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 28, 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:

Jennifer Bai, M.D.
Henry Lee, M.D.
Neil R.P. Ogden
Scott L. Kominsky, Ph.D.

Invited Presenters:

Glenn Cohen J.D.
Adewole Adamson, M.D., MPP

Summation Speakers:

Rudy Andriani, M.S.
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 risk and benefits of skin lesion analyzers (SLAs) 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 nonvoting meeting. Ask our distinguished committee members 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 Chief of the Division of Biostatistics at Cornell Medicine in New York, and I am a biostatistician.

DR. MCGRATH: Good morning. I'm Mary H. McGrath. I'm a Professor of Surgery Emerita at the University of California San Francisco, Division of Plastic Surgery.

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

DR. BURKE: I'm Dr. Karen Burke. I'm a board-certified dermatologist, and I am a Clinical Professor at Mount Sinai Icon School of Medicine in New York.

DR. BOURELLY: Good morning, I'm Paula Bourelly. I'm a private practitioner in the area of dermatology, clinic dermatology, in Olney, 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: Good morning, I'm Neil Farber. I'm Professor Emeritus of Clinical Medicine at University of California San Diego in the Division of General Internal Medicine, and I'm a general internal medicine physician.

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

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: Morning. Steven Skates, I'm Associate Professor of Medicine at Massachusetts General Hospital at Harvard Medical School. I'm a biostatistician by training with a focus on early detection of cancer.

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

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.

MS. HESSER: Good morning. I'm Dineen Hester, a long-term melanoma survivor, an oncology nurse by profession, and I'm here as a patient representative.

DR. CHEN: Good morning. My name is Long Chen. I'm the Acting Division Director for the Division of General Surgical Devices with the Agency.

DR. ASHAR: Good morning, everyone. My name is Binita Ashar. I'm a general surgeon, and I'm the Director of the Office of Surgery and Infection Control Devices at the Center for Devices and Radiological Health (CDRH) at the Food and Drug Administration (FDA), the group that has oversight over these devices. We appreciate your participation today. 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 […], teaching, speaking, writing, patents, and royalties, and primary employment.

For today's agenda, the panel will discuss the topic of skin lesion analyzer technology and its application to detecting skin cancers in various patient care settings. The skin lesion analyzer devices on which the discussion is focused are algorithm-based devices for adjunctive detection of various skin lesions, including skin cancers. We will refer to these computer algorithm-aided devices for adjunctive detection of lesions suspicious for skin cancers as Skin Lesion Analyzes, SLA.

Based on today's meeting and all financial interests reported by the panel members and consultants, no conflict of interest waivers have been issued in accordance with 18 USC subsection 208.

Dr. 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 any

other members or firms not already on the agenda within which an FDA member has a personal or imputed financial interest, the participants need to exclude themselves from such involvement, and their exclusion will be noted for the record. FDA encourages all 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 Panel Meeting on July 28, 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. I would like to invite the FDA to start their presentation. I would like to remind public observers at this meeting that while this meeting is open for public observation, public attendees may not participate except at the specific request of a Panel DR. HARRIS. The FDA will have 1 hours and 15 minutes to present. FDA, you may now begin your presentation.

DR. CHEN: Good morning. My name is Dr. Colin Kejing Chen. I am the Team Leader for the Cancer Diagnosis and Treatment Devices Team in the Office of Surgical and Infection Control Devices at FDA. We are delighted to be here to convene the General and Plastic Surgery Devices Panel. On behalf of the FDA Organizing Team, I would like to welcome everyone in the Advisory Committee, our external speakers, and everyone joining us for the discussion of skin lesion analyzer devices and the key considerations for their regulation. Welcome.

I want to start by clarifying that these next two days will be devoted to two independent meetings. Today's meeting is a general issues meeting. FDA is seeking the Panel’s input for regulations of future devices that may be intended to identify skin lesions. After you hear our presentations and additional presentations from stakeholders, we will ask the Panel to discuss and provide recommendations about three specific aspects. We will ask you how sensitive and specific these devices should be to provide benefit for the user that outweighs the risks. The sensitivity and specificity are determined by comparing the device output to the true diagnosis, or ground truth. We will ask for your input on different options in determining the ground truth. Lastly, devices that assess skin need to be tested for how well they perform across the full

range of individuals in the United States of America. We will discuss and ask for your input on options to regulate the devices so that they perform as intended for the range of patients who will use them.

Tomorrow's panel will be limited to discussions of two device types that are already FDA approved. Both of these were approved with a very specific use: a dermatologist only to provide additional information on whether to biopsy legions suspicious for melanoma. These devices are currently regulated as high risk, or Class 3, devices. We will provide more detailed information on the two devices and the proposed potential changes in how those may be regulated. It is important to note that these are two independent Panel meetings, today and tomorrow. But we will begin today's final meeting by presenting information that is important for you to consider when you discuss the questions this afternoon. We include: A clinical overview of the three most common skin cancers; the existing approaches to diagnosis; and how future skin legion devices, or SLAs, may be used by providers and patients. We will then hear presentations from outside speakers. This afternoon, we will turn our attention to you, our Panelists, as you discuss these topics and answer questions.

Skin cancer is one of the most common malignancies. 20 percent of Americans will develop some form of skin cancer in their lifetime. Early detection is key of good outcomes, particularly for melanoma. Diagnosis of skin cancer relies on examination and biopsy by dermatologists. However, the outreach waiting time to see the dermatologist is long, with one study citing an outreach of 30 days or more for new patients.

Skin lesion analyzers could play an important role in early diagnosis if they are available to non-dermatologist health care providers or even to laypersons. There is growing literature describing skin lesion analyzers powered by artificial intelligence or machine learning, and FDA has received inquiries about the regulation of such devices. Some of the devices reported in the literature analyze photographs or dermoscopic images or a skin lesion to detect visible patterns associated with malignancy, much like dermatologists do with video examination. Other devices may be applied directly to the skin to monitor physiological signals or biochemical changes that can be associated with malignancy.

Some skin analyzers are intended for use by dermatologists. Others are to support non-dermatology health care providers in deciding whether to refer a patient to a skin lesion expert. In addition, there is increasing interest in skin lesion analyzers for use by laypersons for self-monitoring and analysis of skin lesions. Many of those are smart phone-based apps.

Since these users have different levels of knowledge, the way that they rely on these devices may vary. These SLAs provide the user with diagnosis information based on the scientific evidence. It is clear that the level of knowledge is very different for each user group, with dermatologists specially trained for this area, primary care providers relatively broadly trained in this area, and laypersons typically lacking experience in this area. Although, there is a new trend. Laypersons increasingly rely on the internet for initial health inquiries. So what performance levels the device should reach for different users to optimize their performances is an important question.

Skin lesion analyzers may be applied differently. For example, some users may place a device on a specific region that they have identified as suspicious, and ask, “Is this lesion cancerous or not?” or, “Is this lesion a melanoma, or is it a pigmented basal cell carcinoma?” Others may use the skin lesion analyzers to assess any of their moles as a skin cancer screening tool. In this case, the use is to determine whether any of their moles warrant a visit to the doctor. When used for screening, the a priori likelihood of any one lesion being cancerous is lower than when the device is used to assess a selected region that has been deemed to be suspicious. Therefore, devices for screening as an aid to clinical diagnosis may have different considerations for sensitivity and specificity. With these different users and use contacts demand, and with the need to balance the benefits and risks, FDA is communing this final meeting to promote an open public discussion with involvement of our stakeholders.

As you listen to the presenters and outside speakers today, please keep our questions in mind. First, which options for determining the actual, or “ground truth”, diagnosis during clinical trials are appropriate – histology, video diagnosis by single or multiple dermatologists, or other means? Second, what are acceptable thresholds for sensitivity and specificity? Should it be different for melanoma versus other skin cancers? Should the threshold be different if the device will be used by dermatologists, versus a primary care physician, versus a layperson? And third, given the different incidents and a different appearance of skin cancer across the complete U.S. population, what regulatory approaches will support getting accurate devices to market that perform as intended in all the potential United States patients?

Once again, welcome, and thank you all for joining us. We look forward to hearing the Panel's perspective. Next, it will be my colleague, Jennifer Bai, presenting an Overview of Skin Lesions. Thank you.

FDA PRESENTATIONS

DR. BAI: My name is Jennifer Bai, M.D. I am a Medical Officer in the Office of Surgical and Infection Control Devices. I have a background in plastic surgery. I will be providing overview on skin lesions and diagnostics. Today, I will provide an overview on the epidemiology, natural history, diagnosis, and treatment of the three most common skin cancers: basal cell carcinoma, squamous cell carcinoma, and melanoma. Early diagnosis of skin cancers is important because timely treatment reduces morbidity and improves survival. I will also provide an overview of the typical clinical workflow for diagnosis of skin lesions and where skin lesion analyzers fit into those workflows.

Skin cancer can be categorized into two categories: melanoma and non-melanoma skin cancers. The most common non-melanoma skin cancers and basal cell carcinoma and cutaneous squamous cell carcinoma. Both are associated with chronic exposure to ultraviolet light. Basal cell carcinoma accounts for 75 percent of non-melanoma skin cancers, and squamous cell carcinoma accounts for 20 percent of cases. Since basal carcinoma, squamous cell carcinoma, and melanoma are the most common skin cancers, we will focus our discussion on these three types of skin cancers. Basal cell carcinoma and squamous cell carcinoma are more common than

melanoma, with a combined incidence of over 5 million new cases annually. However, these grow slowly and are rarely lethal. Melanoma, though less common, spreads rapidly and results in the greatest number of skin cancer deaths. The estimated number of new cases in 2022 is nearly 100,000 cases, with an estimated 8,000 deaths. The cost of melanoma for the healthcare sector is estimated to be $3 billion a year, with an indirect individual cost of 20 years of potential life and the intangible cost of individual patient pain and suffering. Therefore, skin lesion analyzes prominently focus on identification of melanoma to allow early detection and treatment.

Basal cell carcinoma, the most common type of skin cancer, appears most commonly on the face due to sun exposure. It classically presents as a skin-colored papule with a pearly appearance and prominent capillaries. However, it can have varied presentations in different populations and ethnicities, as shown in the image below, which is an example of a superficial basal cell carcinoma in a darker-skinned individual.

Most basal cell carcinomas occur spontaneously with no precursor lesions. Mimics of basal cell carcinoma include benign nevi, sebaceous hyperplasia, and amelanotic melanoma. Diagnosis is confirmed by biopsy. Basal cell carcinoma is treated by excision or Mohs Micrographic surgery with 90 to 99 percent cure rates. Basal cell carcinoma is slow-growing and rarely metastasizes.

Cutaneous squamous cell carcinoma is the second most common skin cancer. It presents as a scaly, this, erythematous lesion and also occurs commonly in sun-exposed areas, but can develop anywhere. Squamous cell carcinoma originates from

epidermal carotenocytes. Again, you can see that it may present differently in patients of different ethnicities and skin types.

Squamous cell carcinomas may begin as actinic keratoses, which are small, white, scaly foci of roughness that arise on chronically sun-damaged areas. Actinic keratoses are generally treated with liquid nitrogen or a topical drug without confirmatory biopsy. A portion of actinic keratosis may progress to squamous cell carcinoma. Diagnosis is confirmed by biopsy. Common mimics include: inflammatory disorders, common warts, and inflamed benign lesions. Invasive squamous cell carcinoma is surgically excised with 95 to 99 percent cure rates with Mohs Micrographic surgery.

Compared to basal cell carcinoma, squamous cell carcinoma can have higher rates of metastasis, ranging from 2 to 6 percent, especially for cancers in the 'H Zone', which is the area demarcated by the ears and central face. Once metastasis occurs, the five-year cure rate is 34 percent. Patients who are immunosuppressed are at high risk for metastasis and squamous cell carcinoma-related mortality.

Melanoma arises from melanocytes, which are cells that produce pigment. Melanoma can develop in nevi, particularly dysplastic nevi, however, 70 percent of melanomas develop de novo on normal skin. Despite its association with ultraviolet light and tendency to burn after sun exposure, melanoma can develop in both light and dark skin and on any part of the body, including palms, soles, and under the nail. Clinical assessments for melanoma include the ABCD rule, which stands for asymmetry, order irregularity, color variation, and large diameter. The letters E and F

were added to include evolution and funny-looking. Evolution is any change in size, shape, and/or color of the lesion over time. The 'ugly duckling sign' is also useful for identifying lesions suspicious for melanoma, which refers to any lesion that stands out as distinctly different from the rest of a patient's skin lesions.

The gold standard for diagnosis of melanoma is biopsy. Histological evaluation is necessary to differentiate melanoma from other lesion that may mimic melanoma, which include benign and dysplastic nevi, seborrheic keratosis, and pigmented basal cell carcinoma. Histological examination also assesses the measured thickness of the lesion, which guides treatment decisions, as well as the number of mitoses and the presence of ulceration, scarring, and immune cells in the lesion, which are prognostic factors. Definitive treatment of melanoma includes wide excision, with margins proportional to the depth of the lesion. Patients may also require several lymph node biopsies depending on the thickness of the tor. Melanoma has a high risk of metastasis and death. Metastatic patients may require additional procedures or systemic oncological therapies.

Cancer stage at time of melanoma diagnosis is critical in determining treatment options and has a strong correlation to overall survival. The combined average five-year survival rate of melanoma in the U.S. is 93.7 percent. In terms of cancer stage, 82 percent of all melanoma cases are diagnosed as localized disease, while nine percent of cases are diagnosed as regional disease, and four percent of cases are diagnosed as distance disease, although this may vary depending on subtype.

The cancer stage at time of initial diagnosis has a big impact on survival. The five-year relative survival for localized disease is 95 percent. This dramatically decreases with later stages of diagnosis, as the five-year survival decreases to 70 percent for regional disease and nearly 30 percent for distant metastatic disease. Therefore, early diagnosis is critical to reduce mortality.

The histological subtype of a melanoma is also an important prognostic factor because different subtypes are associated with different prognosis. Some of the subtypes are illustrated here, and as you can see, the subtypes have very different clinical presentations in addition to differences in prognosis. Understanding the different visual presentations of melanoma among all skin phenotypes and ethnicities represented in the United States is fundamental to ensuring timely diagnosis and effective treatment for the entire United States population.

It is important to understand a significant and nuanced aspect of assessing pigmented lesions: they do not come in just two flavors, benign and malignant. Lesions comprised of melanocytes exhibit a continuous spectrum of atypia. Some are completely benign, some have mildly atypical or dysplastic features, some are moderately or severely dysplastic, and some are frank melanomas. There are no clear boundaries within this spectrum. The red box on this slide outlines what melanocytic lesions would be considered clinically high-risk and appropriate to biopsy. For borderline lesions, dermatopathologists must at times make a clinical judgment call of whether to label a lesion as severely dysplastic or evolving melanoma in the severely dysplastic nevis. When skin lesion analyzers are designed to classify a lesion in a

binary manner, as either benign or malignant, the manufacturer must decide, and the user must know, whether dysplastic nevi are counted as positives: that is, a lesion that should be referred or biopsied. Please consider this today when we discuss ground truth, as well as sensitivity and specificity

This flowchart provides a typical workflow of how skin lesions suspicious for melanoma or other skin cancers enter and navigate the healthcare system. When patients identify a suspicious skin lesion, they may seek advice from their primary care provider or from a dermatologist. While some primary care providers are comfortable managing skin lesion diagnosis, including biopsy, others may refer the patient to a dermatologist. If the provider, either a primary care physician or a dermatologist, feels the lesion is benign, the patient will be reassured. If the lesion is likely but not definitely benign, the patient may be asked to monitor and return for re-evaluation. If the lesion is suspicious for any skin cancer, it will be biopsied. Though it is not shown here, some primary care physicians refer patients directly for excision by plastic surgeons.

FDA has approved two artificial intelligence machine learning devices for assessing pigmented lesions suspicious for melanoma, Melafind and Nevisense. These devices are intended to be used as an adjunct by dermatologists to obtain additional information to aid in a decision to biopsy for lesions suspicious for melanoma. You will hear more about these later today. These two devices will also be the specific topic of discussion at the second panel meeting tomorrow. There are currently no devices cleared or approved for people other than dermatologists in this space.

Here, we provide an example of how different devices may be incorporated into the clinical workflow. Different devices may be intended for different users, from the patients themselves to providers with differing experience in identifying skin cancer. Because of the varying levels of experience of these different intended users, there are important considerations on device accuracy. In addition to different intended users, skin lesion analyzers may have different indications. For example, some may be used to assess only pigmented lesions, whereas others may be used for screening lesions at home. Some devices will be used by patients with outlearned intermediaries, some devices will provide a binary output such as benign versus malignant, others may estimate the probability of a lesion being melanoma, and some devices may provide the specific name diagnosis, for example, 'this lesion is a basal cell carcinoma'. These considerations may also affect regulatory decisions; for example: should a device for melanoma be more specific than a device for basal cell carcinoma?

