UNITED STATES OF AMERICA

DEPARTMENT OF HEALTH AND HAN SERVICES

FOOD AND DRUG ADMINISTRATION

+++

CENTER FOR DEVICES AND RADIOLOGICAL HEALTH

+++

MEDICAL DEVICES ADVISORY COMMITTEE

+++

GENERAL AND PLASTIC SURGERY DEVICES PANEL

+++

JULY 29, 2022

9:00 a.m. EST

+++

Webcast via Microsoft Teams

PANEL MEMBERS:

Hobart W. Harris, M.D., M.P.H.	Voting Chair
Karla V. Ballman, Ph.D.	Voting Member
Mary H. McGrath, M.D., M.P.H.	Voting Member
Murad Alam, M.D.	Voting Member
Karen E. Burke, M.D., Ph.D.	Temporary Non-Voting Member
Paula Bourelly, M.D.	Temporary Non-Voting Member
Maral Skelsey, M.D.	Temporary Non-Voting Member
Paul Pisarik, M.D.	Temporary Non-Voting Member
Maria Suarez-Almazor, M.D.	Temporary Non-Voting Member
Neil Farber, M.D.	Temporary Non-Voting Member
Renata Block, PA-C	Temporary Non-Voting Member
Laura P. Bush, DMSc., PA-C, DFAAPA	Temporary Non-Voting Member
Lisa Gualtieri, Ph.D., ScM.	Temporary Non-Voting Member
Steven J. Skates, Ph.D.	Temporary Non-Voting Member
Katalin Roth, M.D., J.D.	Temporary Non-Voting Member
Veronica Rotemberg, M.D., Ph.D.	Temporary Non-Voting Member
P. LaMont Bryant, Ph.D.	Industry Representative

Deneen Hesser, MSHSA, RN	Patient Representative
Candace Nalls	Designated Federal Officer

FDA Representatives – Silver Spring, MD:

Long H. Chen, M.D. — CDRH/OPEQ/OHTIV
Acting Division Director Division of Health Technology, IVA

Binita Ashar, M.D. — CDRH/ODE
Division Director, Division of Surgical Devices

FDA Presenters:

Ryan Ortega, Ph.D.
Colin Kejing Chen, Ph.D.
Henry Lee, M.D.
Neil R.P. Ogden
Scott L. Kominsky, Ph.D.

Summation Speakers:

Jianting Wang, Ph.D.

INDEX

MEETING

DR. HARRIS: [9:00 a.m.] I would like to call this meeting of the General and Plastic Surgery Devices Panel to order. I am Dr. Hobart W. Harris, the Chairperson for this Panel, and I'm a Professor of Surgery at University of California San Francisco. I note for the record that the members present constitute a quorum as required by 21 CFR Part 14. I would also like to add that the Panel members participating in today's meeting have received training in FDA device law and regulations.

For today's agenda, the Panel will discuss the risks and benefits of skin lesion analyzers for external use. The Panel will be asked to recommend the FDA whether SLAs should be down-classified from Class III to Class II, subject to general and special controls. The Panel will be asked to discuss the types of evidence, including clinical evidence that would be helpful to support certain indications, as well as appropriate special controls necessary to mitigate the risk to health and assure the safety and effectiveness of these devices. Before we begin, I would like to remind the public and Panelists that this is a non-voting meeting. I now ask our distinguished Committee members and the FDA attendees to virtually introduce themselves. Committee members, please turn on your video monitors if you have not already done so and unmute your microphones before you speak. I will call your name. Please state your area of expertise, your position, and affiliation. Carla Ballman.

DR. BALLMAN: Hi. I'm Carla Ballman. I am a Professor and Division Chief of Biostatistics at Cornell Medicine in New York City, and my area of expertise is statistics.

DR. HARRIS: Mary McGrath.

DR. MCGRATH: Good morning. I'm Mary H. McGrath. I'm a plastic surgeon and my position is that of Professor of Surgery Emerita at the University of California San Francisco in the Department of Surgery.

DR. ALAM: Good morning, my name is Murad Alam. I am Professor and Vice Chair of Dermatology at Northwestern University in Chicago, and I'm a dermatologist.

DR. BOURELLY: Good morning. Paula Bourelly. I'm a private practitioner in the area of dermatology, clinic dermatology, in Maryland.

DR. SKELSEY: Good morning. I'm Maral Skelsey. I'm a dermatologist and neurosurgeon in Chevy Chase, Maryland and Clinical Associate Professor of Dermatology at Georgetown University Medical Center.

DR. PASARIK: My name is Paul Pisarik. I'm a private practice board-certified physician in Tulsa, Oklahoma.

DR. SUAREZ ALMAZOR: Good morning, I'm Maria Suarez-Almazor. I'm a professor at the University of Texas, M.D. Anderson Cancer Center. I am a clinical epidemiologist and I am an internist in rheumatology.

DR. FARBER: I'm Neil Farber, Professor Emeritus of Clinical Medicine at University of California San Diego in the Division of General Internal Medicine, and my expertise is in general internal medicine, and I am a member of MDAC.

MS. BLOCK: Good morning. My name is Renata Block. I am a certified physician assistant specializing in dermatology. I work with Dr. Monica Rani in Chicago, Illinois in private practice.

DR. BUSH: Good morning. I'm Laura Bush. I'm a certified physician assistant practicing in dermatology in Fayetteville, Georgia.

DR. GUALTIERI: Good morning. I'm Lisa Gualtieri. I'm an associate professor at Tufts University School of Medicine in the Department of Public Health and Community Medicine.

DR. SKATES: Steven Skates, Associate Professor of Medicine and Biostatistics at Massachusetts General Hospital Clinical School. My expertise is in early detection of cancer.

DR. ROTH: Good morning. My name is Katalin Roth. I am an internist specializing in geriatric and palliative medicine, and a Professor of Medicine at the George Washington University in Washington, D.C.

DR. ROTEMBERG: Good morning. I am Veronica Rotemberg. I'm a dermatologist at Memorial Sloan Kettering Cancer Center and my expertise is dermatology, imaging, and informatics.

MR. BRYANT: Good morning. LaMont Bryant, Worldwide Vice President of Regulatory Affairs, Ethicon, Johnson & Johnson, and I'm the industry representative.

DR. HARRIS: Deneen Hesser.

MS. HESSER: Good morning. I'm Deneen Hesser, a long-term melanoma survivor and college nurse by profession, and I'm here as a Patient Representative.

DR. CHENG: Good morning. This is Long Chen, the Acting Director for the Division of General Surgery Devices, and that is in the Office of Surgical and Infection Devices within the agency.

DR. HARRIS: Binita Ashar.

DR. ASHAR: Good morning, everyone. My name is Binita Ashar. I'm a general surgeon. I'm a Director of the Office of Surgery and Infection Control Devices at the

Center for Devices and Radiological Health at the Food and Drug Administration. Thank you.

DR. HARRIS: Thank you all. Candace Nalls, the Designated Federal Officer for today's General and Plastic Surgery Devices Panel, will make some introductory remarks.

DR. NALLS: Good morning. I will now read the Conflict of Interest Statement. The Food and Drug Administration, FDA, is convening today's meeting of the General and Plastic Surgery Devices Panel of the Medical Devices Advisory Committee under the authority of the Federal Advisory Committee Act, FACA, of 1972. With the exception of the industry representative, all members and consultants of the Panel are special government employees or regular federal employees from other agencies and are subject to federal conflict of interest laws and regulations. The following information on the status of this Panel's compliance with conflict of interest laws covered by, but not limited to, those found at 18-USC subsection 208 are being provided to participants in today's meeting and to the public. FDA has determined that members and consultants of this Panel are in compliance with federal ethics and conflict of interest laws. Under 18 USC subsection 208, Congress has authorized FDA to grant waivers to special government employees and regular federal employees who have financial conflicts when it is determined that the Agency's need for a particular individual’s services outweighs his or her potential financial conflict of interest. Related to the discussions of today's meeting, members and consultants of this Panel who are special government employees or regular employees have been screened for potential financial conflicts of interest of their own, as well as those imputed to them, including those of their spouses

and minor children, and, for purposes of 18 USC subsection 208, their employers. These interests may include investments, consulting, expert witness testimony, contracts, grants, gratis, teaching, speaking, writing, patents, and royalties, and primary employment.

The Committee will discuss the possible reclassification of approved computer-aided melanoma detection Class III devices: (1) MelaFind, a device that uses multispectral imaging and was approved in 2012, and (2) Nevisense, a device that measures impedance and was approved in 2017. Both MelaFind and Nevisense devices are intended for use on cutaneous lesions suspicious for melanoma when a dermatologist chooses to obtain additional information when considering biopsy.

Dr. P. LaMont Bryant is serving as the industry representative acting on behalf of all related industry. Dr. Bryant is employed by Ethicon Inc., a subsidiary of Johnson and Johnson. We would like to remind members and consultants that if the discussions involve members or firms not already on the agenda for which an FDA member has a personal or imputed financial interest, they need to exclude themselves from such involvement and their exclusion will be noted for the record. FDA encourages all other participants to advise the Panel of any financial relationships they may have with any firms at issue. A copy of this statement will be available for review and will be included as part of the official transcript. Thank you.

For the duration of the general and plastic surgery devices meeting on July 29, 2022, Dr. Neil Farber, Paul Pisarik, Katalin Roth, and Maria Suarez Almazor have been appointed to serve as temporary nonvoting members. For the record, Dr. Farber serves as consultant to the Nonprescription Drugs Advisory Committee at the Center of Drug

Evaluation and Research, CDER. Dr. Pisarik and Dr. Roth serve as voting members in the Nonprescription Drugs Advisory Committee in CDER. Dr. Suarez-Almazor serves as a consultant to the Drug Safety and Risk Management Advisory Committee in CDER. These individuals are special government employees who have undergone the customary conflict of interest review and have reviewed the materials to be considered at this meeting. The appointments were authorized by Russell Fortney, Director, Advisory Committee Oversight and Management Staff on June 29, 2022.

Before I turn the meeting back over to Dr. Harris, I would like to make a few general announcements. In order to help the transcriber identify who is speaking, please be sure to identify yourself each and every time that you speak. The press contact for today's meeting is Audra Harrison. Thank you very much. Dr. Harris?

DR. HARRIS: Thank you, Ms. Nalls. At this time, the Panel will hear summations, comments, or clarifications from FDA regarding yesterday's general issue session. You have 10 minutes. Dr. Chen.

DR. CHEN: Thank you, Dr. Harris. This is Long Chen. FDA would like to make a presentation to summarize yesterday’s presentation on scientific aspects of the skin lesion analyzer discussion. Certainly, I want the Panel to understand today, our discussion is really a different object, which is focused on device classification. With that, going to introduce Dr. Jianting Wang for this presentation. Please.

DR. WANG: I will share my screen now. Good morning. My name is Jianting Wang, Director for Light-Based Energy Devices Team at the Office of Surgical and Infection Control Devices at FDA. Yesterday, we had a very informative discussion at the general issue meeting. But before I start today's meeting, I will give a brief

summation on the discussion we had yesterday on the topic of skin lesion analyzer technology and issues regarding its application in detecting skin cancers in various patient care settings. We presented several topics to provide backgrounds of the lesion analyzer technology and regulatory considerations, including an overview of skin lesions, current diagnostic approaches and treatments of skin cancer, lesion device landscape, diagnostic accuracy of skin malignancies by healthcare providers, and the options for ground truth considerations. We also presented benefit and risks of the skin lesion analyzers with considerations of prevalence and different populations. Our outside speakers, Dr. Cohen and Dr. Adamson, gave talks on the challenges and possible solutions with artificial intelligence and machine learning algorithms for skin cancers, especially issues of potential bias and disparity for different ethnic groups given the different epidemiological characteristics in these groups. In the Public Hearing Session, I heard voices from practicing dermatologists, and also a Nevisense user, Dr. William Steffe, and Mr. Simon Grant, who is the CEO of SciBase for Nevisense, regarding the use and patient risks of skin lesion analyzers, key considerations in performance evaluation of skin lesion analyzers, and the importance of adequate regulations of these devices. In the afternoon sessions of Panel deliberation and FDA questions, issues regarding evaluation and regulation of SLAs were further discussed among Panel members. On the question regarding option for ground truth, Panel generally believes histological diagnosis is required, while some believe alternative approaches can be valuable depending on lesion type, alternate use of device, or clinical setting the patient population were specific study design. Regarding performance thresholds, Panel is in favor of a performance threshold that

shows the device improves the performance of the clinical user or improvements in patient benefits, depending on the particular use, condition with certain level of safety insurance. The sensitivity and specificity threshold should be higher for standalone devices compared to devices for adjunct use. The Panel highlighted the needs of other evaluation endpoints in addition to the binary sensitivity and specificity metric. Panel also discussed importance of prospective data from real-world use and post-market studies or surveillance. The Panel also discussed the impact of false-negative and false-positives for different cancer types in different use scenarios, especially for laypersons. For the performance in U.S. populations, Panel agreed that all skin types should be studied, but allowances can be given to some devices to market before adequate data are collected on all populations. Other strategies or measures can be used to promote continuing data collection in low-incidence populations, such as post-approval study requirements, requirements of transparency, and prevalence of data. On behalf of the FDA analyzing team, I would like to thank Advisory Committee, our external speakers, and everyone attending the meeting or submitted comments to us for your attention on this very important topic and for your valuable and thoughtful feedback. We had a very informative discussion and your input will be taken into consideration to inform FDA future decision-making regarding the regulation of this device type. Please note the docket for public comments on these topics is still open. If you have additional comments, please submit by August 29th. Your contribution is very much appreciated. Thank you. I will stop sharing my screen.

DR. HARRIS: Thank you, Doctors Chen and Wang. Before we move on, I would like to have Dr. Burke, Karen Burke, introduce yourself. … Well, perhaps we will return to Dr. Burke once we have her technical issues resolved.

At this point, we will move on to the FDA presentation. I would like FDA to start their presentation. I would also like to remind the public observers at this meeting that, while the meeting is open for public observation, public attendees may not participate except at the specific request of the Panel Chair. The FDA will have one hour and 10 minutes to present. FDA, you may now begin your presentation.

DR. ORTEGA: Hello, everyone. My name is Ryan Ortega and I'm a regulatory advisor within CDRH's Office of Product Evaluation and Quality, or OPEC. I will be providing a high-level overview of medical device classification and re-classification processes, which form the basis for our discussion today. The purpose of this Panel meeting is to seek the Panel's input and recommendations regarding the proposed re-classification of two Class III devices: optical diagnostic devices for melanoma detection, and electrical impedance spectrometers. FDA is proposing to reclassify these devices as Class II devices called computer-aided devices, which provide adjunctive diagnostic information about lesions suspicious for melanoma. Specifically, the Panel will be asked to discuss the available scientific evidence regarding these devices. The Panel also be asked whether they should remain in Class III, or if they can be re-classified to Class II.

So, let's start by explaining the different classes of medical devices. Devices are classified based on the controls necessary to mitigate the risks associated with the device type. Class I devices are only subject to general controls. Class II devices are

subject to both general and special controls. And Class III devices, are subject to general controls and premarket approval. These regulatory controls will be discussed in greater detail in the following slides. Importantly, a device should be placed in the lowest class whose level of control provides a reasonable assurance of safety and effectiveness.

Now, I'll into a bit more detail about each of the classes. Again, Class I devices are those devices for which general controls are sufficient to provide reasonable assurance of the safety and effectiveness of the device. General controls are basic requirements that apply to all medical devices and are outlined in the Federal Food, Drug, and Cosmetic Act. Some examples include meeting establishment registration and device listing requirements, following good manufacturing practices, adhering to recordkeeping and reporting requirements, and ensuring that devices are not misprinted or adulterated. Most Class I devices do not require FDA premarket review prior to being marketed. On the right hand of this slide, you can see a few examples of Class I devices. These include hospital beds, bags, and certain manual surgical instruments. There is also an alternative pathway to determine the devices Class I. Class I devices could also be devices that cannot be classified into Class III because they are not life-sustaining, life-supporting, or of substantial importance to preventing impairment of human health. They do not present potential unreasonable risk of illness or injury, and these devices cannot be classified into Class II, because insufficient information exists to establish special controls to provide a reasonable assurance of safety and effectiveness.

Class II devices are those devices which cannot be classified into Class I, because general controls by themselves are insufficient to provide reasonable assurance of safety and effectiveness for the device, and there's sufficient information to establish controls that provide such assurance. There are many types of special controls, but some examples include performance testing, sterilization validation, or device-specific labeling requirements. These special controls, in combination with the general controls previously described, provide a reasonable assurance of safety and effectiveness for Class II devices. Examples of Class II devices include nasogastric feeding tubes, semi-constrained metal and polymer knee replacements, and surgical sutures. Typically Class II devices require a premarket notification, which is generally referred to as a 510K submission, prior to being marketed in the US. Within these 510K submissions, companies must also provide evidence demonstrating how the special controls for specific device type are met.

Class III devices are those which cannot be classified into Class II because: insufficient information exists to determine that general and special controls are sufficient to provide reasonable assurance of safety and effectiveness and the devices are life-sustaining or life-supporting, or they are of substantial importance of preventing impairment of human health, or they present an unreasonable risk of potential illness or injury. Class III devices typically require premarket approval, or PMA, prior to being marketed. Examples of Class III devices include pacemakers, vascular stents, or implanted urinary and fecal incontinence devices.

Here, you can see a flowchart that walks the general decision-making process for each of the device classes that was just discussed. We start with determining whether

general controls are sufficient. If so, the device can be appropriately regulated in Class I. If not, we ask whether there is sufficient information that allows us to develop special controls. If so, the device can appropriately be regulated in Class II. If not, it would then be Class III if the device is life supporting or life-sustaining, or if it's of substantial importance to preventing impairment of human health, or if it presents a potential unreasonable risk of illness or injury. If the device isn't life-supporting or life-sustaining or of substantial importance of preventing impairment of human health, and it also does not present a potential risk of illness or injury, then we end up back at the Class I designation.

Now, we will shift our focus specifically to the discussion of reclassification of optical diagnostic devices for melanoma detection and electrochemical impedance spectrometers (EIS). These are proposed to be re-classified as computer-aided devices which provide adjunctive diagnostic information about lesions suspicious for melanoma. So, what is the process for reclassification? The decision to initiate this process is based on new information about the device, either on FDA's own initiatives, or upon the petition of an interested person. When going through this process, FDA considers uses which have been reviewed in the context of premarket review. In this case, FDA is proposing to reclassify the subject Class III devices on our own remission.

The first step in the process is to publish a proposed order announcing FDA's proposed classification in seeking public comment. This step has already been completed. The associated proposed order was published in the Federal Register on June 30, 2022, and it's being followed by a 60-day comment period. The second step is to convene Panel meeting to discuss the proposed classification. This step is being

completed today. The final step will be to consider public comments received and all available information, including the Panel recommendations prior to issuing a final order.

What we ask from the Panel today is to review and discuss available scientific evidence regarding the safety and effectiveness of optical diagnostic devices for melanoma detection and electrical impedance spectrometers. The input and recommendations from the Panel should include an identification of the risks to health presented by these devices. A discussion of whether the devices are life-supporting, life-sustaining, of substantial importance of preventing impairment of human health, or if they present a potential unreasonable risk of illness or injury. The Panel's activity today should also include a discussion of whether sufficient information exists to develop special controls, and identification of those special controls in a discussion of whether general controls are sufficient by themselves.

After this Panel meeting, FDA will consider all available evidence, including the info we received today from the Panel, along with any public comments. FDA will then issue a final order, which identifies the appropriate classification of the device. If FDA determines that these devices should be retained in Class III, devices which have already been approved through the PMA process can remain on market. If FDA determines that the devices can be properly regulated as Class II devices, however; existing devices may remain on the market, provided they meet the designated special controls, and may potentially be used as predicate devices for future Class II devices, at the same time and for the same intended use. Further details regarding the specific and limitation strategy will be outlined within the final order. I hope that this has

provided you with sufficient background to set the stage for the forthcoming discussions. Thank you for your time and attention.

DR. CHEN: Good morning. My name is Colin Kejing Chen, Team Leader for Diagnosis and Treatment Devices Team in the Office of Surgical and Infection Control Devices at FDA. Today, I will cover the computer-aided devices FDA has approved for adjunctive use in assessing lesions specific for melanoma. To date, FDA has approved two Class III devices. Before getting into the details of devices, it will be helpful to provide an overview of the device approvals.

