think it's -- I don't think it's sufficient, I would be very much opposed to it.

DR. BLOOMQUIST: We need to quickly move on.

Colin, any last minute advice on running a possibly large six-person study?

DR. BEGG: I'll be brief. Can you hear me?

DR. BLOOMQUIST: Yeah, we can hear you. Go ahead, Colin.

DR. BEGG: Yeah, I think that, you know, we've been discussing a very promising area of investigation, it's challenging but it's an important issue. I think that what this session has identified is that there's lots of issues, challenging issues, not just with randomized trials, but with the case cohort ideas that we discussed earlier. I get the sense that there's broad agreement on most of these issues, although there's some disagreement on individual issues and -- but it would be much more fruitful and productive to be focusing in on actual concrete ideas about particular study designs so we can evaluate and criticize. Hopefully, what we've learned from this session is how to go about doing that.

DR. BLOOMQUIST: And lastly, Constantine, did you have any advice on running a possibly five to six digit person study?

DR. GATSONIS: Well, having run studies like this, the advice is you've got to be very patient and it takes a long time and resources. But I would say the diagnostic accuracy

studies are feasible in this space and there are various ways of doing them and they're not anywhere near as labor intensive and so on as studies of outcomes. When it goes to the studies of outcomes, then first you get realistic about how that mortality, for instance, outcomes are not going to be available. Outcomes like advanced stage of disease, I already tried, but even there the studies are large. Within a large study of tomosynthesis right now that has announced, that's available in four and a half years but it still requires more than a hundred thousand patients. So it's not realistic to do all this.

I would just say, as a follow-on, that for those who want to make claims that ctDNA is useful today as a screening device for X-Y cancers, they have to be prepared to provide the evidence for it. They cannot just say because this is a new thing that that's the way to go. So maybe, maybe the place to start looking is not in the screening context, but to look at it in other uses, for instance, in the surveillance context and so on and back into screening from there rather than just saying we'll go for the big enchilada right now when there's zero evidence and it would take a long time to develop the evidence for this.

DR. BLOOMQUIST: Okay, thank you.

So I just want to thank each of our six panelists, I mean, this was a very interesting discussion. I'm not sure we came to much conclusion. I think even the endpoint could probably go a whole other 2 hours and probably could continue through the next day. So just want to thank the panelists, a very interesting discussion, lots of problems still to be outstanding and that should be an interesting time. So I think Reena will come up with some final comments and then we'll conclude for the day. Thank you.

(Applause.)

DR. PHILIP: Thank you, all. Thank you for staying until the end. So as you all heard, this is an exciting new technology with great potential and it's going to be a paradigm shift in cancer screening. I'm going to just summarize what I heard and I'm sure I'll miss a lot of

it, but we have the transcript coming up so at least, I will cover just some important points, what I heard.

So Session 1, moderated by Dr. Ghosh and Dr. Abukhdeir, along with the panelists, discussed several advantages and disadvantages of the multi-cancer screening test. As you all know, advantages are, it's a lower risk test, it's noninvasive and it's fast but it's got limitations, right, with ctDNA shedding issues with the multi-cancer, depending on the tumor type and then the CHIP mutation, especially with aging. But, one aspect which we all know is that the adoption is low for the current recommended screening methods, due to different reasons like infrastructure, stigma, but then blood tests have attractive feasibility for this effective screening test with high sensitivity. And so if it's an effective test, it could lead to adherence to recommended screening methods.

But there were concerns about the convenience of ctDNA screening tests replacing the currently recommended screening methods. And there are several good screening tests that can reduce mortality, so if there are existing recommended screening tests, what was recommended was a multi-cancer test should have a very high NPV to opt out of the recommended screening method, but if there is no recommended screening approach, like pancreatic cancer that is opt in, then PPV becomes very important.

So if there is a recommended screening approach, the recommendation was to benchmark the new technology to the old technology to see whether multi-cancer tests perform compromisingly. And as you look into the pretest probability and how will this test alter post-tests, and after the post-tests, will lead to any action. And, of course, high level of analytical validity and clinical validity was stressed many times. The technology can be very disruptive, but we should make sure there is benefit with reduced harm. And we heard different trial designs in the Session 1 but what was mentioned was the prospective trial in average-risk population. And moving on to, of course, the case controls to set the

biomarkers and then nested case-controlled study in a cohort population, like prospective studies. We also heard about intended use and intended outcome.