In conclusion, early detection is important for all skin cancers, especially melanoma, since early detection has a significant impact on survival. Skin lesion analyzers are emerging as potential tools to assist in earlier triage of skin cancers; however, there are many considerations for these new devices, such as the threshold for sensitivity and specificity and the clinical impact of false negatives and false positives. We thank you for your recommendations on these questions. Thank you for your time and attention. Next will be Dr. Jianting Wang, who will share the landscape of skin lesion diagnostics, specifically more on skin lesion analyzers.

DR. WANG: Good morning, my name is Jianting Wang. I'm a biomedical engineer, Acting Assistant Director for Life-Based Energy Devices Team in the Office of Surgical and Infection Control Devices. In my presentation today, I will provide an overview of the skin lesion analyzer technologies, both currently marketed and in the literature, to give you some background information on the landscape of skin lesion assessment tools and analyzers that FDA regulates. In this presentation, we will go over the technologies for evaluating skin lesions by the following type of technologies: physical examination aids, optical imaging modalities, non-optical modalities, and devices that apply software to analyze the data to provide lesion classification, becoming skin lesion analyzers. These devices, depending on their functions, feature a range of complexity levels, from simple devices that provide white light images, to more complex technologies providing tissue microstructure images, measurement of other physical properties, or analysis of measured data for detection of melanoma.

In the following slides, I will introduce some examples from each of these device types. Dermatoscopes are frequently used devices for skin lesion examination and are an example of low-risk devices that do not need FDA pre-market review. Many dermatoscopes are available over-the-counter. Conventional dermatoscopes provide white light illumination and magnification to provide better view of the lesions. Some dermatoscopes support image capture and storage to provide images for a user to assess. And nowadays, the microscope's attachments for smartphones are readily available online for lay users to purchase, so patients can take images of skin lesions at home and send thermoscopic images to their doctors with their smartphones.

Dermatoscopes do not analyze the images and they do not classify a lesion or assess risks.

There are also advanced dermatoscopes that provide more functions than conventional ones. These dermatoscopes are for prescription use, and unlike those just discussed, they are subject to FDA pre-market review. Some of these dermatoscopes use multi-spectral lights to obtain spectral information on the tissue in order to generate maps of highlight areas with high melanin, hemoglobin, or collagen content. They only provide a picture; they do not classify lesions or assess risks. In 2011, FDA approved MelaFind, which is an optical, non-invasive device that uses multi-spectral lights to image skin lesion at different light wavelengths. The images generated are analyzed by a built-in artificial intelligence machine learning algorithm to analyze the spectral information, calculate a risk score on a 10-point scale, and classify lesions in a binary way. Scores above a preset threshold are called 'Melafind positive', which means high degree of morphological disorganization, and are likely to be melanoma or high-grade lesions. Lesions with scores that are lower than the preset threshold are called 'Melafind negative', which means low degree of morphological disorganization and low likelihood of being melanoma or high-grade lesions. MelaFind is intended for use on clinically atypical cutaneous pigmented lesions suspicious for melanoma, excluding those with a clinical diagnosis of melanoma or likely melanoma. MelaFind is designed to be used when a dermatologist chooses to obtain additional information for a decision to biopsy. It should not be used to confirm a clinical diagnosis of melanoma.

In the past decades, a number of optical imaging modalities are emerging as useful tools that can image tissue in real time with very high resolution. Some of these optical imaging modalities have been cleared by FDA for general tissue imaging. They're being studied to provide additional information to users to assess skin lesions, but have not been marketed for the specific indication of skin lesion imaging and haven't been widely used clinically due to the need for special equipment and training. An example of these optical imaging modalities is reflectance confocal microscopy, known as RCM. RCM can provide 2D image of tissue cell architecture. As shown here, its penetration depth is typically within 100 micron. Meta-analysis of literature data has reported dermatologists' performance in detecting skin cancer, but reviewing RCM images, as shown in this table, with sensitivity over 90 for basal cell carcinoma in melanoma and lower specificity.

Another example of optical imaging technologies is Optical Coherence Tomography, known as OCT. OCT produces a visual image similar to ultrasound, but instead of sound waves, OCT uses near-impressed laser as the energy source to produce much higher resolution but shallower penetration, typically within a few millimeters. The example image here shows a cross-section image of basal cell carcinoma. A systemic review assessed reported dermatologists' performance with OCT in detecting skin cancer and estimated the sensitivity of OCT for identification of basal cell carcinoma at 95 percent and specificity at 77 percent.

In addition to optical imaging technologies, there are other physical characteristics measured to assess skin lesions. Navisense, approved by the FDA in

2017, is an example of a device which measures electrical impedance of skin lesions and provides an output called the electrical impedance spectroscopy score or EIS score. Electrical impedance is a measure of a material's overall resistance to the flow of alternating electric currents of various frequencies. The principle is that electrical impedance is different in normal versus abnormal tissue. By sampling both normal skin and lesion, the device provides a score to indicate where the score falls on that spectrum of normal and malignant, and thus the possibility of melanoma.

Some examples of other modalities which have been reported in the literature include high frequency ultrasounds. High frequency ultrasound provides cross-sectional images of skin where lesions can be visualized. A meta-analysis of studies evaluating the accuracy of high frequency ultrasounds to assess lesions suspicious for melanoma basal cell carcinoma or squamous cell carcinoma compared to reference standard of histological confirmation or clinical follow-up derived sensitivities with the device were 83 percent with variable specificities ranging from 33 percent to 73 percent. Raman spectroscopy is another emerging technology reported in the literature. Raman spectroscopy measures light shifts induced by molecules in the tissue. It typically provides a spectral signal, as shown on the slides. As you can tell, this data usually needs to be interpreted by software. These technologies are under development; no devices have been approved by the FDA for skin lesion analysis.

Many of the aforementioned devices provide images or measurements for the users to interpret. They are not skin lesion analyzers unless they are equipped with skin lesion analyzer software. Skin lesion analyzer software can be built in with the

device, such as in metal fines or Navisense, or it can be a stand-alone software, known as software. As a medical device, the software inputs may include skin lesion images from certain imaging modalities, such as dermatoscopes, multispectral imaging, confocal spectroscopy OCT, or some even used photos taken by smartphone cameras. Some skin lesion analyzers use other physical parameter inputs, such as aforementioned electrical impedance or Raman spectra. Some software also considers other supporting clinical data, such as a patient's skin type, age, gender, history of lesions. After analyzing the input information, the software outputs its assessments and various forms to provide adjunct information to the users. Some devices provide binary classification; for example: concerning or not. Some classify lesions into multiple classes, such as lesion type or risk level of low, moderate, or high. Some may provide a risk score in different scales. The intended users of these skin lesion analyzers also vary. Some may be for dermatologists to help making decisions on biopsy, some may be for a primary care physicians to make decisions on referral, and some software, most often smartphone apps, may be for laypersons to assess their own visions and decide whether to see a doctor.

To understand the performance of these skin lesion analyzers, it is important to know how they are developed. The software is typically based on artificial intelligence or machine learning technology. The development of the core algorithm is based on treating the algorithm with a set of lesion data, such as lesion images, with known ground tools. The algorithm learns the correlation between the image or data features in the ground tool. The algorithm is then tuned, locked, and then tested with a new set

of unlabeled data to characterize the performance and establish sensitivity and specificity. It should be noted that for diagnostic devices, the sensitivity and specificity are a pair of trade-off performance characteristics that depend on the selection of diagnostic thresholds. If you set the threshold for positive results very low, you would have higher rates of true positives, but also a higher rate of false positives, which means higher sensitivity but lower specificity. If the diagnostic threshold is set very high, then the results will be more specific, but the sensitivity will be lower. Therefore, once the sensitivity is preset the specificity will follow. Given how the skin lesion analyzers are developed, there are potential limitations and sources of bias that should be considered during device evaluation. The performance of skin lesion analyzers largely relies on the training data sites. The data sets used for training may have limited skin photo type lesion types and diagnosis lesion severity, which may generate bias in the software. Therefore, the accuracy may not be generalizable to all population or lesion types. Therefore, this will be important for discussion of which population the device is intended to serve.

In summary, I have presented to you the expansive landscape of skin lesion assessment and analysis devices, with examples of technologies that are either approved or being studied and reported in literature. You can see that there is a wide range of technologies being developed for skin lesion analysis using various optical imaging modalities or non-imaging technologies with artificial intelligence machine learning algorithms, and these devices' use is not limited to dermatologists. Various devices are being developed for use by primary care physicians or laypersons. These

devices give different types of outputs, such as risk evaluation and action recommendation. These rapidly advancing technologies, with their wide range of features, bring potential to improve skin cancer diagnosis. They also pose questions to regulatory approaches how these technologies should be regulated to assure patient benefits and adequate mitigation of risks. In the following presentations today, you will hear more about diagnostic accuracy of these devices, considerations on benefits and risks. Before we enter the discussion session, the next presentation will be by Dr. Henry Lee, who will present special considerations for diagnostic accuracy and ground truth. Thank you for your attention.

DR. LEE: Good morning my name is Henry Lee, and I am an oculoplastic surgeon and Medical Officer in the Office of Surgical and Infection Control Devices. Today, I'll discuss the diagnostic accuracy of healthcare providers for the diagnosis of skin malignancies, as well as the options for determination of ground truth for skin lesions. In today's meeting, the panel members will be asked to comment on accuracy goals for skin lesion analyzers. In order to aid in the determination of appropriate goals, it is important to consider what the current state is for diagnostic accuracy – a review of the range of sensitivities and specificities for diagnosing skin lesions by various healthcare providers, including dermatologists and primary care physicians, will establish their baseline accuracy. This information may then aid decisions regarding minimal accuracy goals for skin lesion analyzers to ensure that the devices provide a public health benefit. In addition, a review of the diagnostic accuracy of dermatitic pathologists will be provided, thereby establishing a baseline accuracy level for

histopathology, which has long been considered the gold standard for ground truth. The review of the accuracy of the gold standard will provide additional contacts to the Panel when determining if alternative methods of ground truth are acceptable in specific situations.

When evaluating skin lesion analyzers, a variety of performance benchmarks may be considered. The overall accuracy of the device, for example. Its sensitivity and specificity for the detection of melanoma can be compared to performance goals such as pre-defined sensitivity and specificity thresholds to identify the prevailing sensitivity and specificity of various healthcare providers. A literature search was completed in order to provide the panel of context for how predefined sensitivity and specificity goals may be established. Alternatively, the performance of the device could be directly compared to that of different providers, such as dermatologists or primary care providers. This could be used to benchmark performance level for devices used by a particular provider. For example, it may be reasonable in some clinical trials to assess whether a device intended to be used by primary care providers provided accuracy that was comparable to the dermatologist assessing the same lesions. In the study, this device could then, for example, be used by primary care physicians. Finally, the device's ability to improve the performance of the user may also be considered as an acceptable potential comparator for the evaluation of the benefit and risk of the device. This is measured by assessing the user's accuracy before seeing the device output and then again after seeing the device output.

In what is called a reader study, findings from a systematic review by the Cochran Skin Cancer Diagnostic Test Accuracy Group in 2018 are presented on this slide. The aim of this systematic review was to determine the diagnostic accuracy of physicians with experience with dermoscopy for the diagnosis of melanoma in comparison to visual inspection of the skin with the naked eye. The review encompassed a total of 104 publications, from which 39 data sets were identified. The review found that the sensitivity and specificity of visual inspection for the diagnosis of melanoma were 76 and 75 percent, respectively. With dermoscopy, the sensitivity and specificity increased to 92 and 95 respectively. This indicates that physicians with training in dermoscopy have high sensitivity and specificity for the detection of melanoma.

In a separate systematic review in 2018 by the Cochrane Skin Cancer Diagnostic Test Accuracy Group, the accuracy of tele-dermatology for the diagnosis of either benign versus malignant lesions, or for the diagnosis of melanoma, was performed. Tele-dermatology in particular has increased in popularity since the advent of the digital camera and the smartphone. The COVID-19 pandemic has also increased the use of telehealth, including tele-dermatology, by both primary care providers and lay people. Therefore, evaluation of tele-dermatology is important to provide further context on the present-day accuracy of providers in diagnosing skin cancer.

The rise of tele-dermatology has indicated that there are still challenges to access, and skin lesion analyzers may help to address the medical need in the general U.S. population. This Cochrane review encompassed a total of 22 studies; the review

found that tele-dermatology with photographs only can yield high sensitivity and specificity for determining whether a lesion is malignant versus benign with a sensitivity of 95 and a specificity of 84. This systematic review also evaluated tele-dermatology and its sensitivity and specificity for diagnosing melanoma. The systematic review showed that there is greater variability and sensitivity and specificity for specific diagnosis of melanoma than for binary classification of whether a lesion is benign or malignant. This may impact the Panel's deliberations regarding performance goals for skin lesion analyzers, which may have different indications for use; for example: to diagnosis of specific lesions such as melanoma, or to simply classify the lesion as benign versus malignant. We sought to assess layperson's ability to self-diagnose skin lesions; however, there are not adequate studies assessing the performance of laypersons in the United States for regulatory purposes. We asse that laypeople have little or no diagnostic ability, and that they would likely rely on the output of a skin lesion analyzer at face value. Currently, there are no FDA cleared or approved skin lesion analyzers for laypersons.

We also sought to compare the relative sensitivity and specificity of dermatologists to those of primary care providers in order to benchmark current clinical accuracy for those providers. As seen in this table, there are wide ranges of sensitivities and specificities for both dermatologists and primary care providers. The overall performance of an individual provider may be dependent on their training experience and geographic location among other variables. While there is considerable overlap between the ranges for sensitivity and specificity for dermatologists and primary

care providers for a variety of skin lesions, less variability was noted for dermatologists. For example, for the diagnosis of melanoma, a smaller range of 67 to 100 percent has to be reported for sensitivity for dermatologists; whereas a larger, more variable range of about 30 to 98 percent is reported for primary care providers. Similar findings are reported for other lesion types such as basal cell carcinoma where sensitivity and specificity are in the mid to upper-90 percentages for dermatologists, but lower or more variable sensitivities and specificities are reported for primary care providers. Overall, dermatologists appear to have higher and/or more consistent sensitivity and specificity for skin malignancies, whereas greater variability is seen with primary care physicians.

We will also discuss ground truth. We define ground truth as the means by which the true diagnosis of a lesion is obtained. A skin lesion analyzer output will be considered correct if it provides the same diagnosis as the one identified by the ground truth test. We will ask the panel about which tests could be appropriate for obtaining ground truth. We then define sensitivity and specificity or accuracy as the percent of lesions that the device identified correctly relative to whichever ground truth approach was accepted for the clinical study. There are several options for establishing ground truth. Histopathology has traditionally been the diagnostic benchmark for skin lesions and is therefore commonly utilized as the ground truth in clinical studies. For skin lesion analyzer studies, one option for ground truth could be the histological diagnosis of a lesion as reported by a single pathologist or as a consensus of a panel of pathologists. For example, in a pivotal study for MelaFind, a central histopathology lab was utilized, and each specimen was evaluated by at least two dermatopathologists. Alternatives to

histopathology have included: a clinical diagnosis made by specialists, such as a dermatologist consensus, diagnosis by panel of dermatologists, or a confirmed benign diagnosis as evidenced by long-term follow-up over a period of months. A hybrid approach, where a histopathologic diagnosis is needed for suspicious lesions, whereas a clinical diagnosis or clinical follow-up is sufficient for benign-appearing lesions, may also be considered. This alternative has been used in reported studies for benign-appearing lesions for which a histological diagnosis is not available because these lesions would not normally undergo biopsy in clinical practice. For example, in the pivotal study for MelaFind, the protocol allowed for the use of clinical follow-up for three months as a means of determining the ground truth for non-suspicious lesions. However, there are published studies that use the clinical diagnosis by a dermatologist or panel of dermatologists as the ground truth even for suspicious appearing lesions.