MelaFind was submitted to the FDA as a PMA and was considered a normal device, whose benefits and risks were not well understood because of its use for melanoma and the concern regarding the consequences of delayed diagnosis. FDA convened a meeting of the General and Plastic Surgery Devices Advisory Panel to discuss the performance and the benefits and risks of the device. Based on Panel recommendations, a number of revisions were made to the product labeling, and the device was approved in 2011. Nevisense was approved in 2017. It was not reviewed by Panel. In the years since MelaFind and Nevisense were approved, minor hardware and software changes have been made and approved, but the intended use for core technologies had not changed. To our knowledge, MelaFind is not currently being marked.

The first approved computer-aided device was MelaFind, which is a non-invasive diagnostic device for use on lesions suspicious for melanoma. It consists of a handheld imager that acquires initial images using light at wavelengths from 430 nm to 950 nm near infrared light. MelaFind captures images at each wavelength. The images are

analyzed through a proprietary artificial intelligence, machine learning-based algorithm to assess the degree of three-dimensional morphological disorganization, and support the disorganization on a 10-point scale.

In the original PMA, the risk was twofold: images above a certain threshold resulted in the screen showing an output of 'MelaFind Positive,' indicating the lesion could have high dysplasia, or it could be melanoma. Below the threshold, the screen showed a result of 'MelaFind Negative,' indicating the lesion does not appear to have high dysplasia or to be melanoma. In subsequent supplements, the risk score was shown on the interface. The labeling and language approved for melanoma are important. The Panel that considered modifying the MelaFind was concerned about the risks of false-negatives and the potential for delayed diagnosis of melanoma. Therefore, the Panel recommended the device be limited for use of dermatologists when trying to decide whether or not biopsy a pigmented lesion. The complete language approved for describing the device is too lengthy to present this slide. It contains language intended to mitigate the risk of over-reliance on the devices.

The key points are as follows. The intended use of MelaFind is to provide adjunctive information to aid in the decision to biopsy a lesion the dermatologist has deemed suspicious for melanoma. The language includes certain limitations such as the body size and size of lesion on which the device can be used. The intended use statement also clearly indicates that MelaFind not be used to confirm diagnosis of melanoma and that the output of MelaFind is one element of overall clinical assessment. MelaFind development proceeded through 6 studies. The pivotal study, which led to device approval, was a prospective, multi-center, blind clinical trial that had

close to 1400 patients. The lesions to be included in the study were photographed under the clinical diagnosis, or level of suspicions for melanoma was recorded. MelaFind was then applied to each lesion, but the output was not revealed to the patient, the investigator, or the pathologist. All study lesions were biopsied and studies were reviewed by at least two pathologists and a core facility. Device accuracy was compared to the consensus of the study dermatologists. No adverse events were reported during the study.

The pivotal study for MelaFind had two primary aims. Primary aim one was to determine whether the sensitivity be at least 95%. The sensitivity was calculated to 98% under lower confidence margin of 9%. In AI-ML diagnosis, one looks at predetermined sensitivity of algorithm. The sensitivity that results is a function of the algorithm and cannot be independent. Therefore, the second aim of the study was to determine whether when the sensitivity was locked. Specificity was at least as good as the clinical trial investigators and a sensitivity of 98% MelaFind provided a specificity of 1.6, which was superior to that of study dermatologists. Under the second primary aim, MelaFind performed additional studies to assess how the device affected the accuracy of dermatologists. This approach is used for assessing artificial intelligence machine-learning basic devices in other diagnostic spaces, such as radiology. A classic reader study works like this. A reader, who is generally qualified to make a diagnosis, is shown in image and is then asked to enter a proposed diagnosis, or in this case when the lesion should be biopsied. A reader, who is generally a healthcare provider qualified to make a diagnosis, is shown an image and is then asked to enter and propose a diagnosis, or, in this case, what the lesion could be and when the lesion should be

biopsied. The reader is then shown the output for that MelaFind and is asked to enter the clinical decision again. If the device output altered their decision, and did so correctly, then the accuracy of the readers after seeing the device output should be better than their accuracy in forcing the approved. This is called an aided study.

In the unaided study for MelaFind, the reader study was a bit simpler. They compared the unaided accuracy of the readers, who were provided photographs of the lesions, dermoscopic images, and a brief history of the patient. The accuracy of the ones — that is to say it did not assess why there is output approved from the performance of unaided users. Only why there is a standalone output a bit more accurate than the users. In this reader study, the sensitivity, defined as the decision to biopsy, was 97% for MelaFind and 72% for dermatologists. The other difference, one that is equally significant, was the fact the providers in the reader study out-performed the device. MelaFind specificity was approximately 10% compared to 50% of dermatologists.

The General Plastic Surgery Devices Panel voiced significant concerns over the risk of false-negatives. Therefore, as already described, risk mitigations were implemented. We include clear discussion of device accuracy in the labelling. Use was limited in dermatologists and FDA approval was limited to providing adjunctive information in decision to biopsy. The manufacturer was also mandated to perform a post-approval study to assess device performance in real-world use. My colleague, Dr. Harry Lee, will discuss post-approval study in his talk later.

The second approved computer-aided device was Nevisense, which is an electrical impedance spectrometer. This device uses small pins applies to apply low

electrical current to a targeted area. It compares tissue impedance in the lesion and in peripheral skin. The signals are analyzed by a proprietary AI-ML algorithm to determine whether there are issues in the skin lesion that could be consistent with malignancy. The device shows a risk score from 1-10, where 3.5 or higher is considered positive. The positive and negative predictive values associated with risk.

Nevisense was approved with intended use similar to MelaFind, that it is intended to provide adjunctive information for a dermatologist's decision to biopsy a suspicious lesion. It should not be used to confirm diagnosis of melanoma. And deciding output is one element of clinical assessment. Malvehy et al. published their experience with Nevisense in 2014. This large study of nearly 2000 patients was performed much like the study I described for MelaFind. All suspicious lesions were photographed. The study device, in this case Nevisense, was applied, keeping the output obscured from the studies. Like the MelaFind study, all lesions were biopsied and histology was revealed by three central pathologists. No serious adverse events were reported. This study reported sensitivity of 96% and specificity of 34%.

In summary, MelaFind and Nevisense are computer-aided adjunctive diagnostic devices. They are the only devices FDA has approved for assessing lesions suspicious for melanoma. Their intended use is very specific. They are limited for use by dermatologists, and lesions specific for melanoma, and to provide adjunctive information for the dermatologists' consideration on whether to biopsy. Outside of melanoma, FDA has cleared AI and ML-based diagnostic devices as Class II for assessing potential malignancy in radiology and gastroenterology. Special controls

were developed for those devices, and those special controls will be reviewed in preparation for the proposed special controls we will discuss later.

Lastly, you may have arrived that they are our AI and out based smart phone apps and use outside the U.S. It is important to note that, in parts of the world where medical devices are recommended differently, some of those apps may be marketed without review of clinical trial data. Based on the available information for current devices and from growing information about the performance of AI/ML in diagnosis, they proposed to reclassify computer-aided impedance spectroscopy devices, that are intended for use by dermatologists to provide additional information to decide whether biopsy for lesions specific for melanoma, from Class III to Class II, with special controls. Next, you will hear from Dr. Harry Lee for post-marketing information for this device's tasks. Thank you very much.

DR. LEE: Good morning. My name is Henry Lee and I'm an oculoplastic surgeon and Medical Officer in the Office of Surgical and Infection Control Devices. Today, I will discuss the post-market safety and effectiveness data for computer-aided devices, which provide information about lesions that are suspicious for melanoma. As we have discussed this morning, MelaFind and Nevisense are the only approved Computer-Aided devices which provide objective information about lesions that are suspicious for melanoma. In order to fully understand the safety and effectiveness of these devices, which may aid the Panel in their consideration of the re-classification of these devices from Class III to Class II, data from post-approval studies, peer-reviewed literature, and medical device adverse event reporting databases and medical device

recall databases were evaluated. The MelaFind device was evaluated in PMAP090012, as described by Dr. Colin Chen.

One of the conditions for approval was a mandated post-approval study to assess how the device operates in real-world use. The study assessed two components: how accurate the device output is, meaning how sensitive to output MelaFind positive is for melanoma and high-grade lesions, and whether the output improves the diagnostic accuracy of the provider. We'll go through both components to illustrate the difference between an adjunctive devices for clinical decision and devices whose output are used as the standalone indicator. Because MelaFind was approved as an adjunct to clinical decision-making, we will discuss MelaFind's effect on provider sensitivity first. Dermatologists in the post-approval reader study were shown a lesion and then asked if they would biopsy it or not. After they recorded their decision, they were shown the MelaFind output. At the time of the study, the approved device had only two possible outputs: MelaFind Positive and MelaFind Negative. The dermatologists were then asked whether they would now biopsy the lesion. In the group of lesion the device called MelaFind positive, the investigator changed the decision from ‘do not biopsy’ to ‘do biopsy’ for 31% of lesions. Among the additional biopsied lesions, 7% were melanoma or high-grade lesions. This led to an overall increase of provider sensitivity of approximately 2%. This means that the device provided value to the provider and patient by increasing diagnosis of melanomas that might have otherwise been missed. However, it also led to additional biopsies for benign lesions that the provider had not wanted to biopsy prior to seeing the output. This is the type of benefit-risk balance that we are asking you to discuss, and if the

proposed special controls outlined mitigate any potential risks of the device used allowing for the re-classification of MelaFind from Class III to Class II.

The post-approval study also assessed the sensitivity and specificity of the device output. This component of the study assessed standalone device performance. The MelaFind system provided a positive output for 79% of all lesions. Of the MelaFind Positive lesions, 19% were histologically melanoma or high-grade lesions. This result shows low specificity for the device. As previously discussed, the low specificity was linked to a high threshold placed on sensitivity, at 95%, and the performance of the algorithm. A literature review was also performed to determine if additional post-market data is available to further evaluate the safety and effectiveness of these devices.

A clinical study in 2014 by Hauschild et al. evaluated the impact of MelaFind on biopsy decisions. In the study, dermatologists in Germany were asked to review 130 pigmented skin lesions. Via an online survey, the dermatologists were provided with visual images and clinical exam information, and were asked if they would biopsy the lesion. A total of 211 dermatologists completed the study. The dermatologists were randomized into two arms. Arm one dermatologists were given clinical exam information and three high-quality digital images. Arm two dermatologists were given the same information, as well as MelaFind results for each lesion. Results found that the dermatologists' sensitivity increased from 69.5% to 78%. Specificity, however, decreased from 55.9% to 45.8%.

Winkelmann et al., in 2017, published a review of 7 studies that evaluated MelaFind. The sensitivity of providers increased from 70% to 88% when they are provided with MelaFind results. Similarly, aggregate specificity improved by 6%, from

52% to 58%. Biopsy accuracy was also evaluated in subset of studies. MelaFind usage was found to increase biopsy accuracy from about 60% to about 70%.

Finally, a query of FDA databases was performed. FDA receives medical device reports from manufacturers, healthcare providers, and patients. The reporting system aids in establishing a qualitative snapshot of adverse events for specific medical device or device type. It is useful in detecting actual or potential device problems used in a real-world setting, including rare, serious, or unexpected adverse events, user error, and off-label use. There are limitations with the reporting system, which include incomplete, inaccurate, untimely, unverified, or biased data. Submissions of reports are voluntary; therefore, a lack of reports may indicate there were no adverse events or outcomes with device use. The Adverse Event Reporting Databases revealed no MDRs for either MelaFind or Nevisense. The Recall Database search revealed one recall from MelaFind due to a software change for the user interface which had not been approved. This recall affected a total of 65 units and ended via a supplement with the revised user interface, which was approved in 2016.

In conclusion, both premarket and post-market evaluations of the two approved melanoma detectors did not reveal any significant safety concerns. This lends further supports re-classification of these devices in Class II. Special controls will provide a reasonable assurance of safe and effective use of the devices, and will mitigate the risks to help. Thank you very much for your attention. Next, Neil Ogden will provide an overview on the device classification and re-classification.

NEIL OGDEN: Good morning, distinguished Panel members, and other important attendees. My name is Neil Ogden; I am the Assistant Director for the Cancer

Diagnosis and Treatment Team within the Office of Surgical and Infection Control Devices. We heard presentations on FDA classifications and details about two device types, MelaFind and Nevisense, today. I will briefly discuss device classification and then give an overview of the re-classification for computer-aided devices, which provide adjunctive diagnostic information about lesions suspicious for melanoma. As you heard earlier from Ryan Ortega, FDA uses three classifications for devices, based on information available for the technologies and their intended use. Class I devices have simple designs with well understood technologies. General controls are sufficient to regulate them. Class II devices are more complex technologies that are sufficiently understood so that special controls can be written. Class II devices often have general controls and special controls. Class III devices are the most complex technologies that are life-sustaining, life supporting, or not well understood, and therefore, special controls cannot be developed. When the device is not Class I or Class II, it is a Class III device requiring a premarket application.

One of the FDA's Center for Devices and Radiological Health's priorities is to strike the right balance between premarket and post-market collection. Safety and effectiveness data developed for a single device on its own is typically reserved for Class III technologies that are not well understood. This lack of understanding drives the need for detailed clinical performance data in the premarket to demonstrate benefits and risks of the device. FDA reviewed how these computer-aided devices, which provide adjunctive diagnostic information about lesions specific for melanoma, are used in clinical decision-making, the benefits and risks reported in the literature, how the performance of these device types is evaluated, and the information needed to

determine whether the device provides a reasonable assurance of safety and effectiveness. We have learned about significant advances in artificial intelligence and machine-learning technologies in recent years. We understand the importance of training and validation, the risks of false-positive and negative output, sensitivity, specificity, and user improvement performance measures. FDA should be consistent and is currently regulating other artificial intelligence and machine learning radiology at the gastroenterology adjunctive information for cancerous lesions under Class II with special controls.

After our analysis of the computer-aided diagnostic information about lesions specific for melanoma, we believe that special controls and general controls can sufficiently characterize the benefits and mitigate the risks. These device types provide additional adjunctive diagnostic information that dermatologists can use when deciding to biopsy a skin lesion suspicious for melanoma. We understand the risks of these technologies and their diagnostic information output, and we understand how to assess the performance of these technologies pre-market. FDA has the technical expertise to review intricate and complex technologies used to interrogate skin lesions specific for melanoma. The risks of the output diagnostic information are the false-negative and the false-positive results. The greater risk is the false-negative output. It could cause delayed treatment of the lesion. Delayed treatment could result in delayed treatment, with additional patient morbidity and mortality. The false-positive output could cause unneeded treatment, like a biopsy or excision of a lesion.

Assessing performance for these devices includes using valid scientific evidence to describe the technical aspects of the device for safety, and to determine the

sensitivity and specificity for the device performance in the intended population, assuring that the sensitivity and specificity are high enough to provide benefits to the user that outweigh the risks and the device technology is safe to use. We have identified appropriate special controls to ensure FDA has sufficient information to assess performance. You will hear more about the proposed special controls from Dr. Scott Kominsky later today. FDA is proposing to create a separate classification regulation for computer-aided devices for adjective diagnostic information about lesions suspicious for melanoma that will be reclassified from Class III to II. Under this proposed re-classification, these computer-aided devices, which provide adjunctive diagnostic information about lesions suspicious for melanoma, will be prescription use. FDA has determined that premarket notification is necessary for these devices to provide a reasonable assurance of the safety and effectiveness. In the next presentation, you will hear from Dr. Scott Kominsky about the proposed re-classification and regulatory special controls. Thank you for your attention.

DR. KOMINSKY: Good morning. My name is Scott Kominsky, and I'm a Biologist Lead Reviewer in the Cancer Team within the Office of Surgical and Infection Control Devices. This morning, I'm going to be sharing some information regarding the proposed re-classification and regulatory controls for currently approved computer-aided devices, which provide adjunctive diagnostic information about lesions suspicious for melanoma. As you heard previously, FDA considers a device Class II when general and special controls are sufficient to provide a reasonable assurance of its safety and effectiveness. For computer-aided devices, which provide adjunctive diagnostic information about lesions suspicious for melanoma, FDA believes the available

evidence, presented earlier, suggest that special controls can be used to provide a reasonable assurance of safety and effectiveness. That is to say when used the device properly, the probable benefits to health outweigh the probable risks. There is not an unreasonable risk of illness or injury, and the device will provide clinically significant results in a significant portion of the target population.

Potential special controls employed by FDA may include performance standards, performance testing, post-market surveillance, patient registries, development and dissemination of guidelines, recommendations, and other appropriate actions as the Commissioner deems necessary to provide such assurance. In general, FDA proposes that special controls for computer-aided devices, which provide adjunctive diagnostic information about lesions suspicious for melanoma would include both performance testing elements, as well as labeling requirements. When evaluating the adequacy of special controls, it is important to understand that FDA relies on the ability of each special controls identified to mitigate an identified risk to health.

The first risk I would like to address is that a false-negative and false-positive results. False-negative results may lead to delayed diagnosis, which may result in poor disease outcome, while false-positive results may result in the increased use of healthcare resources and unnecessary medical procedures. These risks can be mitigated through clinical performance testing, which may include, for example: testing whether standalone device output meets acceptable performance thresholds, such as sensitivity and specificity; side-by-side comparisons; and/or a reader study to determine whether a device improves the performance of providers, as applicable. The clinical performance testing must demonstrate that the device improves assisted read detection

and/or diagnostic characterization of lesions suspicious for melanoma, compared to characterization of lesions without the device in the indicated user populations when used in accordance with the instructions for use. Nonclinical performance testing will also be needed, demonstrating that the device performs as intended under the anticipated conditions of use, including testing of safety features intended to mitigate device specific hazards, such as electrical, thermal, mechanical, or light-related hazards. The risk of false-negative and false-positive results can be further mitigated by special controls that require information in the labeling that provides detailed instructions for use and informs the user of the expected device performance on a data set representative of the intended population.

Similar to the risk of false-negative and false-positive results, improper use of the device or errors in its use, such as using it on an unintended lesion or patient, may lead to errors in patient management, resulting in poor disease outcome increased use of healthcare resources, and unnecessary medical procedures. This risk can be mitigated by special controls by requiring that the following information be included in the device labeling: the intended patient population, such as gender and Fitzpatrick skin type; the intended anatomical sites and types of lesions to be assessed; compatible imaging hardware; and compatible hardware for the device. Risk can be further mitigated through special controls that require the device labeling to inform users of foreseeable situations in which the device is likely to fail or not operate at its expected performance level. Improper device use or use error may also occur failure of the user to follow instructions for use or intended reading protocol. Mitigation of this risk can be accomplished by requiring that the device labeling include a device description and

information needed to facilitate the clinical interpretation of all device outputs, such as negative and positive results, and by special controls requiring that the device labeling provide a description of user training required prior to use. This risk can be further mitigated by special controls that require a human factors assessment. Human factors testing focuses on the interactions between people and devices: the processes performed by each and the interface between them. The main goal of the human factors assessment is to minimize use-related hazards and risks and confirm that mitigations are successful, such that users can use the device safely and effectively. In the present case, an assessment that examines the ability of attempted users to operate the device according to instructions for use following training, is proposed.

Another risk is device failure or malfunction, which may result in patient injury or errors in patient management. This risk may be mitigated by requiring non-clinical performance testing to demonstrate that the device performs as intended under the anticipated conditions of use, including testing of safety features intended to mitigate device-specific hazards. Risks can be further mitigated by software verification, validation, and hazard analysis. Some devices may also carry risks of electrical, thermal, mechanical, and light-related injury, which may lead to patient or user injury or discomfort. Mitigation of this risk can be achieved via special controls that require testing to demonstrate electrical, mechanical, and thermal safety. Software verification, validation, and hazard analysis, as well as device labeling that includes instructions on appropriate usage and maintenance of the device. The risk of eye injury due to energy exposure, such as light, can be mitigated by special controls requiring labeling that warns users about use on lesions close to the eye and unsafe exposure to the eyes.

Electrically-powered devices introduce an additional risk because they have the potential to produce radiofrequency or electromagnetic disturbances, which can interfere with the function of other devices within their vicinity. In the case of medical devices critical to life and health, such as implanted cardiac pacemaker and defibrillators, for example, interference from a nearby electrical-powered device can result in significant harm. This risk can be mitigated by requiring testing that demonstrates electromagnetic compatibility; that is, the ability of a device to function safely and effectively in its intended electromagnetic environment, without introducing excessive electromagnetic disturbances that might interfere with other devices.