Okay, moving on to Session 2 - clinical validation session moderated by our great medical officers, Dr. Jha, Dr. Pathak, and Dr. Seidman, and with our great panelists that talked about the tumor of origin, and the consensus was that a multi-cancer test without a tissue of origin is not likely to be useful. We also heard several concerns about whole body imaging, compounded radiation exposure can lead to cancer and also these methods do not pick up small lesions. So without a tumor of origin component, there is more risk so there is additional imaging, additional invasive procedures.

Differentiating on the detected cancer versus incidental cancer, was also was brought up. So if the test gives a positive result, you should be looking for what the test was identified, which means if the tissue of origin is not a hundred percent accurate, still there may be biological segments that could be sorted out. It's still more risky, if the PET CT is going in the wrong direction than what's the right cancer that should have been detected.

So there were questions about electronic health record. As of now, it's got some limitations that's administrative in nature, also its built-in hardwired bias. But FDA is interested in leveraging real-world evidence but reports are largely unstructured data and may be biased, so not maybe yet appropriate. And we heard about in every approach looking for a signal versus noise and in an observational study there is more noise than signal, so we have to be prepared for study designs that detect relatively small signals. And also, we heard many times about over-diagnosis, we have to think of over-diagnosis, curing people that never intended to be cured. And so we need to figure out where to put these blood tests in the process. We still don't know the impact of not identifying precancerous lesions, could cause patients anxiety and so it should be narrowed down to specific tumor types. And so it's always good to have a comparison, to see added value. With respect to

highest and average risk, it was noted that it should be looked at separately and shouldn't extrapolate from high risk to average risk.

And screening test validation is an iterative process and should start with population at higher risk, screening asymptomatic patients to determine if a test has benefit. And for low-prevalent cancers, it would be hard and needs a very large trial and could use nested case-control study.

Moving onto third session, a study design session, it was hard with all the technical difficulties, but the moderators and panelists handled it well. I thank the moderators and the panelists. So the first question was should the primary analysis -- oh, sorry. I didn't mention our moderators. Moderators were Dr. Wei, Dr. Erik Bloomquist, and Dr. Pathak. And the great panelists.

So the first question was should the primary analysis for diagnostic accuracy be based on per cancer type and the panelists thought it depends on the question, but in the scenario it looks like we should look per cancer and it's critical to analyze for each cancer type since each cancer has different benefit-risk ratio. Pooled analysis was brought up and it could be looked at was discussed. Some panelists thought there is some rationale in looking into pooled analysis. And how to validate the diagnostic accuracy for cancer types with recommended screening tests, most panelists thought for those it could be compared to the current gold standard. Cross-sectional study was mentioned, similar to what was done for the single cancer screening FDA-approved tests. And randomized control trial was also brought up if clinical utility has not been shown, but wholistic evaluation of the test was something that was brought up, and looking into a tradeoff between clinical benefit and harm, and paired design we should look into, so sensitivity and specificity for evaluating diagnostic accuracy. So for the ones which don't have recommended screening approach, a large cohort study was mentioned, but the question was how you can power the study in a

way that has minimum sites for cancers of interest, and so I guess the approach is to follow prospectively and the actual number so that way the actual number will be much higher if you look at multi-cancer in that scenario, so maybe looking at a wholistic approach. And also, was mentioned, about supplementing with case control enrichment, such as the age enrichment. Anyway, there is no one-size-fits-all for multi-cancer screening.

With that, the workshop is adjourned and I thank all the FDA speakers and FDA folks who really helped putting this together in this little bit crisis situation and also the panelists, who flew in from different places and those who provided public comments and to everyone who came here in person and those who called in. Thank you, all.

(Applause.)

(Whereupon, at 4:07 p.m., the meeting was adjourned.)

C E R T I F I C A T E

This is to certify that the attached proceedings in the matter of:

PUBLIC WORKSHOP - DETECTING CIRCULATING TUMOR DNA FOR CANCER SCREENING

March 9, 2020

Silver Spring, Maryland

were held as herein appears, and that this is the original transcription thereof for the files of the Food and Drug Administration, Center for Devices and Radiological Health, Medical Devices Advisory Committee.

_____

TOM BOWMAN

Official Reporter

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947