In order to aid the Panel in assessing acceptable methodologies for determining the ground truth, we evaluated the accuracy of the dermatopathologists for assessing melanocytic lesions. The overall accuracy of this gold standard for diagnosis may be considered when discussing whether alternatives to histopathology may be acceptable in specific situations. In 2012, Braun et al. utilized the MelaFind pivotal study data. All lesions that were biopsied during the clinical study were sent for independent evaluation by four dermatopathologists in order to determine the inter-observer variability of dermatologists in diagnosing tissue specimens from clinically difficult melanocytic lesions. A total of about 1,250 pigmented melanocytic

lesions were included. The agreement among expert dermatopathologists was measured via calculation of the kappa value, which was 0.80. The kappa value is a statistical measurement of reliability, and a kappa of 0.61 to 0.8 represents substantial agreement. In 2017, Braun et al. also reported the accuracy of dermatopathologists for an international study for the Nevisense electrical impedance spectroscopy device. A total of five U.S. and 17 European sites and 1900 lesions were included. All lesions were biopsied and evaluated by a local dermatopathologist. In addition, the pathology slides were also reviewed by a panel of three experienced dermatopathologists who are blind to the local dermatopathologist's diagnosis. The sensitivity and specificity of local dermatopathologists were evaluated and compared to the ground truth, which was defined as the consensus diagnosis of the panel of three experienced dermatopathologists. The local dermal pathologists were found to have a sensitivity of 84.9 and a specificity of 98.1 for melanoma.

In conclusion, the literature reports a wide range of sensitivities and specificities for both binary lesion classification, i.e. benign versus malignant, and for specific diagnosis of the lesion for both dermatologists and primary care providers. Dermatologists and experienced dermoscopists were found to have higher and/or more consistent sensitivities and specificities overall, including a sensitivity of 92 percent and a specificity of 95 via dermoscopic examination.

The Panel will be asked to discuss acceptable performance goals for each user: layperson, primary care provider, and dermatologist. These performance goals may vary depending on the specific malignancy; for example: melanoma versus squamous

cell carcinoma versus basal cell carcinoma. The performance goals may also be different for a device that provides binary classification; for example: 'benign' versus 'malignant' or 'biopsy' versus 'do no biopsy.' The performance goals may consist of predefined sensitivity and specificity goals in direct comparison to specialists' such as dermatologists or the device's ability to improve user performance. And, finally, histopathology has long been considered the gold standard for determining the ground truth diagnosis. The sensitivity and specificity of histopathologists for diagnosis of skin lesions such as melanoma may vary depending on the skill and experience of the pathologist. In one study, dermatopathologist sensitivity for melanoma was approximately 85 and the specificity was 98, which is comparable to the performance of experienced dermatopathologists. Therefore, the Panel will be asked to consider if alternatives to biopsy, such as follow-up examinations to confirm a diagnosis, the clinical diagnosis by specialists, or the consensus clinical diagnosis of a panel of specialists are acceptable in specific situations such as for benign-appearing lesions. Thank you very much for your attention. Next, Scott Kominsky will present special considerations with respect to skin lesion analyzers, including a benefit risk assessment and skin cancer prevalence.

DR. KOMINSKY: Good morning, my name is Scott Kominsky, and I am a biologist lead reviewer in the Cancer Diagnosis and Treatment Team within the Office of Surgical and Infection Control Devices. This morning, I'm going to be sharing some information regarding special considerations with respect to skin lesion analyzers.

The first consideration I would like to discuss is that of benefit-risk. A balanced consideration of probable benefits and probable risks is an essential part of FDA's determination that there are reasonable assurances of medical device safety and effectiveness. Benefit-risk assessment takes into account not only evidence of device safety and effectiveness but many other factors as well, including the nature and severity of the condition the device is intended to treat, or, in the case of SLA devices, to detect the benefits and risks of alternatives for diagnosing the condition and any risk management tools that might be necessary to ensure that the benefits of the device outweigh its risks. Here, we provide a general benefit-risk assessment for SLA devices. Benefits include: greater access to diagnostic information, earlier testing and diagnosis, and enhanced assessment as an additional tool, aiding healthcare providers with accurate detection, especially with borderline lesions. These benefits may result in improved disease outcome in the case of malignant lesions and a reduction in performance of unnecessary procedures in the case of non-malignant lesions.

There are also several notable risks due to false positive results. Use of SLA devices may lead to increased health care utilization and performance of unnecessary skin lesion biopsies, which carry risks of scarring pain and infection. Another risk is delay in diagnosis due to false negative results, which may result in poor disease outcome. Lastly, SLA devices may have poor positive predictive value when prevalence of skin cancer is low in a given population. Positive predictive value provides insight into how accurate a positive test result is expected to be, representing the proportion of true positive tests out of all positive test results, taking into account test

accuracy and existing disease prevalence. As the prevalence of skin cancer decreases in a given population, that is, true positives are less common; the likelihood of a false positive result increases

The risks associated with SLA device use may be mitigated in part through the performance threshold required for these devices, which the Panel will be asked to comment on later today. One possible level of performance for SLA devices is shown here using the metric of sensitivity and specificity. As you heard earlier from Dr. Wang, sensitivity and specificity are a pair of trade-off performance characteristics; that is, setting a higher device sensitivity typically comes at the cost of lower device specificity, and vice versa, as noted by the green line in the displayed graph. A higher device sensitivity will result in a higher detection of malignant skin lesions, leading to earlier diagnosis and improved disease outcome, while a higher device specificity will result in fewer unnecessary biopsy procedures, translating to reduced strain on health care resources.

When optimizing levels of sensitivity and specificity, factors that may alter the balance of device benefits and risks should be considered. One such factor is the target diagnosis in the setting of cancer diagnosis. The risk of a false negative error has more severe consequences than a false positive error, thus sensitivity may be the more important parameter in this setting. In addition, due to its higher mortality rate, a false negative error may be of even greater severity in the case of melanoma as compared to basal cell carcinoma and squamous cell carcinoma. These considerations suggest that

different thresholds of sensitivity and specificity may be appropriate for different target diagnoses.

Another factor is the SLA user. It has been reported that primary care providers assess and treat a large portion of dermatological conditions. In practice, with less experience evaluating skin lesions, it is expected that non-dermatologist health care providers may have greater reliance on SLA results when making the decision of whether to refer a patient for further evaluation and potential skin lesion biopsy. It is anticipated that laypersons will have even greater reliance on SLA results, since they are not expected to have diagnostic skills. Given the differences in diagnostic accuracy among different healthcare providers when diagnosing skin lesions, as noted earlier by Dr. Lee, and the anticipated lack of diagnostic skills expected of laypersons, different thresholds of sensitivity and specificity may also be appropriate for different users.

The second consideration I would like to discuss is that of disease prevalence. Skin cancer is more prevalent in certain populations; for example, non-Hispanic white individuals, due to their lighter skin tone. As such, the skin lesion datasets currently used for SLA device training and testing are not anticipated to contain an even proportion of skin cancer lesions occurring in both high- and low-prevalence populations. The underrepresentation of skin cancer lesions from low-prevalence populations – those with brown and black skin tones – could affect the generalizability, or in other words, the accuracy with which results can be transferred to those populations. Later today, the panel will be asked to comment on approaches towards addressing this issue.

One potential approach would be to require that SLA devices be trained and tested using data sets having an equal representation of skin cancer lesions occurring in both high- and low-prevalence populations. However, the length of time required to accrue data from skin cancer lesions in low-prevalence populations may result in a significant delay of device access to those at highest risk. A second option would be to employ a stepwise approach, wherein training and testing using datasets from high-prevalence populations is initially permitted, followed by training and testing using datasets from low-prevalence populations. This approach would allow earlier device access to those at highest risk. However, prior to device training and testing, using data sets from low-prevalence populations, it may also increase the risk of false positive and false negative results when used in these populations, since devices trained and tested on those with lighter skin tones may not perform with the same accuracy in those with darker skin tones.

In summary, towards regulation of SLA devices: it is critical that benefits and risks of device use be weighed. While several benefits have been noted, there are also notable risks which may differ based on various factors including but not limited to the device user and target diagnosis. Such factors should, therefore, be considered when establishing adequate performance thresholds used towards balancing the benefit and risk of SLA device use. Additionally, the issue of disease prevalence should be considered, which may impact diagnostic accuracy as well as device access within the U.S. population. I thank you for your attention.

QUESTIONS FROM THE PANEL

DR. HARRIS: Are there any clarifying questions from the Panel members? Dr. Suarez-Almazor?

DR. SUAREZ-ALMAZOR: Yes, thank you. I was wondering if there is any data on current utilization of some of these devices by different providers.

DR. CHEN: Can you repeat that question again?

DR. SUAREZ-ALMAZOR: Yes, I was wondering if there is any data on current utilization of these devices by different providers. I mean, how often are they used? Are they well accepted by the providers? Just to get a general idea of the use of the available devices right now.

DR. CHEN: Okay, let me run it through our team and see whether we have any data to respond to your question.

DR. ASHAR: Dr. Chen, if I could just provide a comment: as you're uh considering who you might call on from your team, just to clarify for the panel analyzers, that what we're discussing today, in general, are not currently marketed devices. There are only two devices that are currently marketed, and those will be the topic for discussion tomorrow. So essentially what we're trying to do is develop a framework by which we may be able to evaluate skin lesion analyzers, which we are defining as those devices that are not currently marketed.

DR. SUAREZ-ALMAZOR: Thank you, yeah, I understand that. I was just wondering, in general, if this were – I mean not the ones that are not marketed – but

whether the use of devices like dermoscopes or, you know, some of the ones that are available were being used currently and they had good acceptance.

DR. ASHAR: It's my understanding that dermoscopes are in widespread use day-to-day, and we anticipate as these skin lesion analyzers do come to the market, that they will also be used among a variety of individuals. Dr. Chen, do you have anything more that you would like to add?

DR. CHEN: Yeah, thank you for the additional comments. We don't have any additional items to cover. We got into those two devices, already been approved. Additional information will be provided in tomorrow's presentation.

DR. HARRIS: Okay, we'll take the next clarifying question from Dr. Farber.

DR. FARBER: Neil Farber – I guess this would be for Dr. Kominsky: I was wondering if you have considered, in the risk-benefit assessment, the risk to the patient in using — either the patient themselves using it or especially non-dermatologic users using SLAs, and the psychological risk to the patient of a false positive… has that been considered and looked at?

DR. CHEN: I'm going to turn it to Scott.

DR. KOMINSKY: Yes, thank you for that very good question. It was something that was considered, but I would defer to our Medical Officers for further information on that.

DR. ASHER: If I could clarify, the purpose of this panel is to get — the team has put together a good understanding of some of the issues that we would like for the panel to deliberate on and to think about, and so if you have specific recommendations around

the benefit-risk pertaining to these devices and recommendations for FDA on how we may consider the psychological effects that would be very helpful, especially if you have any testing suggestions or advice on how manufacturers may consider the benefit-risk associated with these devices.

DR. FARBER: That would be during the discussion section to recommend those types of things...

DR. CHEN: That certainly would be appropriate.

DR. HARRIS: Next question.

MS. HESSER: Deneen Hesser, the Patient Representative. In doing your literature search reviews, was the FDA able to identify any patient preference studies in SLAs? Was patient perspective collected in any clinical trials that supported the two SLAs? Thank you.

DR. HARRIS: Anyone able to respond in.

DR. ASHAR: Dr. Chen, would you like to take that? … With respect to patient preference, we would appreciate suggestions on how to consider that in our review in skin lesion analyzers. Tomorrow you will hear in more detail how FDA reviewed the data pertaining to the two PMA-approved devices.

MS. HESSER: Thank you.

DR. HARRIS: The next question is from Dr. Skates.

DR. SKATES: Hi, Steven Skates. Thank you very much for the great presentations. I'm keen to actually as a statistician, quantify the benefit to risk ratio so that we can then say, "Does adding the device increase that benefit to risk ratio so that

it is effectively safe, it doesn't make the situation for the patient any worse?" So the benefit to risk is, or the benefits, are due to the true positives of finding melanoma, for example if that's the aim, and the true negatives, which is saying that the patient doesn't have melanoma, if they don't have it, divided the false positives and the false negatives. And the false negatives are the big concern. If you say that patient doesn't have melanoma when they do, that's a huge problem. The false positives are less so, and we need to make some judgment about what the relative trade-off between those two false pieces of information are. In the MelaFind study, there was a ratio of about two false negatives by the dermatologist to about 1400 false positives by the dermatologist, and that's a ratio of about 700, so in working out what that benefit to risk ratio is, I weight the true positive – the melanoma the true finding, the melanoma – seven hundred times greater than the false negative, the true negative. So that that would be then comparable to what a dermatologist could do. So my push here is instead of looking at specificity to for safety look at the benefit-to-risk ratio with a weighted false positive and false negative and true positive and true negative, and make that comparable to what the dermatologists can use, and then work out what the specificity needs to be in the target population. The target population isn't all people in a specific group, it's all people with a lesion, and then you need to work out what the incidence of melanoma in that population with a lesion is. And that is different from the incidence when you divide it by all people in that population, so those two considerations, I think, should be factored into the FDA's regulatory considerations.

DR. HARRIS: I would just like to encourage the panelists that right now we're actually seeking any clarifying questions that could come from the prior presenters. I think the good points that were just made will be best discussed during the actual deliberation portion of our meeting.

DR. SKATES: I was trying to address the benefit-to-risk assessment from Dr. Kominsky and clarify that needs to be quantified.

DR. HARRIS: Is there a question that you would like Dr. Kominsky to respond?

DR. SKATES: Is making a benefit to risk quantification something the FDA would consider? Because I would encourage that.

DR. KOMINSKY: You made an excellent point, and it's definitely something we would consider and look forward to your comments on.

DR. HARRIS: Next question.

DR. BOURELLY: BO Paula Bourelly, M.D. Private practice. In Dr. Lee's presentation, there was a slide titled "Accuracy of Telederm" and showed that the sensitivity and specificity were much higher when we were simply asking by binary question, "Is this benign or malignant?" compared with whether this is melanoma. My question is, for benign versus malignant, is that for all comers, basal cells, squamous cells, and melanoma, or was that just for melanocytic lesions? Thank you.

DR. ASHAR: My understanding is that it is all comers, but let me confirm this with Dr. Lee. Is that accurate?

DR. LEE: That is accurate.

DR. HARRIS: Next question from Dr. Alam.

DR. ALAM: A comment was made by FDA indicating that for benign lesions, non-invasive following of these lesions is sometimes considered to be acceptable in lieu of histopathology. I was wondering if you could clarify what you mean by benign lesions. I guess what I'm confused about is: why would you want to follow a benign lesion at all, unless you are worried it is a suspicious lesion or at least a lesion that's borderline suspicious? Could you clarify what you mean by that? As a dermatologist, I am not familiar with this idea of following them clinically and not biopsying them if you're considering them to be suspicious, and I'm also not familiar with the idea of following a benign lesion in any way if you're not worried about it being suspicious.

DR. ASHAR: The team had described both a test data set and a training data set. The training data set would be the basis by which the algorithm is created, and then the test data set would essentially test the algorithm for clinical trial purposes. We would have incoming patients they would have lesions. Under normal practice, some lesions would be biopsied, but then there would be a cohort of patients for which no biopsy would be necessary. And the question there is, is it appropriate to biopsy a lesion where the provider or the dermatologist felt that a biopsy would not be clinically indicated, and so for that reason, in creation of the training data set and subsequently in the test data set, that's what I think the team is referring to with the following of benign lesions for the purposes of developing uh the device. Hopefully that clarifies.

DR. ALAM: So this is something that was done in the study for study follow-up purposes and not necessarily reflective of what would be done clinically.

DR. ASHAR: Yes.

DR. ALAM: Thank you.

DR. HARRIS: Next question from Dr. Gualtieri.

DR. GUALTIERI: Thank you. Lisa Gualtieri. I was interested in the required or optional training that came with these devices, the impact on the accuracy.

DR. HARRIS: Can any presenters address that question?

DR. CHEN: There are already two approved devices, and they were approved for dermatologic use. We have seen a new trend of applications, mainly intended for other providers and late persons. I believe, well now, those devices had been cleared or approved yet. But the human factors would be one of the important factors when we consider for those OTC uses. When training, in terms of, how easy it can be used. How the decision can be integrated to the workflow. That will be one of the factors that we will consider for particularly OTC BAR devices.

DR. HARRIS: Does that answer your question, Dr. Gualtieri?

DR. GUALTIERI: Yes. It does.

DR. HARRIS: Next question from Dr. McGrath.

DR MCGRATH: I was struck in our reading materials by the fact that diagnostic SLAs, while not available in the US, are being marketed in Australia and New Zealand, and also in some places in Europe. I guess that is not surprising with the presence of skin carcinoma in Australia and New Zealand. But I'm wondering if there is anything we can learn from that experience, if the FDA has any information on that… Namely, in Australia and NZ: Are these devices available for lay people? What's been the

outcome of that in those areas? And I'm just curious: where in Europe? I think that would be helpful to know also. Thank you.