The final two risks I would like to address are those of adverse tissue reaction, and of infection and cross contamination. Adverse tissue reaction refers to the potential for an unacceptable adverse biological response resulting from contact of the component materials of a device with the body. The risk of adverse tissue reaction for patient contacting devices can be mitigated by special controls that require elements of the device which may contact the patient can be demonstrated as compatible; that is, causing no unacceptable biological response. In addition, labeling will be required that includes, in addition to user qualifications needed for safe use of the device, instructions for maintenance, and validated methods and instructions for reprocessing of any re-usable device components. Reusable multi-patient use devices also carry risks of infection and cross-contamination. These risks may be mitigated by special controls that require validation of stabilization, shelf-life testing, and again, labeling that includes validating methods and instructions for reprocessing any reasonable patient-contacting components.

In summary, FDA considers a device class II when general and special controls are sufficient to provide a reasonable assurance of device safety and effectiveness. Following an assessment of public health risks associated with computer-aided devices, which provide adjunctive diagnostic information about lesions suspicious for melanoma, based on available public and nonpublic information, FDA identified risks, including: false-negative or false-positive results; improper device use or use error; device failure or malfunction; electrical, thermal, mechanical, or light-related injury; interference with other devices; adverse tissue reaction; and infection and cross-contamination. These risks may lead to injury of the patient or device user, as well as errors in patient management, resulting in poor disease outcome or unnecessary medical procedures. To address these concerns, FDA proposes special controls comprised of specific performance testing and labeling requirements. Each designed to mitigate each identified risk to health. When combined with general controls, FDA believes that these special controls can be used to provide a reasonable assurance of device safety and effectiveness, thus supporting a Class II designation. Thank you very much for your attention. In the next segment, we will hear clarifying questions from the Panel.

QUESTIONS FROM THE PANEL

DR. HARRIS: Thank you. Now, does any member of the Panel have any brief clarifying questions for FDA? Ms. Block.

MS. BLOCK: Thank you for the enlightening presentation and clarifications on the special considerations in regards to or special controls in regards to Class II. My

question for the FDA is, right now these SLAs are grouped for melanoma. My concern is the non-melanoma skin cancers. Are you taking these into account when you are suggesting re-classifying this to a Class II? My second question is, are you considering the quick technology advances that are happening on a day-to-day basis and how that will affect the efficacy of these products in the future?

DR. CHEN: Neil, you want to make some comments?

NEIL OGDEN: Dr. Harris, may I address the Panel?

DR. HARRIS: Yes.

NEIL OGDEN: Thank you. My name is Neil Ogden. Yes, we are well aware of the advancing technologies and the rapidly evolving advances in this area. But to address her first issue, which is other lesions. So, today, we are talking only about the Nevisense and MelaFind type devices, which are specifically indicated for use only on lesions specific for melanoma. So, today's discussion is just limited to those type of devices. Hopefully that answers part of your question. Thank you.

MS. BLOCK: Hi. Renata Block. Thank you for that, Dr. Ogden. My concern is if it is classified into a II, it really paves the way for future companies to maybe not have as stringent regulations regarding future advances. Not only melanoma, but non-melanoma skin cancer. That was the reason for the question.

DR. HARRIS: Thank you. Dr. Skelsey.

DR. SKELSEY: Thank you for those excellent presentations. My question is similar to Ms. Block's. We are looking at reclassifying currently approved devices. Does this re-classification extend to future devices with the same purpose? So, are we moving to reclassify devices that we are not actually discussing? Secondly, I wanted to

know the presented data on MelaFind sensitivity and specificity that; I think that data is about 10 years old. The device has been in existence for more than 10 years. I've used it extensively. I've used both Nevisense and MelaFind, but I have much more experience with MelaFind. My experience has been that the specificity is something that makes the device unusable because the high number of false-positives.

In addition, there are so many anatomic areas that the devices unusable. It's not only not labeled for use in certain anatomic areas, but it's also very cumbersome and can't be applied to several anatomic areas where you would want to look at skin cancers – anywhere in the skin. It can't be used around the eye, but can't be used in small places like in the conchal bowl, behind the ear, and web spaces. I have a significant concern about — it's very optimistic to talk about labeling and making sure that people don't use it in inappropriate areas, but I'm not as optimistic as perhaps the FDA is about what labeling means. What I'm getting to, though, is I wanted to know if the FDA has further information from real-world users. My real-world experience is that I abandoned the device because it wasn't functional in a dermatology office. So, what updated information do you have for real-world users? Thank you.

DR. CHEN: Dr. Harris, can I get to Dr. Neil Ogden again?

DR. ASHAR: Dr. Chen, if you don't mind I'd like to start answering the question and turn it over to Neil should there be remaining issues.

DR. ASHAR: Thank you so much, Dr. Skelsey, for your comments and feedback containing the device, one of the devices that's under consideration today. You know, this is why we are convening this. We want to understand if we are in the right place in our regulation and review of these technologies. Knowing what we know now, knowing

what you know now, could special controls be created to help address the safety and other issues that you are flagging? So, there is latitude in the special controls that may be created. So, if the Panel has particular risks that they are especially concerned about, they may contemplate the appropriate mitigation and that feedback would be especially helpful regarding what we need to put in place. This is a discussion about Class II and Class III, but really it's a discussion about the risks and mitigations that the Panel would recommend.

DR. SKELSEY: I would love to know what the recommendations would be to mitigate that risk of having a patient come into the office and feel that they're going to get a complete exam because they have an expensive device used on them. So, mitigation measures that are other than labeling, because I don't think that labeling would likely be sufficient. Obviously it's something for the Committee to discuss.

DR. HARRIS: Thank you. Dr. Farber?

DR. FARBER: Thank you very much for those presentations. My understanding – and I just need clarification and then one other issue – I just need clarification. My understanding is that these devices are approved only for use by dermatologists. I heard you say 'types of devices.' That, I assume, means that if other devices besides these two would come along and be in the same type, they would come under the same classification. That's my understanding and I just need to confirm that. But the other thing is, there were data presented about the standalone use of, I believe it was MelaFind. The question I have is: right now my understanding is that these devices are approved for use only by dermatologists. Is the Committee, the FDA considering approval for those other than dermatologists? I.e. non-dermatologic practitioners, as

well as the lay public. If so, would that have to be a separate approval, or would they automatically also come under the Class II classification?

DR. CHEN: Dr. Harris? Can I get Dr. Lee, Dr. Henry Lee, to make some comments?

DR. HARRIS: Yes, please.

DR. CHEN: Okay. Dr. Lee.

DR. LEE: I actually just have a quick comment with regards to the information that Dr. Skelsey had asked for. When it comes to evidence of effectiveness in real-world use, what we presented was pretty much what we found. And it's also important to know that MelaFind is currently not marketed anymore, so that seems to mirror your experience with regards to current use the US population or among dermatologists in the US.

DR. HARRIS: Okay. Dr. Rotemberg. --

DR. FARBER: Excuse me for a minute. My questions have not been answered.

DR. ASHAR: This is Binita Ashar. With respect to your question -- actually, I had the response. Do you mind restating your question?

DR. FARBER: Sorry, sure. The two questions are, first of all, what we are addressing today is MelaFind and Nevisense. My understanding is you are saying these types of AI's, so that this type of AI was — another manufacturer came up with a similar type of AI. They would be automatically under Class II, as well, if I'm understanding that. The other question I had is, right now these are used only by dermatologists. Are you contemplating that it would be used, in the future used by non-dermatologic practitioners, and perhaps laypeople as well? If so, would that have to be

reclassified and how would you assure that there wasn't any adverse events in those populations?

DR. ASHAR: Okay. Yeah. Thank you very much for restating the question. I think to make the Panel's job easier, it would be helpful to simply talk about the two devices at hand. And, you know, that they are devices adjunct for dermatologists in their diagnosis of a melanoma and keep the scope at that. If the Panel chair and others wish to address other situations, and that could be noted for the record, but our issue at hand is the classification of the two devices that have been presented. Now, when it comes to the various AI algorithms and other issues pertaining to risk and how those can be mitigated, we — AI technologies, machine learning exists in devices not just in skin lesion analyzers. The agency does have some experience developing special controls in this arena. If there are certain — in the realm of special controls, the Panel can be very imaginative. If you wish for post-approval studies to be done as a special control to address certain questions, you can communicate back to us. If you wish for certain safeguards to be present and noted during an iterations of these devices within this bracket when it comes to AI and machine learning, you can note those factors for consideration. We are simply — we don't have an opinion. We are proposing an approach for Panel deliberation about the benefits and risks of regulating these devices in this way. The most important thing is the special controls and the risks so that we understand in our regulatory framework how we can address the scientific risks that you are flagging. That’s a lengthy answer. Hopefully that helps clarify.

DR. HARRIS: Okay. Dr. Rotemberg.

DR. ROTEMBERG: So, one of the things that so different from dermatology as compared to radiology is that a radiology reader study is essentially a gold standard, because that is what the radiologist has seen. For dermatology, that's very different, and we saw a lot of data yesterday, and even today, to say that in a reader study, a dermatologist performs significantly worse than they do in real life. So, I guess I would love to hear from the FDA about the justification for why a reader study would be sufficient special controls for these devices.

DR. HARRIS: Anyone from FDA?

DR. ASHAR: I think again, we are not here to justify an approach. We are proposing a picture —

[Multiple speakers]

DR. ROTEMBERG: I'm sorry. I just mean, what is the basis for this?

DR. ASHAR: I think consistently we are always evaluating and re-evaluating the approach we are taking in regulating medical devices to make sure we are giving the appropriate oversight and focusing on the places where we need to focus. And this is simply an exercise in this regard. We recognize that this is a burgeoning area. We have two devices on the market. We are wanting to re-assess those, and yesterday we heard about other applications different from these two and how we are thinking about those. So, we are simply asking the Panel to contemplate whether it's Class II or Class III, what are the scientific issues, what are the risk mitigations, so that we can take that into account as we regulate these devices.

DR. HARRIS: Thank you. Dr. Alam.

DR. ALAM: My questions are similar to those already posed by Dr. Rotemberg. Unlike Dr. Rotemberg, I'm not an expert on AI or machine learning, but I do like to think I know something about skin cancer, which I've worked for about 1/4 of a century. In terms of the re-classification, with the proposed re-classification, I understand what you were saying, Dr. Ashar, that you don't have a strong horse in the race as to this presumptive idea, and you like our feedback on it. I'm still struggling with, why the urgency to do this given the relatively limited data that's currently available? Only two devices have been approved for adjunctive use. I understand there can be special controls and that the FDA has experienced in implementing those in the realm of AI devices, but I guess, based on what we've heard, I'm not convinced that the FDA, or anybody, has much experience in the context of devices for detecting melanoma. I guess that's my concern. I'd be interested in hearing the logic; and maybe there is none and that's why you're trying to ask what our opinions are. But it seems like these special controls are like a Rube Goldberg-like contraption where we don't even know what we don't know and what can go wrong, but we're going to have a lot of controls, and hopefully it will sort of work out. It seems to me a little premature given the lack of clarity in this entire space, but if I'm missing something, and there's good reason why this is imperative so early in the process, I obviously would like to be educated.

DR. CHEN: Anybody from the team?

DR. CHEN: This is Dr. Chen. So, first, I think we have presented a lot of information from the post-market, from the literature, also from the surveillance information. Science shows the risks should be able to be mitigated through special control. Also, about the need, as Dr. Alam pointed out — whether the timing of the

question, about the timing — it's always, like Dr. Ashar pointed out yesterday… there's should be a least-burdensome approach we should consider while we maintain a good quality to control all the potential risks. Also to ensure speedy innovations for future generations and so that good devices can go to a U.S. market in a timely manner, so that with all the information we have gathered about these two devices, also others in other AI/ML-based and software devices like for GI tract detection. Also, other reference information from OES and the regulations in their marketing, and we think that it should be ready to use special controls and to mitigate the risks, also to ensure least burdensome approach to regulation for future innovation technologies on the market.

DR. HARRIS: Thank you. Dr. Roth.

DR. ROTH: Thank you. Katalin Roth. Thank you for your presentations. I wonder if you can give us some information about whether this is a standard procedure at the FDA, after the experience with safety, to down-classify a device from a category III to category II, or whether this is a new procedure. And if it's been done before, could you give us some examples, please, other circumstances in which that's been done? Thank you.

DR. HARRIS: FDA?

DR. CHEN: Well, go back to Neil. Any comments on that?

NEIL OGDEN: Hi. Neil Ogden here with the FDA. I might leverage Ryan Ortega, but I do have personal knowledge of a number of times when the agency has done this. For example, hip implants and knee implants. They used to be Class III. We had enough experience and knowledge base to write special controls and went down-classify them. So, even though they are a permanent implant, excuse me, even though

they are an implant, and one may argue they have significant impact on one's life, we did down-classify knee implants and hip implants. So, it's a fairly common thing. The agency has been doing it for decades, so when we have enough information that we think we can regulate these devices in a lower class, we move to do so because we want to be least burdensome, like Dr. Chen was telling us, and that facilitates getting high-quality devices that are safe and effective to the US population in the least burdensome manner.

DR. ROTH: Thank you very much.

DR. HARRIS: Thank you. Dr. Bourelly.

DR. BOURELLY: Hi, Paula Bourelly. I just wonder, under the labeling mitigation measures, if the FDA considered in the proposal of labeling anything to actually instruct the provider how to protect the device? In other words, do you assume there will be any measure for ensuring that the device is still working properly after X number of treatments or uses? Thank you.

DR. ASHAR: This is Binita Ashar. I think that's a great recommendation and if the Panel would like to make recommendations along those lines, we would appreciate them.

DR. HARRIS: Thank you. Dr. Skates?

DR. SKATES: Steven Skates. Thank you very much for the comprehensive presentations. They were very informative. I do want to go back to the original reason for what I believe the Panel back in 2010 said, which was that because of the risk of false negatives, there's a significant safety concern, and this is the way they mitigated it. What I would like to understand – I didn't see any change in subsequent data to say that

this risk has changed or has lessened. And the safety issues that were addressed was whether the instruments caused an acute, I guess, adverse event on a patient. That was never, back into 2010, a Panel's concern. The concern wasn't the adverse events. It was the risk of falls negatives, and that hasn't changed.

The other issue that has been touted as a reason for going from Class III to Class II is the experience of AI in general. AI has clearly advanced and has made amazing strides. For example, computers learned how to play chess and then beaten all of the chess-playing engines in the world, from AlphaZero, for example. Those situations are where you have millions of data points on both positive and negative. They just played millions of games and they learned from that in the chess game. Or Google translate, for example, that's amazing. They've got millions of books in different languages to cross-compare. My concern in medicine is we don't have millions of cases. We maybe have hundreds or maybe thousands. In a complex situation like medical care, I think we are still very much at the beginning of our understanding of AI and interface between AI and medicine. The fact that AI has made advances in other areas, I don't think that's a strong enough argument within medicine, for down-classing III to II. I would love to see more of that.

My concern is, does downgrading from III to II remove, for example, intended users being dermatologists? Could it be that PCP's and even lay users get to use it because it's a Class II? And then the other big issue is, what is the additional burden anyway for leaving it a Class III compared to Class II? That's very vague in my mind. Those are the thoughts that are going through my mind at the moment with this request to downgrade it to Class II.

DR. CHEN: Can I get to Dr. Colin Chen first?

DR. ASHAR: I'm sorry, if you don't mind, I would like to address Dr. Skates' question, and if I leave something unaddressed, if you, Long Chen, or Colin Chen can step in, I'd appreciate it. To go back to the earlier question pertaining to the agency, FDA, in our Center, has an enormous responsibility in regulating all general and dermatologic infection control devices marketed in the U.S., particularly for this Panel. And that is just not talking about all of the other Offices within the Center. And so, we routinely assess, evaluate, reevaluate whether or not we are in the right place. So, just in the past year, we have been in a – not in the past year, but in the past few years – an up classification of surgical staplers. We discussed devices that are currently unclassified to see what the appropriate classification of those devices are, and this happens to be a proposed discussion around down-classification. So, this is normal business as usual for us. These are normal conversations around, are we mitigating the risks? Can we envision special controls or additional safeguards to address the scientific questions that we have?

Now, if the Panel has concerns, Dr. Skates, pertaining to spillage of the intended user moving from a dermatologist to other users, the Panel can make recommendations along those lines to say that, our comments today pertain to the two devices that we are contemplating. But for whatever reasons that you wish to communicate, we do not see this as appropriate for extrapolation to other users. So, that's up to you and we love to hear your feedback around some of those boundaries and concerns and considerations.

You also mentioned that in the MelaFind experiment, the post-market study was done to assess the risk of false negatives… perhaps that study, in retrospect, didn't

meet expectations. Knowing what we know now, knowing what the Panel knows now, perhaps you would have different recommendations. Those recommendations would be helpful to us because we want to do the right thing. We want to be well-informed. We have to start somewhere, and we did 10 years ago. Knowing what we know now, what should we do? With AI devices, there is capability within our special controls to manage concerns pertaining to artificial intelligence and machine learning. If there are specific concerns, I understand this is a unique case. No device area is just like any other device area. Each is unique, and so we respect that. That's why we have you here, for your expertise. To tell us, what special considerations unique to this area do we need to be mindful of? And so, there is some urgency, I guess, because people are asking these questions. You know, innovation continues. Medical device developers are anxious to plan how they might be able to develop the level of evidence that would meet your expectations. And so, the faster we converge on that, the faster we can start talking about the appropriate studies that are necessary to be able to move these devices forward in a safe and effective way. So, there is no urgency except one that is self-imposed to do the right thing for patients.

DR. SKATES: I guess my point was that the Panel in 2010 felt there was a significant risk, and I haven't seen any data from the FDA to change that significant risk and significant safety concern. I presume that significant safety concern was part of the reason that the FDA classified this as a device III. What I'd like to see his new data that mitigates that safety concern that has changed from 2010, and I haven't seen it yet — this safety concern being risk of false-negatives — being smaller than it was back in 2010. So, that hasn't changed, so why should our conclusions change from the 2010

Panel is essentially my question. And then the other question is the burden. What is the difference in burden to the company for a Class III versus Class II? That's very unclear to me.

DR. HARRIS: I think that these are all obviously excellent issues. I would like to try to confine this session to specific questions rather than the issues that we will have a chance to deliberate, and also reflect in the Committee's open discussion. So, if there is a specific answer to the second half of Dr. Skates' question, this would be a great time to hear that. Anyone from FDA be able to respond to that specific issue on burdens for close to versus Class III devices?

DR. CHEN: Yes. We can do that in afternoon session.

DR. HARRIS: Okay. Thank you. Next, I want to ask Dr. McGrath.

DR. MCGRATH: Yes. I noticed in our notes, and also briefly in your comments, that although this would be down-classified, you still would require a premarket notification for these devices and employ the 510K requirement. My question to you is: what is the FDA thinking when they decided to do that – which is a Class III requirement usually, but to keep it on these devices if they were Class II – are you thinking about future generations when you are saying this? In a way, this is kind of a special control, too, and I'm just trying to understand what you were thinking when you said, "Let's keep the PMA." This would be, I guess, for Dr. Ortega or Dr. Ogden.

DR. CHEN: Yes, Ryan. Dr. Harris, can I bring in Ryan from our team to address this?

DR. HARRIS: Yes.

DR. ORTEGA: Hi. Everyone. This is Ryan Ortega. Re-classifying the device to Class II would mean, for a premarket regulatory submission, it would go through the 510K process, which is also called premarket notification, which is separate from the requirement for a premarket application, or PMA, that a Class III device would go through.

DR. HARRIS: Thank you. Ms. Hesser, you had your hand up. You no longer have a question?

MS. HESSER: The question I was going to ask has been addressed, but as long as I have the opportunity, thank you for the helpful background from the FDA. It was my impression – and please correct me if I misunderstand this – that the one device with a market present currently is not supportive of down-classifying. So, what weight has the FDA given to that feedback? That would be helpful for me.

DR. CHEN: Neil, you want to make any comments?

NEIL OGDEN: Hello. So, I would like to try to redirect the Panel and reiterate what DR. ASHAR has been saying. We really want your input on suggestions on how to best regulate these devices. What special controls would you recommend that we use and why? What are your concerns about these devices and how do you think we should mitigate those risks? So, we proposed a number of mitigations and controls that we think would be adequate, and so we would love to hear your opinions about other ones that you think are necessary. Thank you.