DR. WANG: I can take this question. We know there are a few smartphone apps for melanoma detection available in Europe and Australia, but there is a difference in regulation for these devices in Europe. I think those apps are city-marked, so it's not based on the performance … for those devices, in Europe, they don't require performance testing. As long as they're safe, they're city-marked, so it's not based on performance. So there is also a major concern over that many literature has reported the performance of those apps are not very good. Here for those devices, of course, performance will be needed, and that's what we can discuss today and tomorrow.

DR ASHAR: Thank you, Dr. McGrath, it's an excellent recommendation and we appreciate that advice.

DR. HARRIS: There's no information regarding the usage of these devices by the lay public in Australia or New Zealand?

DR. WANG: There is data reported in literature; some apps show good performance, but there are also a lot of concerns about those studies. I won't go through the details here, but there are literature reviews available.

DR. HARRIS: Thank you. Next question from Dr. Rotenberg.

DR. ROTENBERG: This is Dr. Rotenberg. Would it be okay to answer a question from a patient advocate? Based on my own knowledge of the literature.

DR. HARRIS: Sure.

DR. ROTENBERG: There is one study that interviewed patients about their opinions about SLA devices. Tt is a very small study, it was outside of clinical trials, something like 48 patients, and there was enthusiasm for certain aspects of the devices. For example, more rapid assessment of lesions rather than waiting to see a dermatologist. It's very concordant with what the FDA has presented today in terms of the potential benefits and the potential risks. So there is one study related to that.

DR. BRUMMERT: If I may ask, did it indicate where the risk accessibility was for them. What were those patients comfortable with in terms of accepting risk?

DR. ROTENBERG: It did not quantify that. It was a very small, qualitative study that came out of Boston and it was a very small group of patients, but overall there were certain benefits that were specifically around early diagnosis or early access to dermatologist level care and concerns about accuracy and adjudication of the accuracy. Those were the main takeaways that I had from that paper.

DR. BRUMMERT: I appreciate you sharing that, thank you.

DR. CHEN: That is a great study Dr. Rotenberg just mentioned. Also, we have seen, in general digital health has been the the driving factor for a lot of innovations. We have seen that laypersons have been increasingly relying on mobile-based apps, for at least initial house inquiries. In general, we have seen that trend in the past decade and through a lot of reported literature seeking early detections are one of those factors.

DR. HARRIS: Next question from Dr. Skelsey.

DR. SKELSEY: Thank you, this is Maral Skelsey. I had a question for Dr. Lee regarding the diagnostic accuracy as dermatologists versus primary care. Were those physicians using dermoscopy in previous analysis of the comparison of dermoscopy versus non-dermoscopy, and do you have that same accuracy data using the prior analyzers that have been approved? I know we're discussing tomorrow Melafind and Navisense, but since Melafind has been approved for over ten years now, do you have that same kind of accuracy data?

DR. LEE: With regards to that larger table that was presented comparing dermatologists versus primary care providers… that came from a variety of different studies. It was a mixture of clinical information, so some studies had dermoscopy. Some studies did not and only had clinical exam findings and/or photographs. It was not only accuracy with dermoscopy. With regards to the performance or sensitivity/specificity of providers when using one of the approved skin legion analyzers, we will go through that in significant detail tomorrow. That being said the clinical study for, let's say Melafind, as referenced, had very high specificity in the 90's, as compared to very low sensitivity. We've found that in clinical use, or in these studies, that the sensitivity continues to remain high and does seem to increase provider sensitivity, but it may either positively or negatively affect provider sensitivity, depending on the study.

DR. SKELSEY: And those are real world data, or is that the original approval studies?

DR. LEE: In the original approval sensitivity 90s specificity around 10 um and then there's real world use where it showed that the dermatologist sensitivity uh

increase and specificity again the banana study either increase or decrease. But that's in the real world.

DR. SKELSEY: Thank you.

DR. HARRIS: Next question from Ms. Block.

MS. BLOCK: Good morning, my name is Renata Block. Thank you for the wonderful presentation and for having me on this panel today. My question is regarding ground truth and histological diagnosis by a dermatopathologist. Obviously, we are using the ground truth to establish performance thresholds and everything in regard to SOAs. My question to you is: if a primary care physician doing a biopsy, do they use a dermatopathologist specifically, or another pathologist-performing organization regarding the diagnosis of melanoma? I think that could make a huge difference in the data that is collected and the performance thresholds and the sensitivity and specificity of the data. So my question is: are we looking at dermatopathologists with dermatologists, and are we looking at pathologists with primary care physicians?

DR. ASHER: I think we have an expert panel here that may be able to provide input on this excellent question. I would suggest Dr. Harris ask some of our dermatologists to address this.

DR. HARRIS: Is there anyone who can comment?

DR. ALAM: I could. I would suspect that in some cases, perhaps in most cases, that pathology obtained by primary care providers may not go to board certified dermatopathologists. It is a sub field of both dermatology and pathology... I think I could say this with confidence, this is a high likelihood that a sample obtained by

dermatologists would be checked by a dermatopathologist, with the exception being a situation where the patient's insurance didn't permit that, and I think with primary care physicians, it's probably going to be more mixed in terms of what type of pathologists.

DR. HARRIS: If I can ask Dr. Pisarik to comment.

DR. PISARIK: I am not familiar with the ways these things go. When I do biopsies, I send the biopsies to the central lab, and I assume they are appropriately trained to check that out.

DR. HARRIS: Dr. Borelli, any comments?

DR. BORELLI: I can only speak about my own experience. All of my paths go to board-certified, trained dermatologists, sometimes reviewed by multiple. If it is a really gray zone area, it gets sent to a specialist, oftentimes, out of town.

DR. HARRIS: Dr. Rotenberg, you have a comment?

DR. ROTENBERG: I would agree with Dr. Alam. I have practiced in areas with primary care settings, especially in residency. Those biopsies don't always go to a board certified dermatopathologist and certainly not someone with a pigmented lesion expertise. In clinical trials, most of the pathology gets reviewed centrally. It may not be as much of an issue in prospective trials. I think that is an important consideration for real world data, and I think that was a great question.

DR. HARRIS: Any other questions? Insight regarding family practices... Next question from Dr. Skates.

DR. SKATES: In my experience in early detection of cancer, much of the framework is assessed with randomized control trials, and I want to understand from the

FDA, when you are assessing these devices as an aid to either a primary care physician or a dermatologist, that interaction is complex, and presumably it's going to be hard to capture that role. So only empirical assessment of what the impact of adding SLA to the flow compared to not having SLA in the flow and doing that in a randomized study would really assess what the real world impact is. But there's no consideration in the presentation to date of having a randomized trial, and I'd like to understand why, in melanoma or skin legions in general, that's not considered the way to go.

DR. ASHAR: This is Binita Ashar, I can address that. Excellent question, Dr. Skates. FDA regulations as it pertains to medical devices involves our center taking the least burdensome approach to address the important scientific questions. In establishing that least burdensome approach, we consider valid scientific evidence. Our regulatory definition of valid scientific evidence ranges anywhere from report forms to randomized on control clinical trials.

This is why the panel is so important. If there are key considerations, key scientific questions that need to be addressed, I think this is the place where we are looking to the panel to tell us: what key things need to be considered as part of our least burdensome assessment using valid scientific evidence?

DR. SKATES: I will make it more specific. I am concerned if the device says, no, you don't have melanoma, but the dermatologist says, yes, you do. How is that study going to conclude what the impact of that device would be? I would be surprised if it has enough currency amongst dermatologists to override their judgment that there's melanoma there. Therefore, false negatives on a device study probably don't matter

that much. If they actually do have an impact on the dermatologist and they don't refer them to biopsy and they miss the melanoma… that is a huge impact. It is very unclear to me how you will assess what a false negative will be and its impact in a real-world setting. And RCT will deal with that, but observational study what I understand is being proposed is going to be very uncertain on that

DR. ASHAR: That issue that you're raising gets at the heart of the matter in much of this. If the panel has recommendations on brainstorming how the clinical trial should be conducted, if there should be a certain rubric or protocol embedded in the study to help manage the circumstance you are talking about, if there are considerations after the study is done in weighing benefit versus risk, elucidating to us what those considerations are is very helpful... If in the post-market arena, there are lingering concerns, how those concerns may be addressed... Things that may be very helpful for our device team as we move forward together on getting the appropriate amount of clinical testing data to ensure the safety and effectiveness of these devices.

EXTERNAL SPEAKER PRESENTATIONS

DR. HARRIS: We will move on now to the guest speaker presentations. The first speaker will be Dr. Glenn Cohen followed by Dr. Adewole Adamson. They both have been granted ten minutes to speak. Dr. Cohen, you may begin.

DR COHEN: Thank you for having me. Today, I will talk about the ethics and law of creating AI models and apply particularly to the question of bias and racial bias. I will

start by having my disclosures up here. I want to start at the broadest level to expose you to how an ethicist views the perils in each stage of building and implementing and AI model. The first phase is acquiring the data. Does it matter if there was a stripping of the 18 HIPAA identifiers?

Among the issues: do patients need to be explicitly consented to the use of their data? Whether there has been a stripping of the 18 HIPAA identifiers? Would more or less be good? Is front door consent good enough? Is that too broad or too general? Do they need to be re-consented, that is the patient, for each potential use in the future, or is notice about potential use is good enough? How representative is the data? If racial and other minorities are underrepresented in the dataset, the model's predictions will be off for them, potentially hurting them. Can statistical corrections be made in such a way that can overcome this problem? If not, what resources and sense of carrots, or requirements and sense of sticks, are in place to ensure a representative data set? Now, this question bias will be the focus of my remarks today, but I want to put it in the context of the larger questions that are legal and ethical about building and implementing AI models.

Also in this bucket: what role should patients have in the governance of their health data? Should datasets be treated by laws/trusts with the trustee and executor of fiduciary duties as a union? Is there a patient's steering committee doing some governance work. How can that become a meaningful opportunity for patient engagement?

You're now at phase two: building and validating the model. How will we know if the model works well enough to be used on real patients? What standards of validation should be put in place? How should we be doing risk classification? Can we know ahead of time about possible cascade effects as a particular model is built into a device or hospital system? Who's doing most of the validating? And here, there's interesting questions about the tension between trade secrecy on the one hand and transparency on the other, and roles of third-party auditing versus governmental review.

The third phase: testing the model in real world settings. What, if anything, will patients be told about the fact that a predictive analytics model is being used to partially direct their care? Does there need to be separate informed consent for the use of the AI parts of a device? Can patients opt out? What about cases where an analytic might steer a patient towards or away from a rival risk resource and therefore opting out might be a problem? What about when analytics is working in the background? It's going to affect physician or nurse time — kind of invisible allocation. How do we think about all this as compared to informed consent for decision aids, where physicians typically don't ever consent to patient in an explicit way? There's also questions about liability: who should pay? Will there be victim compensation funds as we do with vaccines? How is liability being allocated between model makers on the one hand and physician, users, patients, and hospital systems? Also very interesting questions about choice architecture: how many overrides, how the human loop is integrated, etc.

Finally, the last phase broad dissemination. Once you have a product that works your

AI is doing a great job for patients. Now we run into a problem with equitable access. If all patients have contributed the data across the United States to build the model, the model is developed by a private sector actor and may not be accessible to all patients at all hospitals. Are there obligations to ensure access to all individuals whose data has been used to build the model? How do we actually effectuate and guarantee that?

It was important to give you the whole panoply of legalese as I see it, because I wanted to locate the one very specific one we're going to talk about today, which has to do with bias. You may have heard that AI can be biased or discriminatory, but actually, that label is a bit oversimplifying in terms of what we mean. The easiest version of the problem to see goes something like follows. An AI app built into your phone scans moles to determine which ones might be cancerous and be sent for follow-up to a dermatologist. The app has been trained largely on the skin of white men and women so the AI's predictions are quite good for those individuals, but not for Black or South Asian patients. Now here the problem is very easy to see: the training data is not as diverse as the deployment data; the problem though is typically much harder to solve, but at least we can see what it is we need to do.

In other instances, though, in fact the problem itself is much more subtle, and I'm going to tell you, this is the more modal situation. On the screen I've put this famous paper by Obermeyer and co-authors, one of the best done and most famous examinations of the problem in the healthcare space of the issue. And just to set it up, here's how they set it up in the paper, and I'm quoting now.

"Large health systems and payers rely on algorithms to target patients for high-risk care management programs these programs seek to improve the care of patients with complex health needs by providing additional resources, including greater attention from trained providers to help ensure that care is well-coordinated. Most health systems use these as the cornerstone of population health management efforts and they are widely considered effective at improving outcomes and satisfaction while reducing costs, because the programs are themselves expensive, with costs going towards teams, dedicated nurses, extra primary care appointments, loss, and other scare resources. Health systems rely expensively on algorithms to identify patients who will benefit the most. Identifying patients who will derive the greatest benefit from these programs is a challenging, causal interference problem that requires estimation of individual treatment effects. To solve this problem health systems make a key assumption: those with the greatest care needs will benefit the most from the program. Under this assumption, the targeting problem becomes a pure prediction policy problem that developers then build algorithms that rely on past data, to build a predictor of future health care needs."

Okay, so far sounds good. This sounds like a very good problem to try to solve and a good way to solve it. What do they find though? They find the model. They look at that, quote, "For each patient in the data set, they calculated a, quote, 'overall health status,' the number of active chronic conditions, or 'comorbidity score,' a metric used extensively in medical research to provide a comprehensive view of patient's health by race conditional and algorithmic risk score at the same level algorithmic predicted risk

Blacks have significantly more illness burden than whites." End quote. What do these prediction differences actually mean for patients? Quote, "Algorithm scores are a key input to decisions about future enrollment in a care coordination program, so as we might expect with less healthy Blacks scored at a similar risk score to more healthy whites, we find evidence of substantial disparities in program screening." Unquote. Why do we have this source of difference? It's because the model looks at a host of demographics, like age and sex, insurance type, procedure, codes, medications, detailed costs, but explicitly, the algorithm excludes race. Instead the problem is not because the algorithm is fed race, it's because it's a result of what it tries to predict. Quote, "The algorithm takes total medical expenditures or cost as its target, thus the algorithm's prediction on health need is in fact a prediction on health costs." Unquote. Now, prima facie, if I asked you, "Is cost a good target for this kind of algorithm to aim for?" Most people would say, yeah that seems like a pretty good target." But in actuality, it turns out that black and white patients differ tremendously on costs, and that's what produces the disparity in the re-admission algorithm, namely, and again quoting from the paper, "We find substantial disparities in health conditional and risk, but little disparity in costs. On the one hand, this is surprising. Health care costs and health needs are highly correlated, as sicker patients need and receive more care on average. On the other hand, there are many opportunities for a wedge to creep in between needing health care and receiving health care. And crucially, we find that wedge to be correlated with race at a given level of health, again measured by a number of chronic illnesses. Blacks generate lower costs than whites, on average 801

dollars less per year, holding constant the number of chronic illnesses. Or 1144 dollars per year less, if we instead hold constant the specific individual illnesses that contribute to the sum black patients generate very different kinds of costs, for example fewer inpatient surgical and outpatient specialist costs, and more costs related to emergency visits and dialysis.

These results suggest the driving force behind the bias we detect is that black patients generate lesser medical expenses, conditional on health, even when we account for specific comorbidities. As a result accurate prediction of costly necessarily means being racially biased on health." Unquote. Notice that we have an algorithm that produces racially discordant results, but adopted a target that is prima facie plausible and might have been viewed as entirely reasonable. And the problem is not data set bias, it's not that this is a data set trained largely on white middle-aged men or the like, instead it's about the parameterization. That's a much more subtle and much more difficult problem to see, unless and until you do this kind of inventory and analysis on the back end. Okay, well, what can be done? Obermeyer and colleagues at Booth in Chicago put out something I highly recommend, called the Algorithmic Bias Playbook, and they suggest in general four phases thinking about managing the bias for an institution.

The first step is just to inventory the algorithms, to talk to relevant stakeholders about how and when algorithms are used, to create of broad a list of algorithms abused, consider broad definitions of algorithms, ask questions, and then to designate a steward

to maintain and update the inventory, somebody who's going to be maintaining the inventory in consultation with a diverse group. That's the first step.

The second step, and a crucial one, is to screen for bias. You have to articulate what they call the 'ideal target': what the algorithm should be predicting, versus the 'actual target,' what it is actually predicting, and you have to think ahead of time about what kinds of mismatches could occur that could cause the bias. They call this analyzing and interrogating the bias. They say we should choose comparison groups like race and perform some basic checks of how well the algorithm predicts its actual target, then think about how label choice might create bias and how well the algorithm predicts the ideal target.