DR. HARRIS: Thank you. Dr. Skelsey.

DR. SKELSEY: Thank you, it's Maral Skelsey. Talking about mitigations for the devices we are proposing: is it in the FDA's opinion that we are looking at these devices

for the lay population? Because it's still not clear to me whether or not you are proposing these changes, not just a re-classification, but using them, having the lay population utilize this. And we talked about how little data there is in terms of the re-classification for other devices. Dr. Ogden said that there was a lot of information, and that was why other devices have been reclassified. You are talking about something that's not on the market. How do you propose getting further studies if it's no longer on the market? So, what kind of studies would you think you can propose and then get some relevant data in the future if there are no users? Thank you.

DR. ASHAR: Okay. This is Binita Ashar. I want to go to Ms. Hesser's question and move to Dr. Skelsey. Ms. Hesser, we consider all feedback we obtain from everyone. We understand medical device review and regulation is not a unilateral sort of endeavor. We need your feedback. We need the feedback of all stakeholders, including the device manufacturer and others, patients, other groups. So, there's no particular algorithm or weight on what's very helpful is an understanding of what's missing. Often times, it's very easy to look on the page and understand everything adds up and makes sense, but what's missing?

I've heard a couple of times that perhaps what the concern is the application of the special controls to groups other than dermatologists. Knowing that is outside the scope of your discussion and you are thinking is helpful. Understanding why is even more helpful. If there are particular special controls that you'd like to see, and perhaps the language special controls is problematic, but if you have a particular concern as the device is being translated from the bench to the bedside and wish to see post-market studies, how would you want those post-market studies to be developed? Do you have

specific concerns or recommendations there? Hopefully that helps redirect the Panel to the issue we have at hand.

DR. ROTH: If I can address Dr. Ashar's response, and thank you for that clarity. One topic that's stood out for me in this discussion is that I have not seen a requirement for patient education. There is patient labeling. I'm not comfortable with the level at which risk is explained in that patient brochure. Perhaps looking forward, if we are going to help mitigate some risk, a little bit more robust patient education requirements would be helpful. Thank you.

DR. CHEN: Yes. This is exactly the comments we're looking for, letting us know what is missing in what we propose in the special controls. Thank you.

DR. HARRIS: Dr. Bush.

DR. SKELSEY: I just had a follow-up question. Regarding the lay user devices, is the FDA thinking they're going to be Class III or Class II? And secondly, how would you propose getting more data when something's off the market again?

DR. ASHAR: Okay. So, I think we — it's difficult to speculate how we would regulate devices that have not come into us. At this point, we are just looking for the Panel's input on the two devices we have proposed. Like I was saying before, if you wish to limit your feedback just to the two devices we proposed that are being contemplated for down classification in saying why our comments do not apply to comments for lay users, that's a different conversation with a different set of risks and risk mitigation. That would be helpful for us to understand. Hopefully that helps clarify. And with respect to premarket and post-market studies, yes, we are looking at this group of devices like the MelaFind device. We understand that's not currently

marketed, but if another device like MelaFind would come forward it to us, what would be the Panel's recommendations regarding the special controls to help mitigate risk for such a special device knowing what we know now?

DR. SKELSEY: So, we are going to be talking about, then, these devices that are not there yet? That's the question.

DR. ASHAR: Well, just because the manufacturer stopped marketing it, it's still classified as a Class III. So, for that reason, we have to talk about MelaFind and its down-classification. It still exists on our record as a Class III device, despite the company has taken a business decision not to market it any longer. Hopefully that helps.

DR. ASHAR: Dr. Bush.

MS. BUSH: Hi. Laura Bush. I have three quick comments. One is requiring compatibility with implantable devices… Is that just going to be in the location of use, or that the person just generally has an implantable device? So, if they have a defibrillator, is it going to be that they just have one, or that the lesion is overlying the area or close to the area? Second comment is on software updates. Because this is essentially software driven, is there going to be some sort of routine verification and validation at certain levels that are certain time periods under special controls? In the last thing would be, is there a special control to require that it be used by someone trained in dermatology? Those are my 3 comments.

DR. HARRIS: So, before we get a response from FDA, I want to remind the Panel that we have the opportunity to make any and all recommendations that we as a committee like rather than questioning the FDA about what the future decision my knee,

I think they are really just imploring us to provide guidance and insight as to what we think that decision may look like. Any responses from FDA?

DR. CHEN: Thank you.

DR. ASHAR: Yeah. I think that any recommendations that you have in any of the 3 categories you have outlined I think would be helpful.

MS. BUSH: Okay. Laura Bush. I would say in the region of an implantable device, I would feel uncomfortable using anything electrical over a defibrillator. And then also on software updates, we all know technology changes rapidly, and I would say that there should be some form of updates required in special controls at a regular interval. Possibly, I mean, I could throw out something every 3 years or whatever, but just something that states they have to update it at a routine basis, or at least check it. And then, as far as the special control trained on dermatology, I am trained on dermatology and I feel like this is an important matter, so I would suggest that.

DR. HARRIS: Thank you. I would like to try and conclude the portion of our discussion where we are asking brief qualifying questions for the next 7 minutes. So, brevity would be appreciated. Dr. Alam.

DR. ALAM: Thank you. I will try to be brief. My understanding of Class II to try to think of it in a different context, is it's something like making a drug generic, or something like that. Please forgive the analogy. I guess my concern is that — or a device generic, like an Nd:YAG laser that anyone can make. But the question is, or maybe the comment is, I'm not convinced that we are at the point where the technology is so well-defined that basically anything is similar. The example of the knee or hip implant, I mean, those are metal things you stick in there, and as long as it looks more

or less the same, they'll work more or less the same. But here, it's a black box, and the technology is rapidly evolving. We are not at a steady-state where a similar device will function in a similar way, and therefore, it's going to be inadequate for special controls. Thank you.

DR. HARRIS: Thank you. Dr. Rotemberg.

DR. ROTEMBERG: Thank you, and I know were going to discuss this a little bit later. I think one of the things I would say to Ms. Bush and Dr. Alam's comments to add to that, you know, we do have some information from the way AI technology works about data drift and changes in algorithm performance over time, and so I would add on to what Ms. Bush said, that updates should be targeted to when accuracy has fallen for certain benchmarks. And then additional validation studies be performed with additional software updates to bring the accuracy of the algorithm back to those that were approved. So that could be an opportunity for continued improvement and regulation of the algorithms after they are approved.

The other thing I would say, to Dr. Alam's point, is we do need a lot more transparency about the data that's used to validate these algorithms. We talked yesterday about impacts of race and skin tone, impact of gender and anatomic site. So, in order to really consider the use of these algorithms, we need a lot of transparency about how they were validated and in what populations that might address some of the concerns about the black box. At least we would know the accuracy under conditions that we already know affect the algorithm performance.

Finally, I would like to ask FDA – this is the question I was trying to get to, I promise, Dr. Harris. There is extensive literature on AI validation guidelines, including

at least two papers in dermatology. I wonder if it would be okay with the FDA if we revisit those papers this afternoon to try and guide some of this discussion, because I do think some key aspects from some guidelines that are written are missing from special controls.

DR. HARRIS: Okay. I will certainly investigate. I don't know exactly how we will handle that, but we'll get back to that in that regard. Next, Ms. Block.

MS. BLOCK: Thank you, kindly. Renata Block. I just have a quick comment Dr. Skelsey, you made a great point. I think the FDA should consider this in regards to the classification change is a real world abandonment of the MelaFind and really taken that data from physicians like yourself in consideration of why that was abandoned and getting more information in regards to manufacturing or stop manufacturing of the MelaFind altogether and adding that to special controls.

DR. HARRIS: Thank you. Dr. Skates.

DR. SKATES: I will try and be brief. I would like to see at least SLA devices being evaluated with the dermatologist – added to a dermatologist, and then the dermatologist without the SLA and to see a significant increase, at least a statistically significant increase, at least in one of the parameters. There's been both sensitivity increases in both the MelaFind — sorry, in the Nevisense, and decrease in specificity on the Nevisense. So, that prospect, such a prospective study… what I'm concerned about is, if we go down from Class III to Class II, the Class II doesn't require that. That's why what I would be interested in is understanding what the difference between a PMA requirement is. And I understand the two different classes and there's less burdensome

on the 510K, but other than that, I have no idea what that means. If it means removing such a prospect of study, I would be against downgrading.

DR. HARRIS: Can we get some clarification regarding that distinction? The 510K versus the PMA, and whether or not a 510K could be required to provide additional clinical data?

DR. ASHAR: This is Binita Ashar. I can address that and ask Ryan Ortega to fill in anything I may miss. Premarket clinical studies is considered performance testing data. As we review the 510K, we consider the least burdensome approach and get appropriate performance testing data. We first consider whether bench testing is acceptable. If that won't address our questions, we ask whether animal testing would suffice, and if that's not suitable, we move on to human clinical studies. An IDE study can be performed to support either a 510K or a PMA. I think many 510K's don't require a clinical study, but many do. So, that is one thing to note. Post-market studies can also be done for 510Ks. It's less frequent, but that could be a special control. Or, there's 522 studies performed for safety, so that's also possible.

So, fundamentally what the difference is, with the PMA, the data rests entirely on that device. Everything is about that device. None of that data can be leveraged for another device to support its substantial equivalence. The PMA data rests on its own for each and every device. With the 510K, there is capability for the manufacturer, say, of a particular device as they are making an iterative change. If those iterative changes rise to the level of warranting a new 510K, it could claim substantial equivalence with some of the changes they are proposing, and other manufacturers could leverage the predicate device from another manufacturer, claim substantial equivalence. But the

new manufacturer, or any, would have to abide by the special controls that this group is recommending. So, that's the fundamental differences and easiest way I can communicate it, but I'm going to ask Ryan Ortega to comment on anything else I should add, to make sure I'm in complete alignment with FDA's regulatory policies.

DR. ORTEGA: Thank you. Ryan Ortega, FDA. I think that was spot-on, and I will just reiterate: as far as options for special controls are concerned, the option is wide. If the Panel has specific recommendations, even creative ones, for special controls, we really want to hear that.

DR. SKATES: So, special controls can include requirement for a prospective study comparing the device in dermatologists with and without the device, and requiring significant increase? Because that's what I would like to maintain. I have – the description of the 510K I've heard so far leaves a lot of leeway for that not to happen. I'm quite concerned about that.

DR. HARRIS: Okay. Thank you. Before we proceed with the open portion hearing of today's meeting, I also, again, like to offer an opportunity for Dr. Karen Burke to introduce herself.

DR. BURKE: Can you hear me? Thank you. I am so sorry. I'm Dr. Karen Burke. I'm a dermatologist, a clinical professor at Mount Sinai Hospital in New York, and have done research on skin cancer for 40 years. Thank you.

DR. HARRIS: Thank you. We will proceed with the Open Public Hearing portion of the meeting. Public attendees are given an opportunity to address the Panel to present data information or views relevant to the meeting agenda. Ms. Nalls will read the Open Public Hearing Disclosure Process Statement.

OPEN PUBLIC HEARING

DR. NALLS: Both the Food and Drug Administration, FDA, and the public believe in a transparent process for information gathering and decision-making. To ensure such transparency at the Open Public Hearing at the Advisory Committee meeting, FDA believes it is important to understand the context of an individual's presentation. For this reason, FDA encourages you, the Open Public Hearing speaker, at the beginning of your written or oral statement to advise the committee of any financial relationship that you may have with any company or group that may be affected by the topic of this meeting. For example, this financial information may include a company's or group's payment of your travel, lodging, or other expenses in connection with your attendance at the meeting. Likewise, FDA encourages you at the beginning of your statement to advise the committee if you do not have any financial relationships. If you choose not to address this issue of financial relationships at the beginning of your statement, it will not preclude you from speaking.

DR. HARRIS: Thank you, Ms. Nalls. FDA has received 3 requests. Each speaker will be given 5 minutes to speak. The first speaker is Mr. Simon Grant.

MR. GRANT: All right. I'm Simon Grant, and I'm the CEO of SciBase. I would like to start by thanking the Panel for the valuable input yesterday and this morning. As we saw from the discussion, these are complex issues affecting everything from approved devices to future apps for different types of skin cancers. But we are here to talk about melanoma, which is high-risk, where we believe rigorous FDA control is

needed to ensure safety for patients. We are talking about reclassifying two medical devices, regulated under two distinct product codes, from Class III to Class II. It's not just about MelaFind and Nevisense. As we understand it, FDA is proposing lowering the bar on all adjunct SLAs for melanoma detection. We have many concerns, but we have three major concerns with this, all linked in one way or another to patient safety.

Down-classification to Class II means a standard set of rules will be put in place for all new devices of this new code. This means standardized validation study design and limited involvement from FDA. We think this would be a serious mistake. As we saw in yesterday's Panel discussion, these studies are complex, and the devil is in the details. You can influence the outcome of the studies through its design, which is why the FDA's input is essential. The sensitivity results of a study where mostly later-stage melanomas were included cannot be compared to a study like ours, where 85% of the melanoma were in situ or T1. We worked for months with the FDA to design a validation study, and I think their input was essential. A Class II process, even with special controls, would not include that close FDA input.

Our second concern is that different technologies have different characteristics and different clinical strengths and weaknesses. And so, they do require customized studies and different controls. MelaFind works with different light frequencies; Nevisense works with skin impedance. We don't care what the lesion looks like. There's no significant equivalence between the technologies, and so regulating them under the same rules makes no sense. Technology use can affect performance in many ways. Ethnicity and Fitzpatrick of the patient, lesion size, lesion site – ankle, non-ankle – and even measurement environment. As an example, we note environmental

lighting can affect imaging technologies, but it has absolutely no effect on Nevisense measurements. I can finish with an analogy. In cardiology, different technologies can be used to investigate certain disease, such as ECG and ultrasound, but there's no one who would say ECG and ultrasound devices should be regulated by the same performance requirements.

The third concern relates to the question raised by Dr. Skates yesterday, regarding the carrots and sticks that FDA has available to it after a product is approved. This is an excellent question and raises the question of how much control FDA has in Class III versus Class II products after improvement. I personally worked on 2 Class III products and about 20 Class II products in my 25 years in this industry. I can tell you that, practically, FDA has many fewer sticks, much less control when it comes to Class II devices. I'll give you an example. As a manufacturer of a Class II device, I decide whether a change I make to my product is significant. Perhaps it's a lens coating change or sense of calibration – something small. I write a letter to my file and release it to the market. I don't have to tell FDA, ever. And then a year later, I realize the change had unintended consequences. And with AI systems, small changes can result in large and nonlinear, or even random, effects. But it's too late; the product is on the market and the harm has already occurred. With a Class III product, all changes must be reported to the FDA, and the bar is a significant change is much lower.

These are major concerns, but would also like to comment on the literature search on which FDA has based their decision to reclassify. There are nine publications listed, and they are discussed today. One is outcome is our pivotal trial, the basis for our Class II approval. Three are reader studies from MelaFind. One

addresses apps with no performance evaluation, and the other four are reviews or overview articles, where the most relevant is a broad UK review of skin cancer systems, the Cochrane report. The conclusion of the Cochrane report states incomplete reporting of studies made it difficult for us to judge how reliable they were. Many studies had important limitations. Some studies only include particular types of skin lesions, or excluded lesions that were considered difficult to diagnose. It went on to say these characteristics may result in computer-aided diagnosis systems appearing more or less accurate than they actually are. And that's exactly what we mean. SciBase does not understand how these publications can form the basis for reclassification. We need well-designed prospective trials, not reader studies or reviews.

One final point we would like to repeat from yesterday is: it's still very early for this type of product. Normally, FDA considers reclassification when there are years of expense with multiple products, many products, and they have built up an understanding of the risks and realities of the product class. That's not the situation we have here. Our only real experience is from our product, Nevisense, and until recently, that experience was very limited. It's just too early to consider reclassifying.

So, if I summarize our three concerns, you can't use a cookie-cutter approach for clinical validation studies for melanoma. We believe you need direct FDA involvement to ensure the results can be trusted. You can't regulate very different technologies effectively with the same set of performance characteristics and requirements. That's why FDA put MelaFind and Nevisense under different product codes to start with. FDA control of device changes after approval is essential. Moving to Class II, this control of the number of sticks that FDA has at their disposal is greatly reduced. And in

conclusion, you know, at SciBase, we work with other indications. An example is eczema. There, we believe a Class II designation is correct, and we would even fight for a Class II designation. But this is melanoma. And melanoma is different. If we get it wrong, or another company gets it wrong, people will die. Yes, it was hard to navigate the Class III approval process. Yes, it took time and large studies and inspections and all that stuff, but we did it. If a small Swedish company like SciBase can do it, I don't see why other companies can't as well. Patients deserve no less. Thank you.

DR. HARRIS: Next speaker is Mr. James Castro.

JAMES CASTRO ARGUETA: Good morning. I am James Castro Argueta. I'm a medical student completing my degree at George Washington School of Medicine and Health Sciences. I appreciate the opportunity to speak today on behalf of the National Center for Health Research. Our nonprofit research center conducts research and analyzes scientific data to provide objective health information to patients, health professionals, and policymakers. We do not accept funding from drug, medical device, or tobacco companies, so I have no conflicts of interest. We do not support the proposed reclassification of the two skin lesion analyzers, MelaFind and Nevisense, from Class III to Class II. We have two major concerns.

Number one, there's no guarantee that newly-developed devices cleared through the 510K devices would be accurate as those currently on the market. For that reason, newly-developed devices should be reviewed through the PMA process. Number two, there is a clear risk of false-positive and false-negative results, misuse, and device failure, and the FDA has not provided evidence that these risks can be adequately mitigated with Class II general and special controls listed in the executive summary.

As described, the FDA cannot ensure there will be reasonable assurance of safety or effectiveness over the lifespan of the newly submitted devices. For example, the FDA mentions that clinical testing and non-clinical testing needs to be performed to determine whether the device has an acceptable specificity and sensitivity, but that does not specify what will be considered acceptable, nor does the Agency specify how often this testing should take place. This is a problem because both of these devices are much more dynamic than was described by the FDA summary document. Each of these devices will be subject to future software and hardware updates that may change sensitivity and/or specificity. This week's high-risk recall, Baxter Healthcare's Abacus software, is a useful reminder that software glitches can seriously harm patients. Where is the FDA evidence that special controls will keep up with future software updates…? It seems to be lacking. As the software undergoes updates that may change the interface or use, it is essential to ensure patient safety, and that can be best done through the PMA process that requires post-market studies. This is especially crucial since these updates can affect the devices under poorly-understood AI/ML learning algorithm.

Furthermore, the special controls require that adequate information be provided on device labels regarding its operation, use, users, patients, lesions, and more. However, in practice, we all recognize that it's only helpful if the label is carefully read. In reality, many healthcare providers take their cues from device sales reps, which definitely isn't sufficient as a special control. For that reason, I would suggest that the software itself be required to make clear what the intended use is, who should perform the assessment as a physician operates the system. This would help ensure that each

operator is routinely reminded of the limitations of the device. Additionally, the background material state that an experienced dermatologist, who has taken a training program, should operate these devices. However, there's no clarity on who would be considered an experienced dermatologist, what the requirements for the training program are, who is teaching the program, and what are the requirements for passing the training program. We know that online training for physicians is often started, but not completed; even when completed, it's not necessarily an effective teaching tool.

I also reviewed the studies and I was concerned about a lack of patient diversity. One of the initial studies done on MelaFind had a patient population that was nearly 98% white. Other studies are conducted in countries such as U.K., Sweden, and Germany on populations that may not be representative of the U.S. in terms of race, ethnicity, and other demographic variables that may affect testing. We agree with the new article that concluded that new studies are needed to evaluate these types of diagnostic aids in more representative populations.

Bottom line, Class III devices are held to a higher standard than Class II devices, and melanoma is a potentially fatal disease. Down-classifying these devices to Class II would mean no clinical trials or evidence of accuracy would be required for new devices of this type. Although the FDA believes that the current devices seem to have an acceptable specificity and sensitivity, there's no guarantee that future devices of this type would be as accurate as the current devices. Inaccurate diagnosis of melanoma, either through false positives or false negatives, fits the definition of being high-risk, and therefore, belong in Class III, and not Class II. Thank you all for your time and giving me the opportunity to speak here today.

DR. HARRIS: Thank you. Our next speaker is Dr. Lily Peng.

DR. PENG: Hi. I'm Dr. Lily Peng, and I'm a Director of Product Management for our Health AI team at Google. Thank you for having me today. We share the FDA's enthusiasm in ensuring that consumer preferences are well-represented in the design of healthcare products. Today, we would like to share a few of our learnings about what consumers with a skin issue need and prefer.

Skin diseases are enormous global burden. Two billion people are affected worldwide, but half the world's population faces shortages of dermatologists. At Google, we see a lot of interest from consumers in this area, with 10 billion derm-related searches each year. So, what kind of conditions are consumers searching for? Here are a few common aggregated anonymized searches from google.com. As you can see, it's not just skin cancer. It's also other skin diseases. In fact, the most common queries are things like poison ivy, herpes, acne, eczema. Generally low acuity conditions, but nevertheless bothersome.

But describing what you have is really challenging. Consumers spend hours researching their issues on the Internet and talking to strangers on forums. In a study of over 1,100 cases, we saw that patients arrived at a relevant condition only 13% of the time. Many participants completely unprompted express a desire to search by image. For example, some wish there was an app, like plant identification app, that can help them take pictures and identify what they may have. Current search-by-image products are not optimized for this use case, and there's clearly an opportunity to address an unmet need with broad impact. To better understand what consumers want in a tool that helps them with their search, we conducted a survey where we asked them to

imagine an unbranded skin tool. We surveyed 2,541 individuals who conducted a search for a skin issue in the past 12 months. We wanted to make sure we have a good mix of folks; for example, across income, race, ethnicity, education, age, etc.

One of the big takeaways from the study is that consumers wanted a stepping stone between conducting an Internet search and doctor's visit. For example, they expected the hypothetical skin tool to be more accurate than other tools, but less accurate than healthcare providers. For consumers, Internet searches are often one of the first options they turn to for help. Beginning relevant results is time-consuming and often frustrating. On the other side of the spectrum, we have a doctor's visit and help directly resolve an issue. This option is much more difficult for many to access and afford. In many cases, it's also not necessary and too many unnecessary appointment can make it much harder for those who need care to actually get it in a timely manner. Consumers prefer to have a stepping stone between these two existing ends of the spectrum. One that provides options to guide them. They expect the results from the tool to require further interpretation by healthcare providers and serve as a step toward resolution rather than being able to directly resolve the issue itself. Skin lesion analyzers, or SLAs, can play a role in any step along this progression, and depending on the SLA's role, the risk-benefit trade-off may be quite different.

So, takeaways: One, there are many unmet needs for consumers experiencing skin issues, many of which are lower acuity conditions. There's a big opportunity to increase accessibility and relevance of health journeys for consumers, and in particular, we have heard from consumers that they would like to have a self-help tool for non-serious conditions so they can decide when to seek medical condition. We hope that

these insights into consumer preferences are useful for this Committee. We believe that providing tools that help consumers, prior to a traditional screening or triage by a healthcare professional, can empower consumers to proactively participate in their health journey. Thanks.

DR. HARRIS: Thank you to the previous speakers. Now, I would like to ask if any of the Panel members have questions for one of our speakers. Dr. Alam.

DR. ALAM: For the last speaker, I wasn't really sure what the speaker's position on the issue. Before FDA was and if they would like to clarify that.

DR. PENG: Thanks, Dr. Alam. Actually, just to clarify, the video submitted wasn't aimed at talking about reclassification. I'm actually a little curious why it was in this particular session as well. This is actually more around getting information about what we think consumers actually want and need and sharing data that we have. So, it's very much not related to reclassification whatsoever. So, apologies if that wasn't clear and apologies if there was a mistake and a break in everyone's flow.

DR. ALAM: It was very informative and interesting. Thank you for that research. Thanks.

DR. HARRIS: Ms. Block.

MS. BLOCK: Hi. Renata Block here. Ms. Lily Peng, thank you for the presentation. It was enlightening for the research that you did and what consumers want, which tells me as a medical provider, dermatology assistant, that we need to be better at educating these consumers of the opportunities that they have to see a dermatology provider or their advanced practice practitioner, such as an assistant or dermatologist. As a dermatology assistant, we are available a little bit quicker than a

dermatologist is, but we work very closely with our collaborating physician in regarding to diagnosis and treatment. You also pointed out that the consumers are looking for identification of non-serious conditions, and as you know, these apps, the SLAs, are concentrating on melanoma, which is a very serious diagnosis in the eyes of dermatologists and dermatology PAs. I think the data is great and eye-opening, and I think, as dermatology physician assistant, we just have be better at educating consumers on the opportunities that they do have besides using apps. Thank you.

DR. HARRIS: Thank you. Ms. Peng, if you have any comments, please just introduce your name before making them. Thank you.

DR. PENG: My name is Lily Peng. Thank you, Dr. Block, for the comment. Again, the major continent of the committee is on skin lesion analyzers, understood. Again, apologies. This really talking about the beginning of the user's journey, not toward the end when they're already in the care of a physician. This is really about finding and helping people navigate to the right kind of care, including physicians' assistants and including healthcare providers, and just for people to understand the research that we did. We also looked at access to healthcare providers of different backgrounds, you know, dermatology, as well as others. It wasn't just, you know, dermatology. So, again, thanks so much for reflecting on that.

MS. BLOCK: And Ms. Peng, no apologies needed. The information was enlightening. Thank you so much for your time.

DR. HARRIS: Thank you. Next, Dr. Bush.

DR. BUSH: Laura Bush. Ms. Peng, thank you so much for the valuable information. The last thing you said was the non-serious conditions. I would wonder if

Google has done any studies on serious conditions, such as melanoma, more specifically. I know they did not give you that slide deck and I understand that. That would be useful information if that was available. Thank you.

DR. PENG: Thanks, Dr. Bush.

DR. HARRIS: Please announce your name again.

DR. PENG: I'm so sorry. My name is Lily Peng. To answer your question, Dr. Bush, we actually looked at the query. As you can see, for different trends that we have, and melanoma was one of the queries. I think the information we were sharing was to say, “Hey, as a Panel considering different kinds of skin analyzers, don't forget about the stuff that we can be doing to help consumers at the end are beginning of their journey.” Many, many of which – the vast majority of which – have acne, and a small number will have melanoma. We need to address those queries and force them into the right pathway. So, how do we think about the risk and benefit of consumers in the beginning of this journey, accessing some additional better help than they are actually currently getting. The current situation with searches is that you go around, you search for hours, you get frustrated, you search some more, and then you finally present in the wrong place, or were either too early or too late. So, how do we reduce that kind of patient frustration? Again, the people who need help with serious conditions, care faster, but the people who don't need help, they can actually not clog the system with unnecessary appointments.

DR. BUSH: Laura Bush. Totally agree. I think in the future, this can be definitely a pathway to angling these patients in the right direction, to the right sources. Maybe

the right nonprofit groups for the condition they are searching, or pathways to dermatologist's office. But thank you so much.

DR. HARRIS: Do any of the other Panel members have any questions for the Open Public Hearing speakers? If not, I now pronounce the Open Public Hearing to be officially closed. We will now take a one hour lunch break. Panel members, please do not discuss the meeting topic during lunch amongst yourselves or anyone attending virtually, and we will resume at 12:30 p.m. Eastern Standard Time. Thank you.

[Lunch break]

DR. HARRIS: It is now 12:30 p.m. Eastern standard Time, and I would like to resume this Panel meeting. Before we began discussing the questions from FDA, I would like to provide Dr. Ryan Ortega an opportunity to provide some additional information regarding the difference between Class II and Class III devices. Dr. Ortega.

DR. ORTEGA: Thank you. Hi. This is Ryan Ortega from FDA. Just a brief overview of some of the differences between Class II and Class III. So, excuse me. For Class III devices, for which a PMA is required, the review of the devices is essentially, you can think of it as [inaudible] — and this is contrasting to a 510K review for a Class II device that has special controls. Generally for a PMA, we say that this premarket review for a PMA is what's required in order to have a reasonable assurance of safety and effectiveness with this device. I believe as Dr. Ashar said earlier, each of these PMA reviews is assessed almost in isolation, on its own, own its own merit, but for a Class II device that has certain controls associated with it, what this says is that every device of that device type will have to meet this set of special controls. Meeting those special controls is what gives us that reasonable assurance of safety and

effectiveness of that device, along with the general controls that all devices are subject to. I think it is important, particularly for this discussion, to note that the special controls for a Class II device, every device has to meet those controls. But the way each device meets those controls may have some variation. An example might be, say, if clinical data is a special control, every device would want to meet that special control, but perhaps the design of the trial might be different, depending on the very specific indications for use, or the specific labeling claims that the sponsor of that device wants to make.

Additionally, you know, we do try to take a least-burdensome approach to regulation and try to strike the right balance between what we are asking as far as evidence is concerned for assessing each device. There is a difference between kind of what a PMA looks like and what a 510K review looks like. Generally review timelines for Class II devices are a little faster. I believe the user fees associated with Class II devices are less, as well.

I will say it's also probably important to consider how we look at changes for Class III and Class II devices. So, for a Class III device that has a PMA, if the sponsor of the device would like to make a change that would affect the safety and effectiveness of the device, that generally would require a PMA supplement. However, if there is a change that doesn't affect the safety or effectiveness of the device, often times that is just reportable in annual report, rather than through submission of a PMA supplement. For a 510K-cleared device, on the other hand, generally if a sponsor wants to make a change to a 510K device that could significantly affect the safety and effectiveness of that device, then that is generally when a new 510K would be needed for that modified

device. If the change wouldn't significantly affect the safety and effectiveness of the device, in that case, that would be documented by the device sponsor and could potentially be part of an inspection, say, whenever an FDA inspector comes to inspect a facility. Dr. Asher, is there anything else that you think I should mention as far as the difference between Class II and Class III?

DR. ASHAR: Now. I think, Dr. Ortega, that was very helpful. Actually, I might add my comments, Dr. Hobart, if that's acceptable to you.

DR. HARRIS: By all means.

DR. ASHAR: So, you know, kind of taking what Dr. Ortega said step further, in the space of machine-learning and artificial intelligence, because I understand the Panel has concerns about that. It's important to understand that our goal is to assure that medical devices, there is reasonable assurance of their safety and effectiveness. The presence or absence of artificial intelligence or machine learning does not drive the regulatory classification. We have devices in every class with this technology. Rather, what we are trying to do is understand what the Panel wishes to know about the AI and machine-learning capabilities to help ensure a reasonable assurance of safety and effectiveness. So, for example, I'm going to list some things, but these are food for thought. If you had a special interest in making sure that the training data set had certain aspects associated with it, whether it be diversity, or the type of melanomas, or anything else, that would be helpful. If you felt that the test data set had to demonstrate a certain amount of improvement over clinical practice, knowing that would be helpful. If you wanted updates at regular intervals, I think that was mentioned. If you wanted labeling that communicated whether the patient was within the scope or outside the

scope, whether it be with respect to skin type or other characteristics, those things will be helpful.

At the back end, you know, our goal is always to assure reasonable assurance of safety and effectiveness. If you are putting something forward that cannot be accomplished through special controls, we are taking this as advice and I think we are going to have to determine, from a regulatory perspective, whether or not we can accomplish what it is that you desire to have understood, tested, evaluated, publicly available regarding these technologies. And our responsibility is to take your advice and put it into the appropriate regulatory paradigm. If it doesn't work for a special control in Class II, it doesn't work. I think, hopefully, that causes you to think critically about the data and the scientific information that you wish to obtain in the premarket, the post-market, both, as the classification of this device is contemplated. Thank you.

DR. HARRIS: I would like to ask one quick question, and perhaps direct this to Mr. Bryant. As the industry representative, can you give us any sense of how companies look at Class II versus Class III device requirements when thinking about developing a product, or thinking about how competitors may be developing products? In other words, would change in the classification influence companies thinking around their future competition and their willingness or interest to produce a device?

MR. BRYANT: Thanks. LaMont Bryant. Specifically around how companies look at the classifications, I will give you a general perspective. Typically, when companies or firms are looking at developing products, they first and foremost look at the product, procedure, and the level of risk, and then the investment in time. Meaning, really focusing on promoting and protecting public health, what are going to be the

requirements generating the data? What is your timeframe? And then what is the probability of success? The actual classification doesn't really play a critical —is not the deciding factor, but it is to be considered as we are assessing.

The second question you asked, specifically around competitive landscape... After you have the conversation around your claims, the product and procedure, the competitive landscape, then you think about what do you truly want to claim as it relates to differentiating your product? As you know, the 510K process allows for companies to be able to claim substantial equivalence if there is a patient advantage and strategic advantage to being able to generate the data to give you a stronger claim… that companies sometimes air toward generating the additional data. Dr. Harris, did that give you a general landscape assessment?

DR. HARRIS: It does. So, I see that Dr. Rotemberg has a question relative to this discussion of classification distinctions. Please.

DR. ROTEMBERG: Thanks, Dr. Harris. This is just a general comment about the statement that the FDA has made about how there is enough information in the literature to consider this reclassification. And I just wonder if it might be helpful to this Panel. I'm an AI researcher. My expertise is in factors that influence AI application to melanoma detection specifically. And I just wonder if it might be helpful to this Panel for me to try to give, like, a 5 minute overview of the latest research in this area that I believe the FDA is resting some of their claims on. But it could be at a different time if you think there's a different opportunity in the schedule.

DR. HARRIS: No. I think that would be both beneficial and appropriate at this time. By all means.

DR. ROTEMBERG: Okay. I am going to do my best. Essentially, what I want to talk about today is the factors that we already know influence artificial intelligence performance and practice, specifically as it relates to melanoma detection and how it relates to the proposed special controls. So, first of all, I think – and we talked about this a lot – I know Dr. Skates and Dr. Alam, in addition to lots of people on this Panel, have talked about the dermatologist interaction with the technology. I would say we know that sensitivity and specificity on static data, even if it's collected prospectively, is not the same as the way that it is in practice when interacted with dermatologists or dermatology providers. We also know that the specific output, whether it's melanoma probability, or multi-class probability, or just a binary class, actually impacts the way that the dermatologist interacts with it. So, these seemingly minor details suggest to me that static assessment, you know, purely sensitivity assessment, is not going to be sufficient for us to understand how well these devices perform in practice. That's the first one. I made a list.

The second thing is, and I think this is really important for us and we discussed it yesterday, is around concerns about ethics and transparency. We know that it's really important to understand the data that algorithms have been trained on. There's a lot of recent literature that suggests that algorithm performance in certain populations, skin tones, but not just limited to that... In addition, unusual anatomic sites, unusual melanoma subtypes… All these are going to influence performance. And I would suggest what we really need is transparency from the developers on the specific data that's been used for training and validation, because it's not going to be possible for us to tell the FDA, you know, you need 5% melanoma and you need 20% this, we actually

need to know the distribution and understand, if there's going to be disparate performance, and that's well-known and well-characterized in the literature — and I also think that coming in oh, from the perspective of algorithm design, we need to know the demographic distribution of the patient population.

The other things that I think are really important for us to consider is that algorithm performance degrades naturally over time, and this has also been well-characterized in lots of domains. We know that demographics will change, the people who come to the dermatologist will change, and also, the use of the device will change, even if it's not intentional. You know, I thought that you gave a really good example, I think it was Dr. Skelsey, about how the use of MelaFind changed over time just by having it in the office; and in that case, it decreased. But in other cases, it might increase as you develop confidence in the technology and start to use it in different domains. We need really careful post-marketing analysis of accuracy in its actual place of use. And I don't even think that's even technologically that challenging with everything these days, with everything so connected, so I don't want us to worry too much about that. Oh, the other — I apologize for this. I will say, we published recently some guidelines for the evaluation of AI technologies, and I am resting heavily on the eDelphi process that we performed. This is not just based on my own assessment. This is based on expert consensus review.

Finally, I think from the technical standpoint, again, we are very worried about both false-positive -- I mean false-negatives and false-positives, but we do need to take into account the difference between severely dysplastic nevus and melanoma, and I think we should take about that in the context of stratifying our concern and gold

standard rating, and I think that's a little bit different from what has been published in some of the previous studies. And we should develop automated safety standards to prevent use in the wrong places, where it hasn't been trained. Those are my major summaries, but I would be also happy to address any questions that Panelists have about the current research into AI performance and factors that influence AI performance, including prospective benchmarking studies for dermatology applications, if Dr. Harris thinks that's appropriate.

DR. HARRIS: Certainly. I see we have one question. Dr. Farber.

DR. FARBER: Yes, Neil Farber. Thank you for that, Dr. Rotemberg. That was very insightful and helpful. The thing you first described in terms of the fact that the data changed depending on the interaction with the dermatologist is not unexpected, given long ago, the Hawthorne effect was discovered in, I guess early 1900s. Where basically anytime somebody tries to measure something and its human behavior involved, it will change, because the fact that they are being observed. In this case, because of the fact that the interaction is the very nature of the beast, it's going to change what data occur. And that means that basically, even with large sets of data, it may not apply to other AI devices.

DR. HARRIS: Okay. Dr. Alam

DR. ALAM: Thank you. I have a couple of questions for Dr. Rotemberg as well. From the standpoint of a company that is potentially, let's say hypothetically there's a process with special controls to get something to market, and there is a predicate device, you have, I think, explained some potential problems that can occur with that. But are there other areas you could detect that — I don't want to impute a bad motive to

a company, but the purpose of a company is to make money and provide a product that's good enough to get through approval, that they can market subsequently. Are there some other areas where you are concerned that, with the current proposed regulatory framework, that it would be easy to produce something – well, not easy, but possible – to produce something that looks good, met the rules, but really didn't have the level of quality that we would want in it?

And then one other question that's similar to that, which is that, do you think the current proposed special controls are enough to ensure that, given this is not a mature technology, AI isn't, and presumably the advances in the future… is it enough to ensure that the quality of these devices will continue to improve, and that there will be motivation to do that? Or will this sort of, to some extent, cast things in stone and keep us where we are, rather than make things better. Thank you.

DR. ROTEMBERG: I'm Veronica Rotemberg. I think the details are outside of my domain, but I would say, as I've advocated before, I think that one of the things I would require is evaluation of accuracy in a prospective intended use trial, including for new devices that would be compared to those that are already approved. And I do not think any standard of just specificity or sensitivity against a retrospective data set is sufficient to understand how these devices would be used in practice, and it's also not sufficient to understand how dermatologist practices, in practice. We do need to get away from reader studies as benchmarks for these kinds of devices, because they just are not the same as what happens in the real world. I think that's one of the things that I would say about this premarket evaluation of these devices.

The other thing that I would say is that the distribution of the validation data is very important to algorithm performance. And so, I think it would be very difficult to express equivalent performance on different test data sets, because these algorithms can sincerely not be compared to each other. And so, and historically, you know, these validation data sets have not been, as we discussed, as diverse or representative as we would like. So, I would expect that to improve over time, and that might mean that our standards improve over time, and that would be something that I recommend that the FDA really take into consideration when they are thinking about designing something that would fit under special controls.

In terms of the current special controls question, I did my best to outline the places where I think that the current proposal from the FDA fall short. I think in broad strokes, that includes clinical validation. I think that's how they called it. It also includes post-market surveillance, which I outlined why it's okay to be so careful, and I think a registry, something that looks like there is, would be extremely appropriate for this type of technology because we have no idea how the drifts will affect performance. And I would also recommend, as we discussed, you know, updating algorithms based on performance benchmarks because that is a clear, clear design that could be easily implemented. But again, those performance benchmarks are going to be very challenging because they're going to be different for every device. Does that answer your question?

DR. ALAM: It does. Thank you very much.

DR. HARRIS: At this time, I'd like to focus our discussion on the FDA questions. Panel members, copies of these questions are in your Panel packs. I would ask that

each of you continue to identify yourselves each time you speak to facilitate transcription. May we please see the first question?

FDA QUESTIONS TO THE PANEL

DR. WANG: Welcome back. I hope you had a good lunch. My name is Jianting Wang. I am a Biomedical Engineer, Acting Assistant Director for Light-Based Energy Devices Team in the Office of Surgical and Infection Control Devices. In this session, we would like to invite the General Plastic Surgery Devices Panel to provide feedback to our questions regarding the reclassification of computer-aided devices, which provide adjunctive diagnostic information about lesions suspicious for melanoma. We have three questions for you. The first one is regarding health risks of these device types. The second one is regarding whether the Class II device criteria apply to these device types. The third question is regarding the proposed special controls. In the following slides, I will ask the questions one by one.