The third step, they say, is to re-train biased algorithms with a growth amount. Try retraining the model in our label closer to the ideal target; that is, you should assess possible mitigations to label choice bias by comparing results between different labels. You should consider alternative options if necessary, and if that won't work, you should consider suspending or discontinuing the use of an algorithm if necessary.

Finally, their fourth step is to set up structures to prevent future bias; to implement best practices for organizations working with algorithms. Under the aegis of the steward and diverse team, they say, conduct recurring audits and ensure rigorous documentation of current and future models.

Okay, so that's advice to an institution. What advice does that yield for FDA in evaluating a particular device with an AI-enabled software, as a medical device or the like. And in particular I'm going to use the dermatology space as my example. As

applied to this area, at a minimum, any product FDA reviews should be required to show it performs relatively well as to any skin tone, any race, any age, and any gender. That should be a minimum. If that minimum is too high for the regulator, out of even lower minimum, the label should reflect the limitations that are shown in terms of this analysis. But given how much we know, there's disregarding of labels and off-label usage. Truthfully, that's probably not enough. What you want is really a demonstration of it – doesn't even have to be exactly the same performance, but acceptable performance as between all these obvious groups. There should also be a commitment to engage in post-market evaluation and looking at operations in the real world across these kinds of cohort groups. So it's not enough that the version that FDA sees before it's out of the barn performs well on all these measures. What's necessary is that the version that is applied in the real world, which involves things like staffing, things like usage, things like human factors, also is able to be demonstrated to perform relatively well across these obvious groups. That is what I think the minimum you should think about when approving this particular kind of technology. I'll just say thank you to you, and thank you to some of the funders of the work that I do, and I hope that has been helpful.

Dr. Adamson: My name is Ade Adamson, and I'm an Assistant Professor in the Division of Dermatology at dell medical school at the University of Texas at Austin, and I direct the Pigmented Lesion Clinic here at Dell Medical School. Today I'm going to talk about the health disparities in skin cancer prevention in the age of artificial intelligence.

Some disclosures: I'm a former member of the American Academy of Dermatology's Augmented Intelligence Task Force, and I'm also a current member of the American Academy of Dermatology, Skin of Color and Skin Cancer Work Group. Nothing that I say today reflects the opinions of either the Task Force or the Work Group.

Skin cancer prevention in skin of color is a challenge, and it's a challenge for two reasons. First, the incidence, or how much disease is out there, is much, much lower in darker skin types, and two, when skin cancer develops in skin of color or patients with skin of color, it's often later, and clinical outcomes are often worse, i.e. morbidity and mortality. How can we approach prevention, and how can AI help these two issues and these two challenges.

So first I'm going to talk about the epidemiology of skin cancer in people with skin of color. I want to focus on three types of skin cancer: basal cell carcinoma and squamous cell carcinoma, together that are known as non-melanoma skin cancer, or more, recently keratinocyte carcinoma, and then finally, I'll discuss melanoma. Basal cell cancers their incidence rates vary significantly by different groups. There's almost a 1,000 to 2,000 fold difference per 100,000 population between black Americans and non-Hispanic white Americans. It's just a tumor that is not especially common in people that have skin of color, i.e. people that do not identify as non-Hispanic white. As you see here, it's about between five and six if you identify as Chinese American, 15 to 17 for 100,000 if you identify as Japanese. It's a bit higher for residents in Hawaii because of sun exposure, and it's a little bit higher in folks that identify as Hispanic. Squamous

cell carcinoma incidence rates also vary tremendously by skin type: it's only three per 100,000 population in people that identify as black, but it's one thousand to fifteen hundred per 100,000 population. So this is a very common tumor if you identify as non-Hispanic white, but not if you identify as black. The differences are not as dramatic in melanoma, where the incidence rates are about 30 times different in people that identify as black versus people that identify as non-Hispanic white. But people that do not identify as being non-Hispanic white, i.e. those that have skin of color, the rate of melanoma incidence is also really low, multiple times lower.

Non-white race is associated with later detection of melanoma. In fact if you're black you have a thirty percent lower chance of being diagnosed with stage one melanoma, compared to if you're white. And if you're black you have a two and a half times the likelihood of presenting with stage four melanoma, compared to if you're white. And the survival curves tell a very similar story. As you can see, the top curve, which are people that identify as white, they have higher survival rates than people that identify as African-American, which is the lower black line. And those of Asian Pacific Islander, American Indian, and Hispanic race or ethnicity are somewhere in between.

So there's a healthcare disparity, and it begs the question: can artificial intelligence and the power that artificial intelligence has helped with these melanoma disparities? In a 2017 article in Nature, authors were able to train an algorithm to classify skin cancers at the level of a dermatologist. At the time this was a truly remarkable feat. Now let me show you one of the key figures from that 2017 Nature paper. Here in blue you have the accuracy of the algorithm, so everything to up and to

the right of that curve, it performs at a higher level than the algorithm. And these small dots are individual dermatologists, and the green cross is the average dermatologist. So you can see, for cancers, for melanomas, and for melanomas under dermoscopic images, the algorithm performed better than the aggregate board-certified dermatologist. This was a quite a remarkable feat as I said, but one of the major problems of the study is that it lacked diversity of skin types the study actually excluded acral melanomas, which are the melanomas that are most commonly seen in darker skin types or people with skin of color. Acral lesions are moles or cancers on the palms of the hands or the soles of the feet.

I've given this talk about AI and bias and skin cancer diagnosis to many different audiences, and one of the things that they say is, I often hear at least, is that skin cancer isn't really a big deal in, you know, darker skin types, so we shouldn't get in the way of this remarkable technology. And what I usually tell them is that this is probably just the tip of the spear, trying to get algorithms to decide what is and is not cancer is just one possible use case. There are other diseases that these AI tools are currently being developed on that aren't cancer, like psoriasis or atomic dermatitis, like inflammatory diseases like that, or sexually transmitted diseases, and all of these all the lesions in these different diagnoses can look different in darker skin versus light skin. Having algorithms not trained on diverse data sets, which are representative of the population, could increase disparities. Google, in fact, developed an AI-powered tool to help users diagnose lesions and skin conditions on their own. What is really troubling is that the algorithm that was used to develop this AI tool was developed on skin that

wasn't very representative. If you look at the development set, it had less than four percent of type 5 and type 6 Fitzpatrick skin types. That's where the majority of folks of color that identify as black or South Asian or, and some folk that identify as Latinx, and certainly people in Africa and the African subcontinent. So you can imagine how this may render some diagnoses in certain skin types incorrectly. And the lack of race or ethnicity in in data sets is a problem that has been highlighted in several reviews. This is a scoping review from 2021 where they showed that the makers of certain algorithms use datasets in which race and ethnicity was reported less than 20 percent of time Fitzpatrick's skin tone information reported 10 percent of the time. In this systematic review, if you look at the characteristics of publicly available skin cancer image data sets you find that almost no reporting of Fitzpatrick skin type or ethnicity, and even more frightening to me is the histopathology ground truth in overall skin cancer lesions isn't even reported. The problems with AI in dermatology are that: machine learning algorithms are only as good as the inputs that are used to train them, and if they're not representative, then we are at risk of worsening outcomes. And yes, your skin cancer is less common and skin of color, but I think that skin cancer and how we approach regulating apps in that space is a test case for what benchmarks that we need to set up, so that uh in other conditions, we don't ignore the fact that certain disorders manifest differently in certain populations. I think we have an opportunity now to intervene before healthcare disparities potentially widen and worsen.

Now there's some potential solutions for AI and darker skin types. We can over-sample skin lesions in skin of color, or over-sample rashes in skin of color as well. We

could design a separate algorithm for darker skin tones; I would say that's less than ideal. Could you imagine having an EKG algorithm on designed on EKGs, and you know for one group of people versus another… we don't do that. They're potentially digital solutions which involve manipulate manipulation techniques to mimic dark skin, although this is also somewhat problematic, even though this has been proposed as a solution, because it doesn't recapitulate the truth . And the truth is what is required to have an algorithm that performs in the best possible way as possible. So I thank you for listening and also want to thank my funders for the research that I do, particularly in skin cancer related to overuse and under-use in dermatology.

CLARIFYING QUESTIONS FROM PANEL

DR. HARRIS: Thank you to the guest speakers. Now for any clarifying questions. And just to give some background, Dr. Cohen unfortunately is not available, but Dr. Adamson is. We will circulate amongst the panelists the two manuscripts that Dr. Cohen referred to in his talk. Are there any clarifying questions for Dr. Adamson?

DR. ALAM: Thank you for the excellent and well-thought-out talk. That gave us a lot to think about. We admire your efforts to make sure the devices are not racially biased. Again, I really enjoyed your talk, and I am curious as to your thoughts on the realm of digital devices, SLAs, for skin cancer detection? Would you prefer for such a device to be withheld from market until an adequate sample of patients with skin of color are able to be enrolled and we had the same level of confidence that we'd be able to

detect cancer in such individuals as in other individuals? Would you consider some other solution or prefer some other solution, like has been raised, such as having a disclaimer that this doesn't work for skin types 5 or 6, or hasn't been tested or designed to work for skin types 5 or 6, or whatever it may be. Would you suggest, like you said, some amount of post-marketing requirement? Would you suggest that an approval be time-limited such that, if the company didn't come up with a dataset including patients with skin of color, then the approval would expire after some period of time? I don't want to put words in your mouth but I'd just like hear your thoughts. Thank you.

DR. ADAMSON: I struggle with this question because, as I showed, the epidemiology is just drastically different. From a practical standpoint, being able to collect enough samples is not especially feasible. And so that the raises the question, should you withhold a potentially useful device for a large number of people for a disorder that is actually pretty rare in darker skin types? And where I've settled on it is, I think that there should at the very least be a disclaimer that this has not been tested in x skin types or racial or ethnic categories. But, you know, it's tough. But that's kind of where I've settled on it.

DR. HARRIS: Next question from Dr. Rotemburg.

DR. ROTEMBERG: Thank you for such a great talk. One of my questions is about the comparison to dermatologists that you showed. We've heard from the FDA earlier today that the accuracy of dermatologists is not the same with an in-person evaluation as compared to a tele- or a remote evaluation. How seriously should we

take those ROC curves, given that it's a reader study as compared to an in-person evaluation?

DR. ADAMSON: So that's an excellent point, and what I would say is that, just about every single AI study I've seen comparing, the technology to dermatologists have been reader studies, which doesn't mimic the dermatology visit, right? Where you look at moles, not only in isolation, but in context with other moles. You also have metadata, like asking the patient, how long has it been there or what kind of sun exposure you have, et cetera et cetera. I think extrapolating those, or making that comparison, isn't necessarily a valid comparison, if you will. And in the real world, those splendid curves that you see in these AI algorithms – I'm sure their decrement will decline.

DR. HARRIS: Next question from Ms. Block.

MS. BLOCK: Dr. Adamson, thank you for an enlightening lecture and really shedding light on the importance of skin of color and representation in dermatology, not only with images of but also histology. I think that's very important for the FDA to consider moving forward with any SLA products on the market. My concern is, and I want your advice on the products on the market now, or FDA approved class 3 devices are not intended to diagnose or help adjunct diagnosis with lesions on special anatomic sites, such as acral palmar and plantar surfaces. As you know, in skin of color, that is typically where melanoma can be found. Is it a concern for you that these devices do not focus on that, and do you feel like more time is needed for the devices to be technologically ready to do so before it is proposed for use in all skin types?

DR. ADAMSON: I think that whatever algorithm gets developed and approved, whatever company is doing that needs to have put forth some good faith effort in getting as diverse population within their dataset as possible, or at least in such a way that is reflective of the epidemiology of the disease as it relates to skin cancer, as an example. And I do think that there needs to be some emphasis on acral lesions. I mean, at the very least in my opinion because as you said, you know, those are the ones that disproportionately affect folks of color. Or if you don't, say on the disclaimer not to be used on the palm, soles, special sites, et cetera. And honestly that would go for also patients that identify as white, right? Because those patients also would benefit from having a device that found those lesions as well.

DR. HARRIS: Thank you. Next question from Dr. Farber.

DR. FARBER: Neil Farber. Thanks so much for that talk. Going a little bit sort of out of the realm of developing the actual algorithm, might it be useful for changes in the sensitivity and specificity, specifically for patients of color, so that basically if a lesion was considered, it would be oriented more towards looking at higher sensitivity and less specificity, so that the lesion would be addressed. I'm asking that because of the fact that you mentioned that a lot of times you hear, well, "lesions in people of color are not any big deal anyway," which is absolutely not the truth. And so I was wondering how we could address that, and might it be addressed not specifically in developing the — well, one thing would be to increase the number of people of color in the testing but also to increase sensitivity for acral lesions — but in addition, looking at a way of setting the parameters for the AI differently in people of color.

DR. ADAMSON: I think that will be a challenge, because if you don't have an examples in your training set, then I think it will be hard to really tune anything very much if your data is sparse to begin with. I don't know whether setting the sensitivity gain higher or lower would help that much if, like I showed you from the Google app, they have almost nobody of color in the study.

DR. HARRIS: Next question. Dr. Bryan.

DR. BRYANT: Yes. LaMont Bryan. Dr. Farber kind of touched on it a little bit, but I'll go back specifically to clarify a point, Dr. Adamson. Earlier, during Dr. Alam's questions specifically around representation and labelling, I just want to clarify: In your response, you mentioned feasibility as it relates to making sure we had all skin types representative, so I want to thank you for your data. Very enlightening, but then, two, I guess I want to dig down and discuss how representative data collection is difficult but feasible, and the fact that you do need the representation as articulated by your point on Google. Can you clarify your feasibility point?

DR. ADAMSON: If we're talking about developing algorithms for skin disease at large, that is a different question than developing an algorithm for skin cancer, because in skin cancer, I would say — if we just take melanoma as an example, the amount of new melanomas diagnosed in the United States among Black Americans is in the hundreds. It's maybe 400, 500, right? In folks that identify as white, it's something like, if you include melanoma in situ, it approaches 200,000. Okay? So just by that pure epidemiologic mathematics, it would be hard for you to get all of these lesions in the Black population, in order to power an algorithm, right? But if you think about skin

disease at large, people of all — I mean, humans have skin problems that need diagnosing and being sure that these algorithms take that into account in their development is important as a larger issue. But if we're talking about specifically skin cancer, that feasibility I'm talking about is being able to train an algorithm in a disease that's pretty rare in, let's say Black people, is a challenge.

DR. BRYANT: A challenge, versus not feasible. I understand.

DR. HARRIS: Next question, Dr. Bourelly.

DR. BOURELLY: Thank you for that presentation. I have one quick clarifying question. We tend to go to the extremes and talk about very White skin vs. Black skin, and you were in clear in saying there is some degree of variation when you talk about Latinx, when you talk about Middle Eastern, when you talk about very fair- skinned Black people. So I believe that we're talking about Fitzpatrick's 3 through 6 rather than just 5 and 6. It would potentially be left out of the evaluation, I mean, if we really want to look at the population that's most affected, we classify them as Fitzpatrick 1 and 2. That's a pretty big chunk of people that will not be included in the evaluation.

Really quickly also, I just want to re-define what most vulnerable sounds like. We think of most vulnerable as being the highest number in a given population who's affected by this, but I would argue that sometimes the most vulnerable is a person who doesn't get diagnosed, who gets late detection, because the consequences of that lack of diagnoses are going to be more dire. And quickly, the last thing is, one thing I try to do in my own practice which is very helpful is I try to get everybody undressed. Not everybody goes for it. But you come in for acne, I also offer a skin exam and by doing

that, I actually increase the number of patients that I can actually evaluate for skin cancer at the same time, even if they're low risk. I have diagnosed melanoma, acral melanoma, on an African- American man who came in for a contact dermatitis from a boot. He had no idea it was there. So I think if we all get into the practice of screening everybody, before you knew it, you will have that population. If you're waiting just for a Black person or a dark-skinned person to come in for a skin cancer review, I agree with you 100% that won't happen, but if they're coming in for something else and we decide to surveil them, it takes all of five minutes. All of a sudden you have a population right there from which to draw. Thank you.

DR. HARRIS: And our last clarifying question from Dr. Bush.

DR. BUSH: Thank you. Excellent presentation. And I love the comments of Dr. Bourelly. I agree with you. Would you suggest studies mirror the U.S. population skin of color making it for feasible to enroll these patients, taking in account the changing landscape of our population over time?