Question 1. FDA has identified the following risks to health for computer-aided devices, which provide adjunctive diagnostic information to dermatologists about lesions suspicious for melanoma based on available information for these devices, including data in P090012, available to FDA under section 520-H4 of the Food Drug Cosmetic Act, input from the 2010 Panel on P090012, published peer-reviewed literature, and post-market experience associated with use of these devices:

False-negative or false-positive results. False-negative results could result in complications, such as incorrect or delayed diagnosis and delays and biopsy decisions

and melanoma treatments, which may allow an undetected condition to worsen, and potentially increase morbidity and mortality. False-positive results may result in complications, such as incorrect management of the patients, including unnecessary additional invasive biopsy procedures and more frequent screenings, as well as the potential administration of inappropriate treatments and/or the withholding of appropriate treatments with adverse effects.

User error/improper device use. The device could be misused to analyze images from an unintended patient population and on anatomic sites or lesions having an unintended attributes, or to analyze images acquired with incompatible imaging hardware or incompatible image acquisition parameters, resulting in a device not operating at its expected performance level. The device could also be misused if the user does not follow the appropriate reading protocol for using the device to assess lesions of interest, which may lead to lower accuracy. Inaccurate results may result in the same complications associated with false-negative or false-positive results, as discussed above.

Device failure/malfunction. Device failure or malfunction result in the absence or delay of device outputs or incorrect device outputs, which could lead to inaccurate patient assessments. Inaccurate results may result in the same complications associated with false-negative or false-positive results, as discussed above.

Electrical, thermal, mechanical, or light related injury. While in operation, the device may discharge electricity that could shock the user or patients. Electrical discharge or exposure to device-generated heat may cause thermal injury or discomfort. Moving parts may cause mechanical injury. For devices that utilize light to provide

adjunctive diagnostic information, accidental eye exposure to the light source could cause eye injury.

Interference with other devices. Individuals with electrically-powered implants could experience adverse interaction with the device due to electromagnetic interference or radiofrequency interference.

Adverse interaction. A patient could experience skin irritation or allergic reaction associated with using the use and operation of the device via the use of non-biocompatible materials and patient-contacting devices, infection, or cross-contamination. If certain components of the device are not adequately sterilized or if reusable components are not adequately reprocessed between uses, the device may introduce pathogenic organisms to patients and cause an infection.

Please comment on whether this list completely and accurately identifies the risks to health presented by computer-aided devices which provide adjunctive diagnostic information to dermatologists about lesions suspicious for melanoma

Please comment on whether you disagree with inclusion of any of these risks or whether you believe that any other risks should be included in overall risk assessment of this device type.

DR. HARRIS: Thank you. I would like to now hear the Panel addressing these specific questions. First, Dr. Burke.

DR. BURKE: I would just like to know if we could learn… I think this is a very excellent and complete list. And it was mentioned about implants, particularly pacemakers. I think it's important to know, can this be used in a patient with a pacemaker, and if so, can it be used near the site of the pacemaker? And similarly for

other devices, even metal implants when they hit replacements that are metallic or other synthetic materials.

DR. HARRIS: If I understand your comment, you would feel that should be one of the requirements that a manufacturer would need to provide, is the safety profile in patients with pacemaker or other implants, and correlating that to the actual distance from that implant that the device is being used?

DR. BURKE: Yes.

DR. HARRIS: Okay. Dr. Alam.

DR. ALAM: Yes, I think I can start. I think, with probably the help of people like Dr. Rotemberg and others who understand what's inside the black box, but some of the things that seem apparent are that somehow you have to make sure that all of these devices – I don't know what the correct term is, norm-standardized on an identical or very similar set of patients. You don't want to teach to the test, so you have to find some way of ensuring that.

You want to make sure that the quality of the output and, in particular, false-negatives are very low for different types of melanoma, for melanoma at different anatomic sites, for melanoma in patients with different ethnicities, for melanoma at different stages of evolution including before they happen when they're severely atypical nevi, and that this would be robust over time. I have heard several mentions of post-marketing surveillance and, perhaps I'm in a minority, but I'm very skeptical about the value of that. I think that's kind of something we suggest when we don't know what else to do: we're just like, "Let's have post-marketing surveillance."

But the device is already approved and out there, and the fact that you have a registry that's collecting data over however many years really doesn't impact how that device is going to be used. So, I don't think that's adequate. But other methods might be implemented to make sure that these are similar. I would generally agree with doctor – well, several physicians who have spoken, and others – that because this is such a complex issue, and the potential of harm is quite substantial if we get it wrong, something like a prospective trial would be helpful. The benefits of prospective, RCTs of course, are they allow us to control for unknown confounders and I suspect we have unknown confounders here. Thank you.

DR. HARRIS: Thank you very much. Dr. Farber.

DR. FARBER: Neil Farber. Thank you for that. I agree there needs to be more testing. And especially around the issue of false-negatives and false-positives, because it's such a risk and because this is such a complex issue. Because the fact that you have an AI device in one hand and interaction with dermatologists on the other. It makes a very complex... and I think we need some real-world, real-use types of studies and those could have to be prospective. I would defer to Dr. Rotemberg.

The other thing that's necessary, regarding the false-negatives and positives that have not -- has not really been addressed until I addressed it yesterday, is the psychological affects that either false-negative or false-positive could make happen. It's obviously more prominent when it's a layperson using the device. And that's why I'm not going there because that's a whole other world. But even in the issue of these devices it needs to be explored and it has not been explored.

DR. HARRIS: Thank you very much. Dr. Rotemberg.

DR. ROTEMBERG: Veronica Rotemberg. I agree with a lot of what has been said. I think what is missing on this list is the risk of bias. I think that it is an overarching thing that includes false-positives and false-negatives and algorithm performance, but also, includes perpetuating systemic problems in the healthcare system, based on bias, algorithm, performance and training and I think it might make sense to add one item related to algorithm bias. And I think that that also encompasses a lot of the concerns related to validation. How do we validate these algorithms against benchmarks that include lesions that are susceptible to bias? Some of the things we know are related to ethnicity and skin tone, some under-represented types of melanoma. We need to measure the performance based on innovative in site II and severe dysplasia separately, since the risks of false-negatives of those are clearly different.

And I know that the Panel may not exactly agree on those specific numbers, but for sure, these types of analysis need to be performed. And I think it also speaks to potential sources of bias of the algorithms.

DR. HARRIS: All right. Thank you. Dr. Skelsey.

DR. SKELSEY: Thank you, Maral Skelsey, picking up where others have already talked about. Dr. Alam mentioned the false-negative and false-positives, are really the heart of this issue. And the lack of data on skin of color, something where it's labeled not to be used for patients with skin of color. Both because of the lack of validation data and because of the inability, at least for one of the devices to be used in areas where patients with skin of color are prone to skin cancers. So I think it's -- it needs to be addressed more specifically. And in addition to increasing anxiety, I think patients' needs to know the often increased cost. Because with devices where we know

that the specificity is low, patients will increase their numbers of biopsies and they need to know that that will increase their cost as Dr. Rotemberg said there's the risk of bias and at least from the prior data, we know that there will be an increase in procedures that may or may not be necessary, but most likely, unnecessary.

And lastly, with use, I think the FDA in its original labeling was correct, in requiring dermatologists use these devices. Specifically not -- I mean, just for lesion identification and identifying higher risk lesions, because of the use of dermoscopy. But also, because the devices require that they not be used on skin that has inflammatory conditions. There's a list of all the conditions that are confounders and I think that placing -- diagnosing those conditions, places of burden on the non-dermatologist increasing the possibly of user. Thank you.

DR. HARRIS: Thank you. Dr. Skates.

DR. SKATES: Dr. Harris, thank you, this is Steven Skates. I also have a brief presentation akin to Dr. Rotemberg that bear on the false-negative and the false-positive issue that Dr. Alum I think, highlighted as the major issue in this list but it is a benefit/risk issue that I'm proposing here to deal with the specificity, especially. I don't know if this is the appropriate time to raise both the risk and the benefit but at some point we need to and this bears on this question of false-negatives and false-positives.

DR. HARRIS: If it does apply to answering this question, by all means. This is an appropriate time.

DR. SKATES: Okay, I'm going to share my screens. I tried to put this together. This is the MelaFind Panel conclusion and this risk of false-negative was a significant safety concern. And safety in the FDA's eye seems to be focused on is there an acute

adverse event. And I think that is missing the big safety concern here, which is that you don't diagnose a melanoma when there's a melanoma there and that's a false-negative that this Panel is concerned with, back in 2010.

In which I'm concerned with and I believe many other people on the Panel are concerned with. What I would like to do is just show you one way of assessing that, because the specificity issue is very hard to gauge. And I'll make a case for this. This is the MelaFind PMA results, there were 173 true positives for the dermatologists, two false-negatives, 82 true negatives and 1375 false-positives from the dermatologists.

The MelaFind changed that. 157 true negatives and less false-positives so it had improved specificity. So they both had very high sensitivities, 98%; that sounds great. But then you look at the specificity, it's 5% doubling up to 11% with MelaFind and as Dr. Skelsey said, the use of MelaFind was so low specificity it you want wasn't helpful. Now I would like to points out that Nevisense specificity was 31%. So 31%, is it enough? Or does it need to be 50% or what? And here is how I would like to suggest reducing that.

Specificity, for detecting disease, but specificity is too high. It impacts the false-positive and you want a balance between the false-positive and the true positives and the false-negatives. It's the number of false-positive balanced -- that's the false-positive rate and this incident can vary between study populations and because their incidents vary so much we don't have a good intuitive instrument for what specificity should be. I suggest we quantify this benefit to risk ratio, this risk meaning the facilities negative, big concern I and Dr. Alum and others have.

And use this benefit/risk ratio of the dermatologists as a metric and goal for other benefit/risk for other settings of SLA. I can see this SLA being eventually used by --

sorry, PCPs and um potentially even the lay public. This is too detailed to go through, but there were 173 positive that the dermatologist found and 2 melanomas they missed and almost 1,400 false-positive. I'm going to suggest a weighted risk between missing melanomas and essentially biopsying lesions that were not melanoma. We can change this to whatever, dermatologists would judge to be right. But that, essentially, says that there were four events that were at risk or 4 equivalent of 4 melanomas missed and that's weighted benefit/risk ratio is 43 melanoma, which is one melanoma missed or 7 hundred biopsy of not melanoma or a combination of those two.

That is what dermatologists can achieve and that was in the MelaFind study. How does dermatologists achieve in the population that is being studied. How many benefits do they get per bad outcome? And if you look at the benefit/risk ratio of MelaFind, there were fewer -- sorry, there were more true negatives, non-melanomas classified as not melanomas. And fewer false-positives but there was one additional melanoma that was missed. And instead of 2 plus 2, you have 3 plus 2. That gives you 35 melanomas detected for bad outcome, instead of 43. And so the comparison here is 43 melanomas detected per bad outcome, which is a dermatologist, which I'm going to assume is an acceptable, benefit-risk ratio because that's what dermatologists practice and achieve, at least in the MelaFind study. The MelaFind device only found 35 melanomas detected in the bad outcome. And to get that specificity -- to get that benefit/risk ratio to 43, you would have to increase the specificity to 54% and not just double it from 5.5 to 11%. The 11% falls way short, and making these equal gives us answer objective way of estimating what the specificity needed.

So my suggestion to FDA is this is a better approach to specificity goal. Choose a ratio that is best for safety. And is a reasonable acceptable state -- standard for patient safety. So then the specificity goal for should be set so that the benefit/risk ratio is the same as the dermatologist and then that will remember sure a patient's safety and address the 2010 Panel's concern about patient risk. So for melanoma detection, you might want to use this 43 benefits per bad outcome, and apply that to SLAs aimed at detected melanoma for other populations, other providers. And with those other study populations, where the instances vary all over the map, you would then at least achieve this benefit/risk ratio that dermatologists achieve by setting the appropriate specificity level.

To enable you to do that you need a study design which is prospective. Compare the dermatologists standard of care, without the SLA and with the SLA and if you want to expand it to other care providers or potentially to the lay public, then I would suggest that safety would want you to get the benefit/risk ratio as the same as the dermatologist, and thereby deal with this hugely varying incidents issue and complicated balance between one minus specificity incidents. This gives you an objective way to set patient safety. That gives you what is needed to set an accurate goal for both the false-negative and the false-positive and the true positive which is the sensitivity. Thank you. So that's my advice to the FDA, on -- and whether that is -- that requirement is part of the 510K, it's very unclear to me. I think it is part of the PMA and I base this these judgments on having done prospective studies with the AI, over the past 30 years, not in dermatology, and I need more collaboration with a dermatologist, but in varying cancer, early detection. So it's based on my experience in that and

setting patient safety by this benefit/risk ratio, minimum standard, that is currently practiced. Thank you.

DR. HARRIS: Thank you. A couple of clarifying questions, if I may. So, the MelaFind study, if I understand your analysis, would indicate that when the -- when the dermatologists use the device, they were actually -- performed less well.

DR. SKATES: That's correct. Well, it's unclear how -- what the action was taken, whether they -- where the action was taken, but the classification was less accurate.

DR. HARRIS: Less accurate.

DR. SKATES: Was less accurate, yes. Whether they followed their judgment or the judgment with the SLA, it's unclear from that study.

DR. HARRIS: And my only other quick question was: If you're suggesting that the use of the device should essentially match the performance of a dermatologist not using the device, what would be the rationale for such a device? Since we spent a lot of time yesterday talking about the need for that to enhance performance in terms of diagnoses and treatment?

DR. SKATES: So I'm setting a minimum safety level, here. The effectiveness could be greater, so um the effectiveness of...

There, MelaFind was an increase of specificity of going from 5.5 to 11%, doubling it. And maintaining the same sensitivity. So, the reason for at least obtaining, I guess my goal is to obtain at least what is acceptable for benefit/risk ratio, where acceptability is defined what dermatologists can currently achieve now. it's a good question, if dermatologists are going to use SLAs then there should be some increase in

one or two -- one or either of those metrics of accuracy. And then the question is: How much a decrease should you accept in the other metric?

Because in the MelaFind study there was increase in sensitivity. So it's the safety that I think should meet the same level as the dermatologist without SLA. And then another metric of effectiveness could increase it.

DR. HARRIS: Thank you. Dr. Bourelly.

DR. BOURELLY: Yes, Paula Bourelly. I ask that the FDA consider in addition to special controls not just a focus on device performance, but on provider performance, I know that was mentioned before. We talked about with updates of algorithm you need to get your device updated and functioning at an acceptable stand Saturday. I would love to see the same thing happen with the provider. I would love to know if there's going to be mandatory training with some sort of establishment of competency. That was brought up this morning. And who gets to do the training? Does that training come from the company but then can a provider in one practice train another one in that is considered allowed? Is there a certificate that goes with that and a certificate that might need to be maintained? Could I train a PA in my office, for example? Trying to focus not just on the performance of the device but the provider.

DR. HARRIS: Not to put you on the spot, Dr. Bourelly, but what would be your suggestions in answering all those questions you posed?

DR. BOURELLY: Yes, I would love to see mandatory training that could be offered by the company. I don't know who should do the training, the company or the developer, but I would love to see that happen.

I would love to see maintenance and certification, so when algorithms are trained -- sorry, updated or changed, not only does the device meet the standard but the provider is also required to demonstrate continued learning and competency. And I would like to have one person do the training. In other words I would prefer that someone who has been trained not be the next trainer. Because I think there could be a lot of things lost there. I would love the company to train everyone in the practice and allow that to be one standardized certification. My thoughts. Thank you.

DR. HARRIS: If I understand you correctly, you would be in favor of non-dermatologists being trained to utilize to use this device?

DR. BOURELLY: Paula Bourelly. I was not going there. I was thinking of multiple dermatologists and practice? Does everyone get trained by the company? Or -- one doctor can train his or her partners or in the case of a dermatologist who has a physician's assistant or nurse practitioner, does that doctor in the practice get to train the physician extenders or should the training come from the company.

DR. HARRIS: Right, but you are saying physician extenders would be --

DR. BOURELLY: Derm physician extenders. I wasn't thinking of primary care.

DR. HARRIS: Thank you. Dr. Roth.

DR. ROTH: Could you please post the question again. This is Katalin Roth. I feel we have gone far away from the question. So my question here -- I agree with some of the things which have been said about how to enhance performance, and I thought we were only discussing using this for physicians; whether primary care, plus dermatologists, or only dermatologists, I think that needs to be decided. But – and I support the idea of training for the users – but I'm not clear whether any of the issues

that we have discussed, impacted on changing the device classification, from Class III to a Class II. Am I missing something here?

DR. HARRIS: Well, I think the first question was actually to determine whether this list of health risks that you see before you, is adequate or not. And Dr. Alam made the point that he felt the key issue needed to be teasing apart the false-negative false-positive issue. And that brought about the remaining discussion. So, I would propose this is directly answering the question.

DR. ROTH: I apologize. This wasn't meant to be a criticism but we are going over some things that we went over yesterday.

And when I had not heard -- we have --

I mean, I think Dr. Skelsey talked very interestingly to me about her personal experience using her device. And you know, as a very capable dermatologist she ended up finding it got in the way. But I haven't heard any information, I don't know if any information has come to the FDA, about any of the kinds of concerns we are talking about, having resulted in patient harm.

No one has mentioned, for example, there has been specific instances of either patient harm, by inappropriate biopsy. Or patient harm in terms of missing melanoma. My brother is a dermatologist and I almost became a dermatologist and I have the greatest respect and interest in dermatology. And taking care of people with a possible melanoma is one of the most challenging things that face dermatologists. So, I would be interested to know whether there have been reports of adverse effects on patient safety, from the device, as we discuss this. And I think I'm going to — I had some more things to say, but I love Dr. Bourelly's suggestion about training. And I do think that if

physicians are asked, for example, to I think this is something that physicians can do. I'm hearing a lot…

I'm very interested, personally, in empowering patients. But I also feel like we are kind of disrespecting the end user of this, when we are assuming that the end user is a dermatologist. And I don't think that would be the case. So, thank you.

DR. HARRIS: Thank you. maybe to answer one part of your -- or to address one part of your comments, FDA did present earlier, issues of reporting through the MO(D) system and others and as I recall, they actually did not have any adverse events to report, but then the reporting systems are not perfect, or necessarily robust, because it's voluntary reporting.

DR. ROTH: But understand that but I think we have ten years of experience here that should temper our um concerns.

DR. HARRIS: Okay. Very good. Dr. Rotemberg.

DR. ROTEMBERG: You know, Dr. Roth, I thought your points were well taken. The thing -- I thought Dr. Skates' analysis was excellent. But a few things that I think are missing from that analysis, are specifically what happens with the false-negatives.

And we don't know. We don't -- we -- first of all don't know if those are severely dysplastic nevi which have 0.2% chance of being melanoma versus an MIS versus invasive melanoma. We don't know if a dermatologist would have biopsied that anyway because it was clearly an invasive melanoma or clearly concerning for another reason.

We don't know if those were lesions of patient concern that would have gotten biopsied anyway. And I think all of that is missing from just saying false-negative is scary, that's why I think Dr. Skates and others have advocated for prospective trials

because that does answer those questions. I'm not taking away from your analysis, I'm saying you know, we don't know the answer and we are not giving dermatologists enough credit and we are not giving patients enough credit. We just do not know and I think that's one of the major things that we need to temp rise our concern about false-negatives by really asking for the data that we need to wage those concerns. If everything that is a negative, gets monitored for 3 months or 6 months, that's a different situation from they get sent away and never see a dermatologist again.

If most of those got biopsied anyway in real life because there was a different reason for concern, again, that's a totally different story. And the degree of invasion is obviously a different story.

The other thing I would say, with caution about Dr. Skates' analysis, those are based on reader studies --

DR. SKATES: The MelaFind was a pivotal study.

DR. ROTEMBERG: Users were not given the MelaFind results.

DR. SKATES: Yeah, I don't know the details but these can be built into a prospective study. My concern is that down-classifying to Class II would minimize the need or [off mic] need for --

DR. HARRIS: Yes, that's Dr. Roth reminded us we are going to talk about that question. Right now let's focus on the discussion on there's a list of potential health risks and do we as a Panel feel the list is complete and if not, how would we like to expand it? Or anything that could be removed?

And then we will get to the other issue of the down classification.

Dr. Ballman.

DR. BALLMAN: Yeah, I think my issue is more the special controls, and I do really feel that we've gotten into that special control sort of thing. This is just a lift. If you look at special controls, it says sort of clinical performance testing to look at sensitivity and specificity will be used. So, I will hold my comments until the special controls, but I hope others do otherwise, I feel my comments won't make sense when we get to it.

DR. HARRIS: Thank you. Dr. Alam.

DR. ALAM: I wanted to address Dr. Roth's comment about harms and how we have not heard about harms. If we look to stay on topic, this list of risks, I wouldn't be adverse to adding specific harms such as missed melanomas, death. But I do think, to your point, Dr. Harris, that it's very unlikely that that would be a fruitful category, quite frankly because so far one device isn't marketed and the other device is used by a handful of people.

And we don't even know what bases those people are making clinical decisions. I doubt they are saying because I have this device, I'm going to close my eyes and do whatever the device says.

And the voluntary reporting. [Off mic] of developed voluntary reporting database for adverse events in cutaneous procedures. Over 3 years we've gotten 15. I can't imagine 15 adverse events in cutaneous that occurred along 6 thousand members of society. It's just not very robust. Thank you.

DR. HARRIS: Thank you. In an effort to try to summarize, Dr. Asher I believe the Panel feels the list of health risk touches upon a number of important areas but there seem to be a specific focus or principle concern regarding the nature of false-

negatives and false-positives as a health risk of such devices and you've heard the discussion of many, many ways, several items or approaches to tease apart and add, greater granularity to the analysis of the risk of false-negatives and analysis positive. The Panel included need for robust data sets, prospective randomized control trials that would reflect real-world experience with the device, the need to incorporate the psychological effects of information that might be delivered by these devices, addressing the risk of bias that may be incorporated within an algorithm, the need for algorithms to be updated and the need for practitioners to be properly trained and have continued training as the device, the technology evolves.

Any other issues that people would want to include in our comments about the appropriateness of this list of health risk?

[Pause]

DR. HARRIS: Okay. Is that adequate, Dr. Ashar?

Yes, it addressed our question. Thank you so much.

DR. HARRIS: Okay, if we could have FDA read question number two for the Panel.

DR. WANG: Question 2. Section 513 of the food drug and cosmetic act states a device should be Class II, if insufficient information exists to determine that general controls are sufficient to provide reasonable assurance of its safety and effectiveness or that application of special controls would provide such assurance.

And if in addition, the device is life-supporting or life sustaining or for a use which is of substantial importance in preventing impairments of human health, or if the device presents a potential unreasonable risk of illness or injury.

A device should be Class II if general controls by themselves are insufficient to provide reasonable assurance of the safety and effectiveness.

And there is sufficient information to establish special controls to provide such assurance.

A device should be Class I if general controls are sufficient to provide reasonable assurance of the safety and effectiveness, or insufficient information exists to determine that general controls are sufficient to provide reasonable assurance of the safety and effectiveness, or establish special controls to provide such assurance but is not purported or represented to be for a use in supporting or sustaining human life or for a use which is of substantial importance in preventing impairments of human health and does not present a potential unreasonable risk of illness or injury.

FDA believes that general controls alone are not sufficient to provide a reasonable assurance of safety and effectiveness for computer-aided devices which provide diagnostic information to dermatologists about lesions suspicion for melanoma. If you disagree, please discuss how general controls alone are sufficient to provide reasonable assurance of safety and effectiveness for the device type. General controls may include prohibition against adulterated or misbranded devices.

Good manufactured practices, registration of manufacturing facilities. Listing of device types. Recordkeeping, etc.

FDA does not believe that computer-aided devices which provide adjunctive diagnostic information to dermatologists about lesions suspicion for melanoma are life-supporting or life-sustaining are of substantial importance in preventing impairment of human health. Do you agree with this assessment? If not, please explain why.

FDA does not believe that computer-aided devices which provides adjunctive diagnostic information to dermatologists about lesions suspicious for melanoma presents a potential unreasonable risk of illness or injury. Do you agree of this assessment? If not, please explain why.

FDA believes sufficient information exists to establish special controls for computer-aided devices which provide adjunctive diagnostic information to dermatologists about lesions suspicious for melanoma. Based on the information presented today, please discuss whether you believe that is sufficient information exists to establish special controls that can provide a reasonable assurance of safety and effectiveness for this device type.

DR. HARRIS: Okay. We've touched upon some of this. But, if possible, I would like to try and go in order. But, obviously, feel free to comment as you feel appropriate. The first question they are asking is: do we agree or disagree that general controls alone are not sufficient to provide a reasonable assurance of safety and effectiveness for these computer-aided devices. So do you agree or disagree with that, and if you do disagree, please tell why. Everybody put their hands down. I'm not meaning to stifle discussion. Dr. Farber.

DR. FARBER: Thank you for that. It's very simple, I agree. There is ample information.

DR. HARRIS: Does anyone have a different opinion? Perfect. All right. So, the next question, that FDA does not believe that computer-aided devices which provide adjunctive diagnostic information about lesions suspicious for melanoma… I'm trying to understand the question… They do not believe that these device are life-supporting or

life-sustaining. In other words, they are believing that these are not Class III devices. I know hands will go up now. If you disagree, please comment. Dr. Alam.

DR. ALAM: I would disagree with the last portion of that statement, of substantial importance in preventing impairment of human health. I think that's what these devices are designed to do, to prevent impairment of human health by allowing or facilitating early detection of a potentially fatal malignancy. So if you extend that, if they don't work well, then they would not succeed at their job in preventing impairment of human health, and human health would be impaired. I do think they meet criteria for a Class III device and they do not meet criteria for a Class II device.

DR. HARRIS: Thank you, Dr. Skelsey.

DR. SKELSEY: I do not agree with the FDA's statement. The purpose of these devices, as Dr. Alam said, is to detect a life-threatening condition. The intent is to prevent impairment of human health, so I don't agree with that statement.

DR. HARRIS: Okay, thank you. Dr. Farber.

DR. FARBER: Yeah, I disagree with the statement. I think the major problem is we don't have sufficient data to say that this is not going to be harmful to people if it's not working well. And we don't know if it's working well; in some ways, we don't know what is happening. For example, to the false-negatives. We don't know what is happening psychologically to these people. So, I don't think we have enough information to say that it's okay to be a Class II rather than a Class III.

DR. HARRIS: Okay, thank you. Dr. Skates.

DR. SKATES: I agree with that. Taking the words out of my mouth. We don't have enough information. I think the studies so far -- but we need the studies that Dr.

Rotemberg and I have been discussing. Until then I think Class III is the appropriate class.

DR. HARRIS: Thank you, Dr. Roth.

DR. ROTH: I respectfully agree with the FDA and disagree with my colleagues. I deal with a lot of people with cancer and I do critical care, and I think what we are missing here is the issue of time. Melanoma is a terrible disease and missing a diagnosis is a terrible thing. But I think special controls would be sufficient to counter the concerns of my colleagues and the committee. I think, even -- I would not be averse requiring that if a device is used patients be afforded a follow-up appoints in 6 months for review of lesions examined. There needs to be caveats, but we have time here. And I understand life-threatening to have a time element. A hip replacement is pretty life-threatening -- the mortality from hip replacements with these approved Class II devices is much higher than what we are talking about here. So I think we need a little bit of perspective, and I think the FDA is not unreasonable in making that statement. Thank you.

DR. HARRIS: Thank you. Dr. Suarez-Almazor.

DR. SUAREZ-ALMAZOR: I respectfully disagree with the FDA. I understand melanoma is a serious disease, and for that same reason, I think that anything that is used for screening and diagnoses has to have high standards. I think the data is very scarce and the quality of the evidence is poor. Eventually it may get better, and then it would be easier to be convinced that going from Class III to Class II is beneficial. And also, if we look at the risk-benefit of the reclassification, per se, the risks of reclassification is clear to me, but the benefits are not as clear. In addition, there are all

the concerns about using populations for which it has not been tested appropriately. So that would be my concern.

DR. HARRIS: Thank you. Dr. Rotemberg.

DR. ROTEMBERG: Thanks. I think my opinion on this reclassification is that it can't be distinguished from the special controls that I think we would need and some of which I don't even think are plausible in a Class II specification. So to me, it's not that we couldn't outline the criteria under which these devices would be reasonably evaluated. We have talked about them already today. We will talk about them with the next question. It's just that some of them are a higher bar even than the devices that are already approved, that have not been tested in prospective trials in the hands of clinicians. I think it's very difficult for us to be able to answer this question, because we are proposing such stringent standards that are so far removed from the Class II with special controls that were given to us in the executive summary.

DR. HARRIS: Okay. Dr. Ballman.

DR. BALLMAN: With respect to the question at hand, about lesions suspicious for melanoma, I agree with the FDA on this particular point. I'm not talking about special controls right now, which is the next one. But on this point, because it's being used by dermatologists – and I would move that the dermatologists are experts enough that if the device is giving them something that they really disagree with – that they would go ahead with their gut and do what is necessary. So for this particular point right here, I do agree with the FDA.

DR. HARRIS: Okay, any other comments regarding this sub-question 2? Okay, moving on to the next portion. FDA does not believe the computer-aided devices which

provide adjunctive diagnostic information to dermatologists about lesions suspicious for melanoma present a potential unreasonable risk for illness or injury. They do not believe it provides a reasonable risk for illness or injury. Comments? Agree, disagree? Thoughts? Dr. McGrath.

DR. MCGRATH: Thank you. I held off on commenting on the last comment because, while I don't think the devices provide life supporting or life sustaining importance, I do think that they carry a possible and reasonable risk of illness or injury. And I really have been thinking a lot about this as a surgeon. You know, hip and knee devices are pieces of metal, and they are --. And we do care about the device characteristics. But then there's all the play of user interface of choosing the indications for use, the technical details of the operation, whether adjuncts like glues and other things and straps are used around them and so forth and so on, which come into what you're talking about, which is the skill set of the dermatologist using a tool.

But there is a difference, I've finally decided. These devices make a diagnosis. I know that it's an adjunct and it's helpful and you don't have to listen to it, but they make a diagnosis which is something that other devices don't do. So I think that does make them different and I appreciate a comment, I believe Dr. Rotemberg said yesterday: Obviously the radiologists use these all the time, they are read ago mammogram, that's got potential and reasonable risk of illness or injury if they read that wrong. They use AI to help them sort out and make the diagnoses, but that is a reader. That's not that interface. In the immediate moment of seeing the patient. I'm struggling where that difference is from that group, gastro that use these devices. I do not support the FDA's assessments that they do not present a potential unreasonable risk of illness or injury.

DR. HARRIS: Great. Thank you. Dr. Skelsey.

DR. SKELSEY: There simply isn't enough information at this point to say that, from my mind, that it's only with special controls that these devices can be de-classified. So, I don't agree. It's an area where we are on the verge and we need these devices. There's a great deal of need for patients to be able to examine themselves; for non-dermatologists to be able to assess lesions. And I think this is a point where, if we are just not ready for prime time with these, even with these special controls — and to move forward is to the detriment of what is coming down the pipeline. So there's a tremendous need for this. But I think we are not ready, and Dr. Rotemberg outlined a lot of the future studies that need to be done, to prospective trials. But to say this is ready I think is harmful to the things that are likely to be around the corner.

DR. HARRIS: Thank you. Okay. Dr. Ballman.

DR. BALLMAN: Here is where I do disagree with the FDA in terms of making it a Class II. It says discussion whether you believe there's sufficient information to establish special controls. And so my concern is that I think these devices need to show an improvement above what a dermatologist can do. And Dr. Skates sort of, you know, went through one way of sort of showing that. I mean, my concern is that right now you know, I don't think we are comfortable with setting what the sensitivity should be, what is the specificity, and even more so, what the relationship between the two should be. I could come up with a more sensitive instrument that says biopsy everyone. Then I would be 100% sensitive. Right? But the cost would be lowering the specificity. So I think you know, it needs to be defined as how much if you increase the sensitivity, which seems to be the most important thing, then there should be you know, at no

expense of specificity, in order to approve the instrument or with a minimal decrease, but you need that risk -- that benefit/risk ratio. And Dr. Skates said one way.

So all this is saying is that you need to have agreed upon performance standards, in order for it to be a Class II, and I don't think we have that at this point, with that agreed upon performance status should be. And so that's why I think that we are not ready to define what the special controls should be.

DR. HARRIS: Thank you. Ms. Hesser.

MS. HESSER: My comment is in relation to the FDA's statement about not presenting an unreasonable risk. I do not agree with their perspective. Simply because we have not done the studies to determine what a patient or groups of patients determine to be unreasonable risks for themselves. And we do know that in other areas of medicine, patients sometimes accept different risk levels than healthcare professionals do. So I think we need validation of the statement before we can actually make it.

DR. HARRIS: Okay, thank you. Dr. Rotemberg.

DR. ROTEMBERG: I have to say for Part (D), I do believe that we know what questions we need to ask. I agree with Dr. Ballman that we can't set explicit sensitivity and specificity thresholds that the comparison needs to be with a dermatologist in its intended use setting. But you know, to Ms. Hesser's point and to everyone's point, I think these devices do have the ability to benefit health. And if we can design the studies that we would need, it would be a worthwhile exercise to do that, and to proceed with testing the devices against those benchmarks.

DR. HARRIS: So you are saying that --

DR. ROTEMBERG: I agree with Number (D).

DR. HARRIS: You agree with (D).

What about (C)?

DR. ROTEMBERG: I also agree with (C).

DR. HARRIS: Okay. Great. Thank you. Dr. Alam.

DR. ALAM: Murad Alam. I agree with (C) but I don't agree with (B). I think I'm, maybe, older and suffered more than Dr. Rotemberg. But I'm just not convinced. I'm sure she's much smarter than me and knows more about AI, so maybe I'm mistaken, but all I can do is provide my best judgment. And my concern is that, while in a perfect would the special controls might theoretically be enough, if there were very, very many of these special controls and extremely well targeted and enormous amount of resources consider to detect those, maybe it would be okay. But in reality, I think, that's not what is going to happen, not because FDA isn't very careful, but because resources are limited, time is limited, and the whole purpose of the classification is to smoothen the process for approval. I don't think they are going to be that efficient to ensure that we get where we need to get.

And Dr. Ballman's comment, I agree, we don't really know at the requisite level of detail what the pre-side performance of characteristics of these should be, so we can specify these. It's sort of a work in progress. I understand the desire to help new devices come into approval status, or premarket notification. But I really do see our primary goal, while it's important not to dampen companies and the enthusiasm and innovation, I see our primary goal here has maintaining patient safety. I think we would be remiss in meeting that goal if we move to special controls at this point, so I would

disagree with that statement and encourage FDA to maintain the current approval process. Thank you.

DR. HARRIS: Okay. I would actually like to hear a little bit from everyone, as would FDA. So not to put anyone on the spot, but I would like to hear just your brief comments on both Part C and Part D if you haven't already spoken. I would like to start with Dr. Burke.

DR. BURKE: Well, I do agree that it does potentially help dermatologists, but for Part D… We have to stay very stringent controls, and it's certainly very specific. I mean, very specific -- perhaps we shouldn't have [off mic] because we don't have enough AI to include that. We certainly have to say that you can't use this on the very anatomic sites that are most serious in Blacks -- so if we were having this, it would have to be very stringent controls and mandatory training with a certificate. And the test would be absolutely necessary as one of the controls. So, I think this should absolutely stay as a Class III device.

DR. HARRIS: Okay. Dr. Bourelly.

DR. BOURELLY: For (C) I agree with the FDA and that's based on my belief that these were meant to be adjunct and not purely diagnostic instruments, I may be misreading that. But for (D), I think it's been brought up a number of times, (D) I would disagree with, that there's still a lot of information that we need to gather before we use that in a practical setting.

DR. HARRIS: All right. Thank you. Dr. Pisarik.

DR. PISARIK: I agree with the other members of the Panel that I don't think this is ready to be a Class II device. I think there's a lot more information that needs to be

brought out. I don't think that the 9 studies published in the medical literature are enough to lessen the amount of information we need to gather on these devices.

DR. HARRIS: Okay. Thank you. Dr. Suarez-Almazor.

DR. SUAREZ-ALMAZOR: Yes, for (C) I don't really could say that I agree or disagree. What (C) says is the statement is whether they present a potential and reasonable risk and I think that we don't know if they present a potential and reasonable risk. So, I couldn't agree or disagree. I think we just don't know if they do or don't. And for (D) I would say that I think that there's not sufficient information for the controls so yeah, I disagree. Thanks.

DR. HARRIS: Thank you. Dr. Farber.

DR. FARBER: Thank you. Looking at it, I would agree that with some of the others that have made statements, stating that I don't feel there's enough information to specify controls so I disagree with (D).

As for (C), I think, again, we don't have some of the information we need to know whether there are people who are being affected, either in term of actual illness, or in progression of melanoma. Or their psychological effects from the devices. So I think there needs to be more information acquired for both (C) and (D) and therefore, I disagree with both (C) and (D).

DR. HARRIS: Thank you. Ms. Block.

MS. BLOCK: Hello, Renata Block. For (C) I go both ways and I'll tell you why. I disagreed based on the fact that I think the experience of the dermatologists, versus coming out of school, versus somebody who has been in practice for quite some time, can make quite a difference on reliability of these devices.

If it's purely used as an adjunct, then absolutely, I agree. I also think, though, that more patient data is needed in regards to solidifying this this decision. As far as (D), I think we've covered a lot of ground in these last two days, and I think the special controls definitely need to be elaborated and become more granular in regards to moving forward with any kind of classification transition to II.

Based on that information, I believe strongly that it should be maintained a Class III. We are not ready for prime time yet. Though the data that is coming down the pipeline and what we have is quite exciting. And I couldn't agree more that our future depends on it. So those are my statements. Thank you.

DR. HARRIS: Just one clarification. So do you think that we currently have enough information to identify the requisite special controls to guarantee safety and effectiveness for such a device?

MS. BLOCK: I do not. I feel we gained a lot of ground but I feel that we need, much more information for that to happen.

DR. HARRIS: Okay. Thank you. Dr. Bush.

MS. BUSH: I agree with (C). Disagree with (D) in that this is an interpretive tool, versus a visualization tool, and probably does need to have study more designed prospective in the hands -- better studies. And I do think we will get there. I don't think we should stifle this. But I this that can be obtained. But at the point I think we need more information to say that's true.

DR. HARRIS: Thank you. Dr. Gualtieri.

DR. GUALTIERI: I appreciate the complexity of the issues that we have been talking about and I think that there's a lot more information that is really needed. Not

just about the use but also, around the training and very perhaps more walls up in terms of who is using it, how it's being used, how information is being reported and accumulated in order to better understand, perhaps, the future not just of these devices but of others. I don't think that it's ready to move to Class II. I think it definitely should stay as a Class III device.

DR. HARRIS: Thank you. Dr. Skates.

DR. SKATES: Hi, Steve Skates. I disagree with the FDA on both (C) and (D). The keyword for me in (C) is "potential."

And because we don't have enough information, there's that important for unreasonable risk for illness and injury. And again the same question, same answers for (D) which is not sufficient information. Thank you.

DR. HARRIS: Thank you. Dr. Roth.

DR. ROTH: I agree with the FDA on both counts, with the caveat that special controls need to be very clear. The device needs to be used by physicians and preferable, my opinion, dermatologists only, and their staff with training, and with appropriate informed consent to patients… understanding that, for both users and patients, that this is an adjunct diagnosis and does not replace the physician's judgment.

DR. HARRIS: Okay. Thank you. Dr. Rotemberg.

DR. ROTEMBERG: I think I already answered this. But I think, I agree with the FDA on (C) and (D). But I have a very long list of special controls that I would like to propose in addition to what they have stated.

DR. HARRIS: Is it too long to review right now, or...

DR. ROTEMBERG: We can review it now or around question 3, which is around the list of special controls.

DR. HARRIS: Perfect. Okay, thank you. Dr. LaMont Bryant.

DR. BRYANT: I'm going to stay in these two questions and defer to the matter experts on these two points.

DR. HARRIS: Okay. Thank you. Ms. Hesser.

MS. HESSER: I've already disagreed respectfully with item (C) and I do disagree with item (D). I have heard the FDA say is that one of their goals is to try to reduce burden in submissions and I think that our list will become more burdensome. We are almost creating a Class II/III. Thank you.

DR. HARRIS: Thank you. So Dr. Ashar, do you have a reasonable appreciation for the Panel's opinion on this question?

DR. ASHAR: Yes, we do. Excellent discussion, thank you so much.

DR. HARRIS: So we will move on to question 3 if we could have that read for the Panel, please.