DR. ADAMSON: I do think that, at least for skin cancer, there should be some representation of folks of color. Maybe it's small, but that reflects the epidemiology of the disease. And so I think that that's important. But I'll also make another point: some of you may know, I'm very interested in screening and what that means over diagnosis, all of this kind of stuff. And I think that because melanoma is so rare in the Black population — actually, if you stack it against the 50 or so cancers that CDER tracks, it is probably the one that kills the least, right? So that's just the data, okay? That doesn't mean that it's not consequential, but when you think about something as labor intensive

as screening… say just screening, everybody Black should come to the dermatologist to get screened for melanoma… I'm not sure if that's the most productive use of the limited resources that we are, as dermatologists. I think perhaps an app maybe could help re-stratify some patients if it actually worked and correctly identified who should come in to get checked and who shouldn't. And so that's just kind of where I sit as it relates to screening in a low- risk population.

OPEN PUBLIC HEARING

DR. HARRIS: Well, thank you very much, Dr. Adamson, for your presentation and response to the questions. We'll now move onto to 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.

MS. NALLS: Both the Food and Drug Administration, the FDA, and the public believe in a transparent process for information-gathering and decision-making. To ensure such transparency at the Open Public Hearing session of the Advisory Committee Meeting, FDA believes that 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 or a 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 such financial relationships. If you choose not to address this issue of financial relationship at the beginning of your statement, it will not preclude you from speaking.

DR. HARRIS: Thank you, Ms. Nalls. FDA has received two requests. Each speaker will be allotted five minutes to speak, and our first speaker is Dr. William Steffes.

DR. STEFFES: Thank you, Mr. Harris. I appreciate the opportunity to speak to everyone today. My name is William Steffes, I'm a private practice dermatologist in the central Florida area. I am being compensated for my time by SciBase to offset the loss of clinic revenue that I'm incurring right now. I've been practicing for about seven years since residency, and I've always had a very strong interest in the early diagnosis and treatment of skin cancer, especially melanoma. And so I think this is a very important topic when we start talking about skin lesion analyzers and how they can help us. In particular, I've been using Nevisense for about two years to obtain EIS measurements on pigmented skin lesions. I believe it to be a very useful device, but of course, it's most useful when added to your clinical and dermoscopic impression of skin lesions. And I think it's really helped me over the last couple of years find even more subtle melanomas and very small lesions since I've started implementing it in my practice. That being said, I think most of the panel members will agree that the diagnosis of skin cancer is complex and it requires the skills of a trained dermatologist, and you have to

consider all clinical parameters including the clinical impression of the lesion, the physical exam, patient risk factors, dermoscopy, et cetera. So I think that these skin lesion analyzers such as Nevisense do provide us with excellent information that we can integrate into our decision-making process to help guide our biopsy selection.

In particular, I find that the devices such as Nevisense are most useful when looking at intermediate or hard to determine lesions, so I don't use them for lesions that are obviously melanoma or I have a very strong suspicion, but in the middle-of-the-road type pigmented lesions is where it's most used. That being said, the most important function for any device is that it performs with high sensitivity and that the data that's being output is trustworthy, because only if the devices are reliable can we be certain that they're not providing false negatives. False negatives can have devastating consequences for patients, not only if you miss the melanoma, but even worse, the patient goes home with a false sense of reassurance and which could lead to an even longer delay in diagnosis. And that can result in metastasis and death when we're talking about melanoma. So dermatologists like me really depend on the FDA to ensure that devices such as Nevisense and others that may come out are adequately tested and that the clinical trials that they perform are inclusive, that they're robust, and that they're performed properly.

I think each individual device should be evaluated individually using agreed-upon histopathological correlations to make sure that they're safe and that they're effective. To be frank, doctors are oftentimes very busy and I don't think it's possible for us as physicians to deeply analyze every clinical trial that leads to the approval of a new

medicine or a new device, and I personally see the FDA as being very essential in this regard, because as clinicians, we need to know that if we choose to use a device, that its going to be trustworthy. And even for importantly, our patients, depend on us to be using devices that have been tested and that are reliable. And the risk to patients' health that could result from using unsafe skin lesion analyzers for example, I think are far too significant to lower regulatory controls for these devices.

So in conclusion, I know I only have a few minutes but I just want to reiterate that as a dermatologist I think it's very, very important that diagnostic devices, skin lesion analyzers, are thoroughly and properly tested through high-quality clinical studies. The information that these devices provide us can be very helpful when used in the right hands, and I think we can save lives by diagnosing melanoma earlier. But I think the consequences of using inaccurate devices could be devastating and could lead to unfortunately to bad outcomes. So I just want to strongly urge the FDA to take responsibility for a rigorous review process and to keep devices, for now, in the class III to make sure that the ones that come to market are being tested adequately. Thank you very much for your time. I appreciate it.

DR. HARRIS: Thank you for your comments, Dr. Steffes. Our next speaker will be Mr. Simon Grant. Is Mr. Grant available? If not, I'm to understand he may be replaced by Mr. Parspinhog.

MR. GRANT: My name is Simon Grant, and I'm the CEO of SciBase, and SciBase is the developer of Nevisense, and SciBase has been working within melanoma detection for nearly 25 years. We have the most experience of any company

globally when it comes to melanoma detection, and we also have the most experience in the U.S. Today, we have over 85 peer-reviewed articles about our technology. Our product, Nevisense, it used for the detection of melanoma at the point of care. Nevisense is an AI-based device used by dermatologists when trying to decide whether to biopsy a lesion or not, and Nevisense is the only skin lesion analyzer available in the U.S., and so SciBase is the only company today that has direct clinical marketing experience of skin lesion analyzers. You discussed the situation in Europe previously, and also in Europe, Nevisense is the only lesion analyzer approved under the new NDR regulations, and I think it can be worth noting that in Europe, the regulations are becoming more stringent, more strict, whereas FDA here is proposing the opposite direction.

Another product, Melafind, has been discussed, and I think it's important that the panel understands that Melafind never really got going in the U.S. and was withdrawn from the market six or seven years ago. So we can't really look so Melafind to understand the clinical realities of skin lesion analyzers in the U.S. today. SciBase is the only company with direct insight into the use of skin lesion analyzers in the U.S., and we feel we can speak to the risks and the challenges of these products. We have been unfortunately only been allocated five minutes to talk, but so we're going to superficially go through several of the areas but we'll try to get through it. Melanoma is a high-risk disease. A false negative, I think as Dr. Steffes said, or a miss melanoma, is just not delayed detection. It can be fatal. In our years of experience, we see that this is the area of dermatologist that clinicians are most concerned about getting it right. It

simply can't be wrong. So when validating new devices, ground truth needs to be as accurate as possible. If the ground truth is incorrect, a study can always state the sensitivity of a new device. Quite simply, it can appear better than it is. For us it is clear that the ground truth must be based on histology.

Histology, though the gold standard, is not perfect, and so a panel of determined pathologist is the safest way to establish ground truth, and this is what was used in Nevisense's validation study. And when that panel disagreed, we had a second panel that we referred to. So this is complex. As part of our approval process, we designed a prospective validation study for our technology together with FDA. That process took months, if not years. FDA considered the patient risk was high, and so we ended up designing the largest ever study within melanoma detection, and it’s our opinion that nothing has really changed from a risk perspective. We think that this cooperation with FDA in study design is one of the key reasons Nevisense is a safe and effective device today. The design of clinical validation studies is absolutely critical, and it's a complex process that requires direct FDA input. FDA has proposed thresholds for sensitivity and specificity. We agree with the sensitivity target. At that sensitivity and negative predictive value, we see are the two host important thresholds. But what is much, much more important are the details of the study design and especially which patients and which lesions are included. Dermatologists often tell us that they can see a late stage melanoma as the patient walks in the door, and, on the other hand, it's fairly easy for them to identify obviously benign pigmented skin lesions. This is not the challenge for dermatologists and it's not where skin lesion analyzers can add value for patients.

Dermatologists need a tool to evaluate atypical lesions, small lesions, lesions that are on the gray side and could be early stage melanomas. Again, it's extremely important that FDA's involved in study design to ensure that adequate numbers of these types of lesions are included, and this is not as easy as it sounds.

Furthermore, different technologies propose different challenges in different clinical situations. Different technologies can be affected by different skin types. Fitzpatrick, ethnicity, age, lesion location, lesion size, ulceration, and previously acral skin was brought up. That's a perfect example; very much technology dependent. A standardized clinical trial designed from vastly different, and even yet-to-be-developed technologies, we don't think is appropriate. In conclusion, SciBase knows that when it comes to designing validation study, the devil is in the details. And we are very concerned that setting standardized performance goals will not be adequate to effectively evaluate new devices based on very different technologies. It's actually very early when it comes to clinical experiences with skin lesion analyzers. Nevisense is the only device where we have U.S. experience, recent U.S. experience. And even that experience is relatively limited.

Finally, we believe that the risk to patients has not changed. It's still high. All this pointless, continued rigorous oversight by the FDA. When there's significant or sufficient experience from multiple technologies and broad clinical use, standardized study guidelines and down-classifications can be revisited. Until then, we're concerned. We believe that the reduced level of oversight that will come from down-classification

and standardized approval guidelines will result in miss-melanomas and even patient deaths. For melanoma it is simply too early. Thank you.

DR. HARRIS: Thank you very much, Mr. Grant. And now I'll see if there are any questions from our panel for the open public hearing speakers. We have a question from Dr. Alam.

DR. ALAM: Thank you, Mr. Grant. You indicated that you felt that if a device were not being used by a dermatologist, but was more free standing if you will, that it should have a higher sensitivity than the 90% benchmark that has been proposed. Do you offhand have a suggestion for what that higher level of sensitivity specifically should be? Thank you.

MR. GRANT: I don't really think that's my position to say, but it definitely should be higher. Remember that 90% that the lower confidence bound so observed sensitivity is actually much higher. So but when you put a device in the market, you know, to be manager bid people who don't understand melanoma I think you have to have a much, much higher sensitivity. You have to have almost 100% sensitivity because you can't send those patients away with a false reassurances as Mr. Steffes discussed. I think it should be very high. I mean, I don't think it's my position to say, but definitely higher.

DR. HARRIS: Okay. Ms. Block?

MS. BLOCK. Thank you so much for the presentations, Mr. Grant and Dr. Steffes. I noticed that Nevisense should not be used clinically on obvious melanoma, which is my concern, using it in primary care physicians or laypeople because, obviously, they don't have the clinical skills of diagnosing or noticing a melanoma.

What's your stance in the future in regarding these types of devices for non-dermatology providers?

DR. STEFFES: Well, I think, you know, it's a good point. As I said in the beginning, if something's an obvious melanoma, we know you're going to biopsy it. If you're worried about it, you should biopsy it. But that's not the sweet spot for these kind of products. So that's why it's not indicated for if it's an obvious melanoma. You know, you should go straight to biopsy, and we've always said that. When it comes to other groups using the device, we think that, providing that there are studies that show it works, that could be something that comes down the line. Today, we only sell to dermatologists, okay? But down the line, if we do the studies that are required, yeah, then we will have a case that we can present that it works in that situation. But it's a much bigger challenge to design studies for general use, you know, for non-dermatologists in a broad set of sort of lesions and these are – our study was 2,000 patients, but you're talking about very, very large studies and where every lesion was biopsied to know exactly what the status of it was. So this would mean, we think, and all we're trying to say is that FDA's input is required. We believe this should remain as rigorously controlled by the FDA.

DR. HARRIS: Our next question, from Dr. Skates.

DR. SKATES: Thanks very much for both of your presentations. For Dr. Steffes, can you give me a sense of what the ratio is between those that you — of the lesions that you were uncertain that you use Nevisense for to guide your judgment, how many false positives that weren't malignant melanoma, versus for every melanoma that you

found in that uncertain group? Did you experience — was it 10 false positives for every melanoma, or 100 o…? It would help us quantify this benefit to risk ratio that you find currently acceptable.

DR. STEFFES: Thank you for the question. It's a very good question. And I think, I don't know the exact number offhand. I think it's how you define a false positive. If we're – because sometimes when I use the Nevisense on a lesion, often times it will be a moderate or severely dysplastic nevus or an atypical melanocytic proliferation. So I think if you're counting those as false positives, I would probably say it's closer to the 10 number that you stated earlier. Again, I think as dermatologists, I think we're pretty experienced and skilled at selecting atypical lesions, and so most of the lesions that I test with Nevisense usually come back with at least dysplasia, and some of them come back as melanoma. I think the trick is, when you see a lesion that you're only mildly suspicious of clinically and you get a higher reading with the Nevisense or a similar SLA that comes down the line, is that it pushes you to do a biopsy that you might not have otherwise done. And therefore, you capture some additional melanomas in that regard. I don't think I totally answered your question, but I would say that – because I don't know the numbers personally offhand; I would have to look it up – but I think if you were not counting dysplasia and atypical melanocytic proliferations I would say 7, 8, 9… to 1. Somewhere in there.

Dr. SKATES: Okay. Great. Thanks. And Dr. Grant, do you have any summary from your study in the U.S. as to what that ratio might be? And the reason I ask this, just to put it in context, I would like to see the SLAs, if they were to be put into the hands

of primary care physicians to at least raise the accuracy level, to what a dermatologist could do without the SLA. Do you have an estimate from your studies of what the ratio of — I'm really interested in false negatives so the way you missed to false positives where you said that the lesion might be a -- it needs a biopsy but turned out not to be malignant and that will help us quantify this benefit to risk ratio. I did find the Melafind study on FDA. They had a link to that, but there wasn't a link for Nevisense, and having those data would be really helpful.

DR. GRANT: The ratio was 7.3 to 1 in our clinical study.

DR. SKATES: Okay. As a biostatistician I'd be very interested in getting access to the same document that Melafind study had access.

DR. GRANT: Yeah. And we would be very happy to supply that to you. We were surprised it was not included as well, so it's… yes. This was a really large well designed study and I think, you know.

DR. SKATES: And I'd have to agree with you that the details of the study design are crucial as a biostatistician and early detection researchers so I applaud that attention to detail and I'd like the FDA to, you know, it's great that the FDA is putting that level of effort into it.

DR. GRANT: Yeah. I think this is what we're trying that say, is that cooperation's essential, we think, because the technologies are different, and so we really need to make sure that you're taking account of those differences when you design the studies.

DR. SKATES: Thank you.

DR HARRIS: Great. Our next clarifying question from Dr. McGrath.

DR. MCGRATH: Mr. Grant, I just want to be crystal clear about something you said about the rigorous FDA governance of devices. I realize we're not discussing the reclassification of Nevisense until tomorrow.

DR. GRANT: Yeah.

DR. MCGRATH: But are you saying — are you just speaking now about forthcoming diagnostic devices, or are you saying that think that devices that are already on the market, like Nevisense, should remain in a Class III category?

DR. GRANT: Yeah. That's what we believe. We believe the Class III is the correct category class for these types of types.

DR. MCGRATH: The ones that are currently on the market.

DR. GRANT: Yes. And new ones. Yeah.

DR. MCGRATH: Thank you.

DR. HARRIS: Thank you. Next question Ms. Hesser.

MS. HESSER: My question is for Mr. Grant. I'm Deneen Hesser, the Patient Representative. Thank you for both of your letters and the presentation you've made. In the real world, many patients engage first with a physician's assistant or a nurse practitioner in a dermatology office. What would it take to extend the use of these devices to that group of health care professionals?

DR: GRANT: Well, the way we look at it is that as long as it's used under the instruction or the supervision of a dermatologist, that's okay. And so P.A.'s can be involved, for example, and are involved today.

MS. HESSER: Do you currently offer specific training for those groups of health care professionals?

DR. GRANT: We do. We'll put it this way: we tailor our training, which is, we have a training program, depending on the audience. If there is someone who is going to be assisting with doing measurements, we focus more on operational aspects, but everyone gets a basic training in the operation of the device, and how it works, and when to use it.

MS. HESSER: Okay. Thank you for your responses.

DR. GRANT: You're welcome.

DR. HARRIS: Okay. So if there are no additional questions from the panel, I will now pronounce the Open Public Hearing to be officially closed. At this point, we will now take a lunch break. Panel members, please do not discuss the meeting topic during lunch amongst yourselves or with anyone attending virtually. We will resume the meeting at 1:00 P.M. Eastern Standard Time. Thank you.

PANEL DELIBERATIONS

DR. HARRIS: It is now 1:00 p.m., and I would like to resume this Panel meeting. We will now begin the Panel Deliberations. Although this portion is open to public observers, public attendees may not participate except at the specific request of the Panel. Additionally, we request that all persons who are asked to speak identify themselves each time. This helps the transcriptionist identify the speakers. During the

next hour, we will open up the floor to questions for the FDA. Is the FDA prepared to respond to panel questions posed today?

DR. ASHER: Hi. Yes.