FDA READER: Question 3. FDA proposes that the following special controls would adequately mitigate the risks to health and provide reasonable assurance of safety and effectiveness for computer-aided decisions which provide adjunctive diagnostic information to dermatologists about lesions suspicious for melanoma.

Clinical performance testing will demonstrate acceptable sensitivity and specificity. Nonclinical performance testing will determine acceptable sensitivity and specificity. Nonclinical testing will demonstrate that the device operates as intended under the ants us pated conditions. Software validation and verification and

cybersecurity testing will be completed in compliance with standards. Thermal, mechanical, electric, electromagnetic and light safety testing will be completed in compliance with the standards. Bio compatibility, shelf life and sterilization processes will be demonstrated to comply with the standards. Human factors testing and hazard analysis will be performed to acceptable standards. Labeling will provide adequate information on distinction operations, intended use, intended users, intended patients, intended lesions, and body site interpretation of outputs, caution against over reliance on output, device maintenance and cleaning and the known sensitivity of the device. Discuss whether these special controls appropriately mitigate the identified risks to health of this device type and whether you recommend additional or different special controls.

DR. HARRIS: Thank you. Again I will be asking each of the Panelists to voice their opinion, but anyone that would like to go first, by all means, raise your hand. Well, I will nominate Dr. Rotemberg.

DR. ROTEMBERG: I guess I could start by going in order. I'm Dr. Rotemberg.

I think a lot of this has been said. I'll try to summarize where I think all of us agree. I think for the first point, clinical performance testing will demonstrate sensitivity and specificity. I think that most of us have expressed concerns that sensitivity and specificity is probably not appropriate, that this needs to be prospectively tested against the standard of care and show both improvement and an acceptable risk benefit ratio or risk benefit calculus.

And I think that significantly more detail is needed for the clinical controls. In terms of nonclinical performance testing, I think that, as far as my expertise with AI and

MO, I would refer to that as the validation steps in the development of the algorithm, and I think those also need sufficient transparency for factors that we know influence AI development. So just acceptable sensitivity and specificity is not sufficient. [Off mic] to be labeled and presented to us for factor such as skin tone, race, ethnicity, and atomic site, types of melanoma, and other things that we know are going to impact the AI performance. if we keep going, I apologize. Please interrupt me when you have questions. In terms of the other types of validation that we have discussed as a Panel, I think completely missing from this list is the, the evaluation of what will happen to false-negatives in practice, that can happen during trials but should also happen during post marketing surveillance. I know that there is mixed enthusiasm for marketing surveillance. Because of data drift we need to develop a novel and reliable mechanism for post marketing surveillance. Something that may not ever have existed before and is more strict than anything that exists, but the accuracy of these algorithms can completely tank over time and we need a system that monitors and um suggests updates when that does happen.

And so that is not in this list and I think we have discussed as a Panel the need for it. I think that we've also developed this sort of checklist for AI development and dermatology. A lot of it is around transparency. I would suggest that we focus on transparency of AI development and validation when we think about mitigation of potential biases and application. Those are my major comments.

DR. HARRIS: Excellent. Thank you very much. Dr. Alam.

DR. ALAM: I would like to thank Dr. Rotemberg for interesting and always very thought-provoking comments.

My view about this question is this answer is "no". And I don't really have a list of special controls that in my opinion, would be sufficient. I think it's -- this is one of those devil is in the detail issue. Will demonstrate acceptable sensitive and specificity. Many of these are in accordance with the standards. That's not transparent. I don't know what specific sensitivity or specificity would be for every tumor for every anatomic site, for every person and over time, would all of those be different? And what would they be?

My sense is that, again, while this theoretically might be possible, in practical term it will be difficult and eloquently said already, this would be trying to create some sort of II/III mechanism that is more cumbersome than the process and even then you won't have confidence it's correct because we don't know what unknown confounders might be. I know there are people highly sophisticated on AI on this call, but I would suggest there might still be some areas of concern that we might not have complete clarity on. So my sense is we have a mechanism for approval that has worked for two devices before. There's no reason to believe it wouldn't continue to work for additional devices.

And the risk of switches a mechanism at this point is high, given that the implementation would be extremely complex and not really lower the bar for everyone, while creating a lot of uncertainties that are not present in the current mechanism.

And I think that the likelihood of having the kind of randomized control trials and collecting data, while it is somewhat cumbersome, other companies have succeeded and I think that will give us the best way to evaluate these devices, going forward.

Bottom line, my answer is "no" and I do feel strongly, but I'm sure others do as well. Thank you.

DR. HARRIS: Thank you. Dr. Farber.

DR. FARBER: Thank you very much. Neil Farber. I think it's notable that the only clinical evaluation, the AI is one study on clinical sensitivity and specificity, and I don't think that's sufficient. It is -- as somebody, I think Dr. Alam you pointed out about a car and wanting to make sure it wasn't on fire or the wheels falling off and you expect that.

And I think they need much more robust studies for the clinical interaction between the AI and the dermatologist. I think they also need psychological studies of the patient, in terms of how they're reacting to the different diagnoses, either -- besides false-positives and false-negatives. So I think that there are other… And I'm not expert at doing clinical studies, in that regard. So, I would refer to some of the others who are much more expert than I. But it's clear that there are more studies needed.

DR. HARRIS: Thank you. Dr. Burke.

DR. BURKE: First, I want to commend all the comments, especially Dr. Rotemberg with your thorough thoughts and Dr. Skates with your amazing analysis of data that I try to do and made it so clear. So, thank you. Especially both of you, and everyone. I'm Dr. Karen Burke. I just -- the whole situation is the kind of… My vote is "no". And I think this is a catch-22 situation. Because if this machine is out, we have to specify that certain skins of color and certain lesions cannot be measured by this particular impedance device, but that there's a particularly, we want to include all races and we want to certainly include the most dangerous types of melanoma. So we would

want a prospective study that could include everyone, and yet we would have to say in that study that the data -- that these measurements can't be done.

We know that future instruments might not be impedance and there might be ways to analyze April and male lesions and there may be ways to evaluate lesions, maybe even in the eye because those are serious. We can't biopsy the eye and the professor of our NYU Dermatology Clinic died of a melanoma in his eye. And he was the person that taught a whole generation of dermatologists, including me and his own son. So we certainly want technology to move forward. We don't want to squelch it, but we have to have the disclaimer for when we understand from the source of the machine learning and the artificial intelligence, we have to know who we should specifically say that the machine may not be accurate for that population. So my vote is "no" but I hope that we do continue, I hope we do prospective studies and I hope all new technologies will come forward and that we will be meeting again, to see a whole new ways of measuring and assessing the data. Thank you. And it's a privilege to have been with all of you on this particular conference. Thank you.

DR. HARRIS: Thank you. Dr. Bush.

MS. BUSH: Laura Bush. I do agree we need more studies and I disagree with (D) but I realize you're looking for opinions on these too, so I do want to make comments. As far as evaluation of false-negatives, I think that's a really important thing to look at, what happened to those people, were they false-negatives for severe DNN or advanced melanomas? Post marketing surveillance and being transparent on demographics and -- our population is changing. There needs to be something to look

at that. In 5 years our population is going to look much different than it is today, even. So we need to have something built into technology, to address that.

The software standards weren't clear to me, who sets those and how well those are kept up. Because we know that's also changing like our demographics, software is changing drastically.

And I'm assuming this is all for this specific sense, this specific device, I think there does need to be more information but I'm also a firm believer that if we don't open up the discussion, we'll never get anywhere. So those are my comments on that, although I do still believe we need a little bit more information to say "yes" to a III. But I'm sorry, "yes" to a II.

DR. HARRIS: Dr. Skelsey.

DR. SKELSEY: We don't have a clear understanding of the benefits of these devices. Moreover, to make a risk-benefit analysis, we can appreciate that the risks really are different for these devices because they are applicable to a much larger population. It's not limited to patients with diabetes or knee replacement. Potentially, anybody. And so the numbers really could be significantly greater. So we don't -- I don't think these mitigation, these controls are sufficient to mitigate. And these devices have been labeled for certain provider population. We don't have the data on that nor do we have the question has risen about whether this should be used by other providers. Those studies need to be done, looking at providers who used dermoscopy, those who don't. Because the next step is the layperson using all of these. Especially when these come out. I think I'm so uncomfortable with talking about the future not knowing what other things are really out there. But if approval of these devices are

translating easily to having a whole new population having access, I think we have to be stringent about the risk-benefit ratio, and considering that a large number of people can be affected because as I said, there's not —Everybody has skin, even though the numbers of skin cancers may be relatively small, if you looked at the numbers of melanoma in comparison to other things, the numbers of patients who are eligible is large. Thank you.

DR. HARRIS: Thank you. Dr. Pisarik.

DR. PISARIK: I don't know if the special controls will make the risks that we have been discussing. I'm not sure what the acceptable risk factor ratio means. I think validation needs to be done for appeal of these skin types, especially for five and six. We need false-negatives, prospective studies to see what happens to these people that have false negatives. Updates to the algorithms, change the testing characteristics of the device and just with the testing over time, I don't think that's before characterized so I don't think they are right to label it as a type II device.

DR. HARRIS: Thank you. Dr. McGrath.

DR. MCGRATH: Thank you, Dr. Harris, you actually didn't ask me before about my thoughts about question (D).

DR. HARRIS: Please.

DR. MCGRATH: I did not agree with the FDA on that special controls would be sufficient. The reason I'm not going to reiterate all the good things that have been said, simply in light of the way that the regulatory process works, I would focus on the issue of future devices. Future devices that are making critical decisions about a very serious diagnoses. If this is a Class II, the device, other devices will have predicate status and

they would -- could be exempt from stringent analysis. Given that, again, maybe this goes back to controls and maybe there's a way to work this under controls. But I think they would have to be some very strong guide rails about future devices and the similarity to the only device that is on the market right now. But overall, in light of the just simply the regulatory mechanism, my personal suggestion would be to leave these as is Class III device.

DR. HARRIS: Thank you. Dr. Ballman.

DR. BALLMAN: Hi, this is Carla Ballman and I also disagree with this. I don't think that the special controls that are proposed are appropriate. and it goes back again to the fact that I don't think -- I think we would need to have the standard sort of clear for what the clinical performance testing needs to achieve. And we don't have that. There's disagreement even on the Panel, you know. There was no sort of risk-benefit thing. It was just that this sensitivity, the specificity seemed very low for what has been approved. So I just don't think it's ready for prime time at this point and should remain a Class III.

DR. HARRIS: Okay, thank you. Dr. Suarez-Almazor.

DR. SUAREZ-ALMAZOR: Thank you. And Maria Suarez-Almazor.

I think the descriptions are vague and especially for performance which we think is very important and also very challenging and difficult to assess -- does not provide sufficient specification to judge what is planned and would be adequate or not. There are issues such as patient population, setting for the evaluation, the follow up to address the effects of a false-negative or a false-positive result are not really explained. And then there are the other issues that Dr. Rotemberg mentioned before, consideration

of data drift and earlier transparency in the training and validation steps for the algorithm that are needed as this is one of the issues that usually not presented in studies considering these devices and seems to be very important with respect to the populations tested.

DR. HARRIS: Thank you. Ms. Block.

MS. BLOCK: First of all I just want to thank everybody on this Panel. Your insights to this have been instrumental in regarding gathering the data to make the best decision that I can, in regards to moving forward. That being said, Dr. Rotemberg, your expertise on this is invaluable and I could hear the passion of where you're going with all of this. And I completely support everything that you say in regards to agreeing with number III in regard to what is needed with this.

My concern is we don't have a lot of devices that have been under Class III, two. One, that is being used, one that is not, and Dr. Skelsey mentioned that she abandoned the one that is not in use anymore, which is concerning to me. I think we need more devices in Class III. This will allow the implementation of more special controls. It will give us a more robust knowledge-base and allow the FDA to tightly control this and have us meet again, hopefully it will be all of us again, to have this discussion. And I can see this happening very quickly, just because of technology moving so quickly. Advancing. And obviously the need for this in our future. Especially with telemedicine popularity coming down the type line.

That being said, I don't believe we are ready just yet. I think we will be ready in the future. It's just not ready now. So at this point, I believe that we need additional

information to move forward and to have this go from a Class III to a Class II. Thank you.

DR. HARRIS: Thank you. Dr. Roth.

DR. ROTH: Thank you. Katalin Roth. I think that with adequate special controls, we can move this device to a Class II. And I think that the special controls I have strongly about is that the user be trained, be a physician and be a dermatologist at this point. I have listened to -- taken everybody's comments. I'm really impressed this Panel has been really -- I've learned a lot sitting with people. But I think -- I'm not clear if moving it from Class III to Class II would change the way it is currently available. And it occurs to me that I haven't heard any restrictions on who can use it now as a Class III device. And you know, it's very expensive. I don't think that these devices are something that people are going to go out and look for. But I think access is important and diagnostic aids to experienced clinicians can be helpful. With the kinds of special controls that Dr. Rotemberg outlines and by the Panel, but specifically restrictions on the providers that can use it, I think it can be a Class II. Thank you.

DR. HARRIS: Thank you. Dr. Bourelly.

DR. BOURELLY: Thank you, Paula Bourelly. I'll start by saying I'm going to say "no" to question 3 and openly acknowledge that yesterday if you would have asked me, I would have said "yes". So I thank the Panel, as well, for expanding my thinking and teaching me things that may not have been on my radar.

I do think the controls set forth were appropriate. Perhaps not the most appropriate, and certainly not sufficient, so I vote "no". Thank you.

DR. HARRIS: Thank you. Dr. Skates.

DR. SKATES: I -- Steven Skates. On this question, it presupposes that we are going to Class II and what are the special controls. So on that presupposition I want to push back and say I don't want to push this down to Class II.

I think that there was an argument made that other devices in diagnostic settings are in Class II. However I would push back a little bit on that because I've been involved with early detection of cancer in other settings. And for most of those applications, FDA has judged it a Class III. And I would put -- in this situation, even though there's a lesion, so it's not an asymptomatic population, but for melanoma is early detection. And for that reason it should remain as a Class III and it parallels FDA's judgments in early detection of many other cancers for asymptomatic population.

DR. HARRIS: Thank you. Ms. Hesser.

MS. HESSER: I have given much thought to industry putting up a yellow flag of caution for us, and that has been done publicly. In my career, that's something that I have personally not come across before, that voice that says, maybe we can slow down our RND a little bit. That's what I heard. I ponder at the same time the ethical aspects of not keeping that guidance. So my vote would be that the time is not right, right now, to down-classify.

DR. HARRIS: Okay. Dr. Gualtieri.

DR. GUALTIERI: Lisa Gualtieri. I think in contrast to some of my esteemed Panelists, I -- there's a certain amount about how the special controls are worded that I find comprehensive. And I think where I -- what I question is how could this be done in a way that was more than just equipped, but thorough and comprehensive. In particular, I think the final point about the labeling is one that troubles me, because I feel

like kind of everything is thrown in there, and it seems more than labeling is needed. I would like to see the rubric, and I would love to see Dr. Rotemberg develop a rubric for assessing all of this, adding in the post-marketing surveillance that would be necessary.

DR. HARRIS: Okay. Thank you very much. Dr. Bryant.

DR. BRYANT: LaMont Bryant. Also [off mic] I start by saying [off mic] the development of the quality of tools. For this case, I think the Panel has been -- the majority of the Panel has been clear around the concerns around the down classification, specifically for the controls. The controls are generally allowing what you would typically see for Class II, but I understand the Panel shared their concerns.

In a broader sense, access to care is essential and as technology advances, our understanding of outcomes with either clinical or real world evidence as it evolves to think to one of the other Panelists, it would be good to have this conversation revisit. But when technologies can be democratized, they should be. When they shouldn't be, it's not appropriate, we should continue to assess. So, I want to commend the FDA for posing this question, for allowing a really good balanced presentation and robust debate. But I'll end there.

DR. HARRIS: Thank you. Before we conclude, there's a request for questions to Dr. Rotemberg from Dr. Ashar. Dr. Ashar, please.

DR. ASHAR: Great. Can you hear me?

DR. HARRIS: Yes.

DR. ASHAR: This has been an excellent discussion. We want to ask provocative questions. So we appreciate everyone's feedback. That's how we converge on getting to the right place and asking the right questions and if this was

easy we wouldn't have convened all of you. Thank you. For Dr. Rotemberg I have two questions. You mentioned transparency of the algorithm a couple of times. I imagine that you know, every device manufacturer, any manufacturer really, in any industry has some proprietary information.

Do you have -- my first question is, any advice to the extent to which transparency is necessary pertaining to these technologies?

And then my second question is: You talked in the capability in that arena. I'm wondering if you have example's from the medical device industry or other industries where post-market surveillance, has worked well as a model for something in this arena. Thank you very much.

DR. ROTEMBERG: thank you so much. And you know, Dr. Ashar, I think your questions also speak to the fact that, regardless of the classification, we're in a totally new technology arena. And so certain controls are going to be new, whether they are Class III or Class II, and I'm excited to contribute to that discussion.

As for transparency, I know that the specific data that is used for training and validation is probably not going to be accessible to the FDA when reviewed, for the purposes of, obviously, keeping some corporate needs private. That said, there have been proposed labels related to the underlying distribution of the data labeling and metadata, such as race and skin color/skin tone. Labeling metadata such as the types of labels that were made, whether it was histopathological review like we discussed yesterday. Whether it was monitoring those types of information, essentially a label for the type of data that was used for training and validation. What kind of distributions were in each of those data sets, I think would be a minimum standard for transparency

and for us to be able to evaluate potential sources of bias. And I think that list should be fairly long. Were all patients recruited from one clinic? From clinics all over the world? What kind of providers were working in collaboration with these devices?

Again, that list is exhaustive and needs to be clarified and more transparent. I know in other applications, the amount of data has been variable in terms of what has been given to the public about the training and validation data. That's something that is especially important in dermatology, where the photographic changes will change and where the impact of ethnicity and skin tone is known to be a challenge. You had a second question which I have forgotten.

Oh, post-marketing surveillance. I think I used VARS as an example. We need to sample data that is evaluated by the algorithms, over time, and compare it to the histologic ground truth and the physician's decision. And that needs to happen after the algorithms are deployed and in their intended use setting.

One of those reasons is because each place that the algorithms are deployed, may be different and have different underlying distributions from where they were tested and validated even in a prospective clinical trial, so they may not perform well even on day one. And surely, they will not perform well 6 months later, due to the contributions of dataset drift that we have talked about. And so we need to have a mechanism for sampling. I think that's really a technology question. I'm a bioinformaticist, so I have lots of ideas about how to do that. The truth is, most of our data is electronic now anyway, so comparing analogy performance with biopsy results, should not be that challenging. Does that answer your two questions?

DR. HARRIS: Dr. Ashar.

DR. ASHAR: Yes it does. Thank you very much.

DR. HARRIS: Perfect. Thank you. I would like to take this opportunity to ask our representatives Dr. Bryant and Dr. Hesser, our industry and patient representatives if they have any additional comments for the Panel meeting. Dr. Bryant, we'll start with you.

DR. BRYANT: No, I think I said it yesterday and I say it today. Thank you to FDA for asking provocative questions to try to make sure that patients have access to quality care. Thank you to the Panel for your insights, your commitment to patients and this robust debate. And Dr. Harris, thank you so much for the job of managing yesterday and today.

DR. HARRIS: Thank you. And Ms. Hesser.

MS. HESSER: Thank you, Dr. Harris. Thank you for the opportunity to bring the patient perspective to these discussions. I'm humbled, truly humbled by the exceptional caliber of expertise represented by this Panel. Much of our discussion has focused on the use of these devices by dermatologists, non-dermatologist healthcare providers, and lay users. But I ask you to consider that the true end user of any health-related technology is the patient. It is our lives that will bear the impact of the design, functionality and reliability of new products.

Regardless of the classification decision that will be made by the FDA, I encourage the inclusion of robust clinical trial data to include patient preference, or patient-reported outcomes, and a patient-education requirement if future SRA development. And thank you for sharing all of this, these past two days.

DR. HARRIS: Thank you. So, I would like to take the opportunity to add my thanks to FDA, to the guest presenters, those participants in the Open Public Hearing portion of our meeting, and of course to each and every one of the Panel members. I think it's really quite impressive and I have nothing but thanks and gratitude and respect for all the time, effort and thought you've shared with us during these past two days.

DR. HARRIS: [2:52 p.m.] If there are no other comments, then the meeting of the General and Plastic Surgery Devices Panel is now adjourned. Thank you.