DR. HARRIS: Great. So, let's open the floor. Do any panel members have a question or comment for the FDA? Perhaps I can -- I don't know if Dr. Skates… Has he joined us yet?

DR SKATES: Yes, I have.

DR. HARRIS: Perhaps you may revisit your earlier comments, because perhaps that would be provocative and useful.

DR. SKATES: Yes. I was actually just writing up a Word slide essentially to give an example from the Melafind study, and I was hoping to actually get the Nevisense results as well, but I think they can wait. In the MelaFind study, the dermatologist had — there were 84 lesions where they thought the melanoma was definitely benign. They were, in fact, melanoma. On the flipside – and those were they thought it was melanoma – sorry, where they thought a biopsy was needed, and there was suspicion of melanoma, there were about 1400 false positives. So, ones which were benign did not have melanoma. That gave in their hands ratio of about 700 biopsies that were melanoma for every one melanoma that they missed. I'm suggesting that is weighing the false negative of missing a melanoma with 700 false positives in doing a biopsy when it wasn't melanoma. In some sense wasn't needed. And if you use that as a weighted benefit risk ratio, then is it possible to share my screen? I do see a share screen. Okay. I will just click on that.

And in the Melafind study, the benefit-risk ratio weighted for dermatologists was true positives to false negatives divided by false positives weighted 700 less than the false negatives gave a benefit-risk ratio of 44. What I'm going to suggest is that is what should drive the benefit-risk ratio for an SLA. That gives the quantification for then that then leads to specificity. And now you can set a specificity to arrive at that benefit risk ratio when using an SLA. So, what I'm trying to suggest here is you don't set specificity in an arbitrary goal like 80% or 90%. What you do is you try to set it so the benefit risk ratio, that's the harm, or that's the relative ratio of benefit to harm… You want to try and obtain the same benefit-risk ratio that dermatologists achieve in their study. And the Nevisense results should also weigh in on this in terms of how many false negatives and how many false positives there were, and that ratio will give us a sense for what dermatologists can achieve. And if we want to apply that to primary care physicians or primary studies, which might have a slightly different study design, then that is how we can incorporate this notion that sensitivity is very important. You want to have this false negative, which is the risk of a false negative be weighed much more highly than the false-positive, which is biopsy when they don't have any melanoma. But missing it if they do have a melanoma is really important. That's reflected in this relative weighting. The overall benefit-risk for the dermatologist is 44, and that will lead in any particular study to a specificity study. I would say if it meets the same level as the dermatologist, then you are obtaining the same benefit-risk, and therefore, it is as safe as what a dermatologist can provide a patient.

I would like to — and then efficacy, or safety and effectiveness… Effectiveness is some measure of sensitivity in what an SLA either achieves or adds to the dermatologist or primary care physician's sensitivity. But my understanding of the FDA's Act is that the Act requires the FDA to assert the device is safe as a first measure -- and this is my measure, I'm going to suggest safety – and it starts with this benefit-risk and it leads to specificity. So, I would like to throw that out there and suggest that that is a principal approach and quantifiable approach for making judgments about safety, whereas specificity doesn't get this ratio of benefit-to-risk explicit. This does. I would like to suggest that be at least a consideration if not a prior step that then leads to specificity in the FDA's deliberations.

DR. HARRIS: Okay. Thank you. Next comment by Dr. Pisarik.

DR. PISARIK: I have a question. It may be kind of obvious, but the purpose of these SOA's is to increase appropriate referrals to dermatologists, decrease wait time, and ultimately decrease mortality of skin cancer. Do we have any evidence that that is happening or that can happen?

DR. HARRIS: Anyone from FDA be able to address that question from their research or presentations?

DR. ASHER: This is Bonita Asher from FDA. We appreciate your comments pertaining to the questions that we have put forward for you. There is a variety of manufacturers that may come to us with any given indication that they wish to market, and it would – the incumbent on us to let them know the level of evidence necessary to justify the proposed indication that they see. So, it could be anything and everything

that you are suggesting there. Really, I think this is an exciting area of innovation, and we're just looking for feedback. If you have specific feedback on a specific indication, that would be helpful. Or if you have just general guiding principles on considerations that we need to keep in mind pertaining to these devices as a whole… that would also be helpful.

DR. HARRIS: Did that answer your question, Dr. Pisarik? Because that wasn't my understanding of your question.

DR. PISARIK: Pretty much. I don't know that there's a lot of data that shows that can happen, but I wanted to see if there's anything out there that said these devices would do any of that. In fact, it can actually increase wait time. A lot of people use the device, and all of a sudden they're concerned; they go to the dermatologist, and the dermatologist, and the dermatologist's wait time could be even higher prior to what it is right now.

DR. HARRIS: Does anyone have any knowledge that can address that issue as to what is the impact of these devices? In terms of mortality rates, morbidity, access, delayed diagnosis? It doesn't appear that anyone has an answer for you...

DR. ROTENBERG: My understanding of what Bonita has said – sorry to interrupt, this is Veronica Rotemberg – is that it is our charge to say, we need studies we need to answer, in order to guide the FDA indications and approvals. And if our goal is decrease mortality, how do we suggest that these companies study — that in terms of their trials rather than, that evidence to my knowledge doesn't exist, but the question for

us is: what evidence would we like to see, keeping in mind these patient outcomes that we care about the potential approval of the devices. But correct me if I'm wrong.

DR. ASHER: That's absolutely correct. Thank you very much.

DR. SKATES: So, Dr. Asher, in my experience with mortality reduction for early detection, I have been involved with early cancer detection, and that required hundreds of thousands of women to be randomized after screening after a 20-year study. The studies that FDA is being proposed here are simple sensitivity/specificity benchmarks that take 1 or 2 years maybe, I don't know, 3 years. If, indeed, mortality reduction is an endpoint the FDA wants to see, that's an order of magnitude greater time and effort and money, and I would say it goes beyond the minimal burdens of criteria that you set out earlier today. So, I think we need to pick our endpoints; that's a combination of clinical utility to the patient, but feasible to the company. There is a balance there and going all the way to the gold standard requires is quite burdensome. I just want to make people aware of that.

DR. HARRIS: Okay. Next question, Dr. Alam for comment.

DR. ALAM: Yeah. Thank you. I just want to make a couple of brief comments to frame my views of some of the issues being raised. I think it's an interesting formulation that Dr. Skates raises in terms of what the benchmark should be. I do think there are a couple of other considerations we have to keep in mind. One is whether such devices are used with a dermatologist or without a dermatologist, because if they are used without a dermatologist, there is really no way to modify the output that's coming out, and that's pretty much what would be the sole decider of further activity. And so maybe

the thresholds for accuracy need to be higher. And I've heard some of the speakers mention that, probably higher in that case

Another thing to consider, and I like Dr. Skates's formulation, is the idea of false negatives and false positives and what that really means to the patient. So, in general, we don't like false positives because they entail a procedure, perhaps, that is not necessary. But I think we have to also remember what the procedure is. It's generally shaved skin biopsy of maybe 3 or 4 millimeters of superficial epidermis and a little part of thick dermis – it's not very different than scratching yourself while unlocking your front heels in 2 or 3 days. It's very different than, for instance, a kidney biopsy or something like that, which would pose substantial risk to the patient, or discomfort, or even anxiety.

I think another consideration there is: we are talking about the benign and malignant as sort of a bimodal distribution. We're looking at how many were false negatives and how many were false positives, but, as has been raised by some other speakers, there's actually a spectrum. We might not be getting a melanoma in some cases, but not all of those, even most of those, are benign. They might be moderately to severely atypical moles. Most dermatologists – if I can make that assumption – most dermatologists believe there's a progression. So, it's not like you have a melanoma or don't have a melanoma, but you could have a lesion that can progress to melanoma. So, there still a benefit in removing a lot of those lesions that are not technically classified as true positives, but really are not an accurate diagnosis because they would've become something. And as we heard before with melanoma, the main way to prevent harm is actually to get them early. Once you have gotten a nodular melanoma,

even if you detect it, there's not much you can do in regard to subsequent mortality. I think those are some considerations, which I think should impact sensitivity and risk, or where my sense would be if a device is being used by a non-dermatologist to make a primary determination. And given the low potential of harm and potential benefits of detecting lesions that are still in the process of going bad, I'd really like to bias it quite substantially – bias is the wrong word. I'd really like to make sure that the device is very, very sensitive or has a very high risk-benefit ratio. Thank you.

DR. SKATES: So, Doctor, can I just suggest that my 700 ratio of cost of a false negative, which is you missed the melanoma, is 700 times as bad as a false positive where you do a biopsy on someone who didn't need it or made some judgment of not needing it. So that's a very high ratio, and that will require a very low false negative rate. And if I can just share my screen one more time. This is from the MelaFind table of how the dermatologist performed. You can see there's only 2 false negatives, 82 true negatives, but 1400 false positives. That ratio 2 to 1400 is where I get that 7 hundredfold ratio. If we can achieve in other people's hands a similar ratio, then I would say that's a reasonable judgment to say that that device is helping the other person or other provider achieve the same level of safety as a dermatologist. So, I'm just hoping that's in the ballpark of what you are saying.

DR. ALAM: Sure. I would agree with that. I'm just saying it might be even higher because that's with the dermatologist, but potentially without a dermatologist. And that's assuming all of those 700 were really wrong. The other 699, whereas, in fact, any of those other 699 might've actually been correct because they could've been severely

atypical or moderate. I think that's a good starting point, but I think if it's a freestanding device or number of lesions, the number could be higher for freestanding device used by a patient. Thank you.

DR. HARRIS: Next question, Dr. Skelsey.

DR. SKELSEY: Thank you all, Dr. Alam and Dr. Skates, for these comments, because they are helping to clarify some things in my mind. One of the things I've been thinking about is if we are going to be also kind of making a decision or making recommendations about the output, in terms of whether it's only binary. Are we only answering the question of something yes or no, should we biopsy or not biopsy? Or these performance thresholds applicable to devices that are going to also give a differential, and maybe with percentages? Because those are very, very different. What I'd also like to know is, as we look at the question of safety, what was the reason CDRH required a post-approval study to determine when they were looking at with MelaFind over 10 years ago? It's important for us because the same thresholds are possibly going to be applied to new devices. So, that would be helpful for me, and perhaps other members of the Committee, to get a better understanding of, what were those safety issues? What were the studies that were followed up, and what was the data afterwards?

DR. ASHER: This is Bonita Asher. I just want to communicate, the team has provided the background information that's less contemporary for the Panel just to be complete. MelaFind was brought to an advisory committee at the time. The panel at that time deliberated over on the pre-market study and made recommendations

regarding the post-market expectations. And so, this Panel here is not intended as to what the safety question was pertaining to the MelaFind device.

DR. SKELSEY: Because we are looking at performance thresholds for all of these devices, and potentially using the same ones and trying to get new data and make new parameters. And whether it's a number needed to biopsy, are we getting the right number? Both Dr. Skates and [ ] brought up sensitivity, and also about balancing these with having a reasonable number. So, were those the issues? And if so, how did you tease these out so that when you have the next iteration of these devices, how did you collect that in particular? Those pieces of information? Because was that useful? Because I think Dr. Skates pointed out maybe it's not straight specificity. Maybe we need to look at something else, like this very large number of lesions that ended up being false positive.

DR. ASHER: This is Bonita Asher again. My best advice to the panel would be to — I think this is a great discussion. It's a complicated topic and there's a variety of indications that any one manufacturer may choose to pursue. Each indication has complexities on where in the treatment care path they are targeting device use, and precisely what they seek to say, in terms of cancer, the value added associated with the device. And so, we look at that in context with a benefit-risk profile. Certainly with cancers and issues where there are good treatment options available, there may be little room for a device to move the needle and change the risk profile for the patient in the scenario we are talking about. So, I would not necessarily say that we are only looking at overall survival, long population studies, but I think it may be helpful in your

discussions to put in everybody's mind the same scenario that you all are thinking about. I'm looking at a scenario, for example, of layperson use of a device that intends to indicate to the user whether or not they should seek immediate medical attention. That might be a very different device than another panel member may be envisioning as they are developing their comments regarding one that, perhaps, is supposed to prevent people from dying from melanoma. So, I know that my response is a little bit off point. It's not directly addressing your question. Again, I will ask the team to obtain the information more specifically around melanoma, but I do believe that the panel has all of the relevant information of available to them to make recommendations around the questions we have posed.

DR. HARRIS: Next question, Dr. Farber.

DR. FARBER: Thanks so much. Neil Farber. So, there's a couple of comments I'd like to make. One is — and thank you, Dr. Skates and Dr. Alam and everybody else for your comments. What I've seen so far is everybody addressing the issue of, is this melanoma or not? We –at least from what I read in terms of the briefing materials – we were supposed to address not only melanoma, but also basal cell cancer and squamous cell cancer. And I think that basically the risk-benefit ratio has to be different for looking at cancers that are not as dangerous to people as is melanoma. And that is so because of two things: one, I don't think anybody would disagree that, in a melanoma, if there is a suspicion of melanoma and an SLA comes up with that determination, no one would argue that there should be less false negatives. Or rather, more false negatives, sorry, than were seen in the studies that were done. And that is

because of the fact that we don't want to risk anything. However, in a situation where you are looking at, is this a basal cell carcinoma, it's perfectly acceptable to have false negatives for the sake of specificity. And the reason I say that is that it's not just about the anxiety about the biopsy, or the use of healthcare for a biopsy, or the trauma of a biopsy, but it's also the anxiety of a patient hearing they have a possibility of having cancer. Perhaps, especially if this is a lay population using it, not knowing that carcinoma is not much of anything in terms of a danger to them. And therefore, until they get to see somebody, having a great deal of anxiety about it. And so, I think it has to be very clear what population we're talking about. Not all oranges or apples. They are both and we have to assess that.

The other thing, I think that we have to assess at some point in time are, what are the psychological risks for the patients when they are having a false positive? And that can be done in lots of different ways. One can do it even without having patients being involved in the study. For example, you can do hypothetical scenarios and assess their likelihood of being concerned about it. Like thinking cancer, and can you die from it, etc. You can also survey individuals who actually did have the experience of going through and having a positive, whether it's a false positive or true positive. And as well as false positives. And that can be done also through various survey instruments. I think those things have to be looked at.

DR. HARRIS: Thank you. Before we go on to our next question or comment, I would just like to ask the panelists to refer if at all possible to the materials that have been presented to you, and how that is relating to or generating your questions or

comments to help review the division kind of categorize the discussion. Our next question or comment from Dr. Rotemberg.

DR. ROTEMBERG: Thank you so much. I'm Veronica Rotemberg. I think a lot of these comments are very relevant, and I'm glad to be going after you, Dr. Farber, because one of the things that's so important here is the nuance of the intended use and setting. When we talk about, in a dermatologist's hands, dermatologists vary their number needed to biopsy between two benigns to one malignant, to 30 benigns for melanoma specifically. It's different for non-melanoma skin cancer. So, a 700 to 1 is probably not a reasonable comparator to a dermatologist and probably places significant undue burden on the healthcare system. However, that's obviously different for the intended use of a layperson or primary care physician. And I think one of the things that I was hoping to talk about when we talk about intended accuracy is the measured improvement over the current standard of care in prospective studies, because that's what's really going to matter. It's not going to be the absolute specificity or sensitivity of something on a static image that, as Dr. Adamson already discussed, is not comparable to an in-person assessment by physician or provider, and it's not going to be comparable to anything except what is happening in the intended use setting.

I would suggest we think about what percentage improvement in sensitivity or specificity. It's hard to measure sensitivity in real-world practice because, of course, we are just looking at what providers have selected for biopsy, but overall, what improvement in specificity we would like, as compared to a potential loss of one or two false negatives? And then I think the discussion of false negatives really lacks nuance,

as well, and that's something we should add as a panel. Because a false negative severe – which has a less than 1% chance of being an MIS on pathologic review when it's resized – is very different from a negative invasive melanoma, and we cannot just consider those to be the same. I think even though it is more work, I would challenge all of us to think about these complex issues in a very nuanced fashion, because it's going to really matter. And I'm really building on everything that everyone has said.

The last thing that I would add to this discussion is that the differential output, as has also been said, binary malignant classification versus multi-class... Those things have been shown in studies with dermatologists and other providers to really impact what the outcome is, so, when a dermatologist is faced with the classifier that says 'benign/malignant' versus multi-class output, even if it's the exact same classifier, their behavior is going to be different, based on what just the interface shows them, and we have some fairly good in silico studies that tell us that. So, we can't adjudicate any specificity or sensitivity material in a vacuum, because we also need to know how the intended user is going to interact with the system.

DR. HARRIS: Any comments from FDA?

DR. CHEN: This is Colin Chen. There is standalone improvements by certain skin lesion imaging devices. Also, there is the concern whether that real-world use really improves performance parameters, or sometimes it could impede the real-world uses. For example, with MelaFind, we had discussion of how sensitive and how specific a MelaFind device was at detecting melanoma. Also, there was a reader study to show how it increased or decreased dermatologists' performance reviews in the

clinical setting. Also, as a condition of approval, there was a post-study that follows up to study the device use in real-world to see how good it is. So, yeah. It is complicated situation. Not just by the device's standalone performance, but also other factors: graphic user interface is a factor, for example. The output format, how it looks like… that could affect a user's performance. That was considered as well. But also, at the same time, we have seen in research that is reporting it could be important to consider all the approaches together, whether there could be any info I know to a study design that would be more efficient is more important. And that's one of the questions we are bringing out at the Panel, to have some future deice design.

DR HARRIS: Thank you, Dr. Chen. To get back to the question pertaining to MelaFind, the purpose of that study was to evaluate whether MelaFind increases the sensitivity of physicians in diagnosing melanoma and high-grade lesions while the false positive rate was not substantially elevated. Essentially, whether the accuracy of the device can be confirmed based on real-world evidence. So, that was the history there. Pertaining to the comments that the Panel is providing, and we sincerely appreciate the considerable thought that is going to this very complex topic, it will be very helpful if you could please stratify your responses as to whether or not you are speaking to malignant melanoma or non-melanoma, as well as whether you are speaking to a device used by laypersons, non-dermatology, healthcare providers, or dermatologists. Perhaps using those categorizations would be helpful in your discussions among yourselves, as well as we go through the Panel transcript and take your comments to heart. Thank you very much.

DR. HARRIS: Next question from Dr. Burke.

DR. BURKE: Thank you. I am Dr. Karen Burke in New York at Mount Sinai Hospital. First of all, I think lots of the previous comments address things I was going to say. I just want to point out several things. First of all, Dr. Skates did a great analysis of the MelaFind. I just want to point out that MelaFind, in doing this study, they had a classification of absolutely certainly benign and absolutely certainly malignant, and within that, they had three categories of 30%, 30%, 30%: minimally suspicious, moderately suspicious, or highly suspicious. I think all of those were part of the false positives. Perhaps it would be interesting, perhaps, to do their data — and I also called to see if we can get the same kind of data on — I talked to the FDA about getting this kind of report for a good comparison. So, just maybe the data could be fine-tuned a bit, and it might not be quite 700 to 1, because minimally suspicious might almost be considered benign. And then I think, also, what Dr. Alam said is just very important: that first of all, these biopsies are very minimal and the main name of the game is not to miss a melanoma. I tell all my residents that everything you suspect is the cancer comes back as the cancer as, you're probably missing something. All of us dermatologists biopsy something that is so called the “ugly duckling,” or the different lesion that really looks like a keratosis, and it comes back as a melanoma or a wart on the foot. I think it's actually good to be more suspicious, and the Nevisense they said they had 7 biopsies to one melanoma. That's extraordinary and I think dermatologists might like that because we shouldn't over-biopsy. The problem with over-biopsies is,

especially in the dysplastic nevus syndrome patient, where everything you would biopsy on a patient of hundreds would come back as dysplastic… but we have to survey.

My other comment is that of the ones that are false-negative, they are not melanoma. But once a patient has hyperplasia or dysplastic nevus, they are already in a different category. And that's like a category. If someone is light skinned, had multiple blistering sunburns… these are all extra factors that must be weighed into a patient. So, if a patient comes with one little abnormal mole that is very dark on a light-skinned person, I think most of us would biopsy. Likewise, if a dark person, a black skinned person comes in with a dark mole that doesn't have any other of the ABCDE criteria, we probably would not biopsy. But in finding a dysplastic nevus, that puts a patient in a category of having to be surveyed maybe at least once a year. Especially if they are continuing to have sun damage.

And the other thing I just want to address is anxiety. That's so different to quantify, and to me, the basic thing that we want is to not miss a melanoma, or even a very dysplastic lesion, or melanoma site 2 that could progress. And everyone anytime has a test, I think there's a degree of anxiety. If you have high lipids, you're worried about your lipid count. If you had prostate prostatic hypertrophy you're worried when you get your blood test. Usually, a biopsy comes back in a week, and if the doctor says, "Well, I'm not sure and I'm getting a consult," then the patient should be extra happy for the extra surveillance. And some people are, as we all know, some patients are anxious despite multiple reassurance, and other patients are little less suspicious even if they have a dysplastic nevus or whatever. Anxiety is certainly important personally for

each patient, but I think that is what the doctor-patient relationship should be, to quell that. Because here we are talking about a machine that measures a quantitative thing, and any anxiety involved with waiting for the results of either this measurement or the biopsy. That's a sort of difficult parameter to include. What we're aiming for is 100% sensitivity, and maybe specificity of 50%. If sort of half of what we biopsy needn't be biopsied, that's not so bad, because certainly a dermatologist with a trained eye does not biopsy things that are obviously benign. But the spectrum between has clinical consequences for surveillance, and that is also important.

DR. HARRIS: Any comments from FDA? Okay. Our next question or comment would be Dr. Bourelly. Before she speaks, everyone just ensure that your hand raise is contemporary and you didn't just forget to put it back; that you have another question. Great. Dr. Bourelly?

DR. BOURELLY: Thanks. In reference to the comment just made, the dermatologist who maybe doesn't have much experience, or maybe the layperson who chooses an SLA and develops anxiety because they think they have something, and they come into the derm and sort of demand a biopsy. If the derm thinks, I don't think it needs be biopsied, or we can monitor, are we now putting an undue burden on the provider, whether it's a dermatologist or someone else who decides to biopsy to either use that same SLA, do I have to purchase one now, or do I have to biopsy because the patient is very anxious, because their version of it told them that this is something that needs to be biopsied. I just wondered if anyone from the FDA has thought about that scenario and how it might impact us in real, practical terms. Thank you.

DR. ASHER: This is Binita Asher. I think it's an excellent issue to flag. How would you recommend that be addressed? Others on the panel. Including Dr. Bourelly.

DR. BOURELLY: Well, really quickly, I would say I would make sure it gets in the right hands. Again, I'm kind of in favor of derms using devices like this. I understand the necessity of maybe branching out, and I'm totally open to that. I do think it would be very important to add the information that this is just adjunct. This is not diagnostic. And I'm not sure the layperson will hear that. I do have great concerns about the layperson sort of feeling overly confident in reducing the provider to a technician. I think I need a biopsy. I'm coming in. I'm going to get a biopsy. If you don't do the biopsy, what sort of tension are you creating between yourself and the patient…?

DR. BALLMAN: So, just a quick comment on that. This is Carla Ballman speaking. I think for the layperson, it should not be a result that is, 'biopsy or not', but should be a result, 'see a specialist' or not, or, 'see a dermatologist or not. Take that 'biopsy' out of there.

DR. FARBER: And this is Neil Farber. I agree, it shouldn't be a 'biopsy,' or not for the laypeople. The other thing is that it requires, unfortunately, the dermatologist to spend some time and use good communication skills to be able to talk with the patient and explain whether it is something that does really need to be biopsied or not. So, there are ways of addressing it, but they are I think outside of the issue of how does the FDA assess the performance of the SLAs.

DR. ASHER: Let me -- so, you know, I think we are converging on it. Typically a manufacturer comes to us seeking to make a particular claim relating to their device.

And so, it's not FDA's role to say this is not an indication that you can never have, but rather to tell them the level of evidence that we would need to justify that indication could be legally marketed. So, I think you are saying perhaps it may not be appropriate for a layperson to receive information on whether or not a biopsy is necessary because certainly that may be difficult to justify. But can you imagine scenarios, for example, where top diagnoses are shared with the layperson saying, you know, this is the likelihood of a malignant melanoma. It's X percent or in the top categories on these are the top diagnoses, and all of those diagnoses are ones where possibly the next step would be to biopsy. You know, can you converge on this a little more deeply, on explaining what would FDA say to manufacturers that are seeking to make claims along these lines, regarding the types of studies they would need to do both in the pre-market and post-market.

DR. ROTEMBERG: Just to directly address this. This is Veronica Rotemberg. We have the tools to answer these questions, and the tools are prospective studies. We don't know what would happen if we get these perspectives to patients. Will they call their dermatologist? Will they call their primary care doctor? What will they say? But we do have the tools to ask manufacturers to study this prospectively and quantify how many additional biopsies of benign lesions are performed? What is the additional burden on dermatologists? I think that would be my opinion about what we have to do, is define what prospective information and the intended use setting we need to feel comfortable saying that these tools could be in a layperson's hand or primary care person's hand. I want to also clarify that we know that if you give percentages versus

just benign malignant, the outcome is different. The way that patients interpret that is different, and that also needs to be studied prospectively. There's too much nuance here to be able to say, "This is what would happen without testing it for sure." And these studies do not need to be big, especially in non-melanoma skin cancer, which is very common. I do not think these prospective studies would be very burdensome, but they would help us understand what the burden would be and what the cost would be in with the potential harms would be.

DR. ALAM: I would agree with that, if I could weigh in. I think it is nuanced, and, Dr. Asher I think maybe what we were saying would be, indirectly in the case of, to specify as per your previous recommendations, of the layperson using a freestanding device. I would say even more specifically primary prevention. This is not a person who has a history of skin cancer; he was just trying to monitor. Someone who doesn't know much about skin cancer, who has this app, they have it at home, they are trying to do something good… in that case, I think that what we are saying is the bar for someone to say 'biopsy or not' or 'melanoma or not' is exceedingly high. And we are not convinced – at least, I'm not –that current technology is up to the task of saying 'biopsy or not,' 'melanoma or not.' If I may use a diversion example, it's like self-driving cars. Sounds great in principle. If you read the predictions, 20 years ago should've already happened, but we are still struggling with that because there's serious points of failure, that, some were anticipated, and then to a point that I think was previously raised by Dr. Rotemberg, there's a lot of known unknowns. And there's so many unknown unknowns that the only device I would feel comfortable saying definitely is

okay… that situation is something that has 100% sensitivity and I'd be happy with it, even if the specificity wasn't great. But, as you know, you can make a device that has 100% sensitivity tomorrow. You just make a device that, whenever you image, it says 'biopsy,' but that's not a very good device either. So, in that case. Now, if it's a case where it's supplementing a dermatologist doing something, that's great. If I's not primary prevention, where it's just someone who has a lot of these already thinking "should I go in earlier because it's changing." highly sophisticated, maybe highly educated, maybe a lower bar. But if it's going to be in the hands of somebody to look at themselves, they don't have much knowledge, then we have to be really careful in terms of what information is provided and, in particular, avoid anxiety, false sense of security. Thank you.

DR. FARBER: And, if I may weigh in, thank you, Dr. Alam. This is Neil Farber.

DR. HARRIS: Excuse me. Can I make a quick comment because I don't want other people who may have other things to say not have a chance to speak. Quick comment and then we're going to move on. Thank you.

DR. FARBER: All I will say is looking at anxiety is not difficult. It's very easy through different types of studies. And in addition, I don't know anybody who can quell a patient's anxiety. It's something that needs to be dealt with.

DR. HARRIS: Great. So, very patiently, Dr. Suarez. Your comment, please.

DR. SUAREZ: Yes, thank you. These have been very appropriate. It seems, being a non- dermatologist, that there's actually two layers, or frameworks, that one is discussing. The actual accuracy of the device, for which I think the highest standards

and the best possible ground truth should be used. And then the whole other layer of performance thresholds and all of that. That's more in the context of how it's used to practice. It's almost there's two different types of studies that need to be done. And those are going to be different according to, who is the user, and what's the definition they are being used. So, on the perspective of what we are being asked to do and the questions the panel needs to address, I was wondering, is that what the FDA has in mind? There's a sort of a two layer process where there's a first layer, where you want to make sure that the device is as accurate as it can be, then a second layer, where new studies to address how these were performing in an actual work setting. For instance, say we are talking about melanoma and the false negative rate if we have primary care physicians that have sensitivity of 50%. So, there's going to be 50% of melanoma missed. Even a sensitivity of 7% would be a vast improvement over what is being the standard of care at this time. So, again, I guess my comment or question to the FDA is, are we thinking in terms of two layers or two stages of studies, where the first one is really about the accuracy of the device, but the second one is the actual performance in real-world settings before a decision is made as to how these devices are approved or permitted to be used.

DR. ASHER: Yeah, I think you raise an excellent point. Certainly we want to characterize the capability of the device, and doing that in clinical trials in the pre-market space is typically how that is performed. Although, more and more we are looking at real-world evidence where it's possible to obtain it to help guide our regulatory decisions. And then if there's questions in the post market, you know, the real-world

experience is very informative. Not only to confirm or better understand what the device is capable of doing, but also as device manufacturers seem to modify the technology and make iterations that are more useful for patients.

Since I'm answering this question, I did want to just circle back. I sincerely appreciated the comments, as I think many of our team members did, about considering the psychological aspects, the anxiety associated with the use of these devices. It cuts both ways, correct? If you use the device and have access to an app, and you may not have the resources to see a dermatologist, it may be that this reduces your anxiety. Alternatively, you may use the skin lesion analyzer; it may heighten your level of concern, causing greater anxiety. So, if you have specific suggestions or ideas on how we might be able to even-handedly assess that and put that into context as we evaluate whether a favorable benefit-risk profile exists, that would be very helpful. Ideally, such a tool can be used not only in the pre-market, but also the post-market if we conduct real-world evidence evaluations it could be nomenclature in our discussions pertaining to these technologies. So, any such advice would be appreciated.

DR. HARRIS: Dr. Skelsey.

DR. SKELSEY: Thank you. It's Morrell Skelsey. In terms of real-world, I think will be useful in terms of the FDA and companies is having an assessment of their reliability and ability to diagnose other skin cancers. They may be completely unable to, but as a dermatologist, we all know that patients feel they have – they report they've had a melanoma and they might have had only a basal cell or squamous cell. Something that's labeled as a diagnostic tool for melanoma has the potential to give people a false

since of security. In addition, there's tumors. Depending upon the patient population they may be more susceptible to squamous cells and the device gives them a false sense of security that they actually don't have a cancer. They don't have a melanoma. I think it would be important in terms of discussions with the companies to give some idea in terms of other common tumors that they are able to, perhaps, assess, but also the uncommon pigmented lesions of the darkly pigmented -- for instance. Where does that technology lie -- to make a recommendation at all regarding other tumors?

DR. HARRIS: Any comments from FDA?

DR. ASHER: We appreciate that comment. I think you make a great point, and we will continue to keep that in mind.

DR. HARRIS: Next comment or question from Dr. Rotemberg.

DR. ROTEMBERG: I really liked Dr. Skelsey's comment because it goes to one of the things that I also wanted to talk about now, which is what was brought up by Dr. Adamson and Dr. Cohen around labeling. It does make sense for devices that are marketed to laypeople or PCP's to not be as narrow as some of the devices that are marketed to dermatologists, because it may not be possible for a layperson, as you pointed out, to not know the difference between a pigmented basal cell or something that is melanocytic, for example.

Another thing that I think is going to be important is how these devices communicate to us. This is not in scope for this particular device. For a dermatologist or provider, it may be sufficient to have that be in the label for a user of an app. It might need to be more obvious, more automated, and more significant to say, you know, this

looks like an acral lesion, and we are not trained to identify acral lesions. This looks like a patient with skin type IV or V that we have not had the training to analyze. It may need to be more automatically communicated, this type of labeling, then we might expect in our office or office tools. We do a lot of procedures that we already know the expected population and the way that an app probably shouldn't. And so, I did want to make sure that we talked about that and talked about labeling for specific indications and to reflect the data that was used for training and validation.

The other thing that you brought up, Dr. Asher, that I think is really important, is the post-marketing surveillance. One thing we know about AI tools is that they are susceptible to data drift in a way that sometimes lab tests may not be. So, as for example, cameras are updated; those cameras may not have been used to take photos that are in the training and develop phase. How are we going to learn that the accuracy of these algorithms have decreased over time due to drift? And I do want to suggest that we think about a way to do post-market surveillance that continues to assess models for accuracy and benefit the way that we suggest, over time.

DR. HARRIS: Any comments from FDA?

DR. ASHER: No. Thank you very much.

DR. HARRIS: Next, Dr. Bush.

DR. BUSH: Laura Bush here. My comments have already been touched on by Dr. Rotemberg. Dr. Asher, I was going to ask you I see patients for days a week or every week I think the devil is in the details on the labeling because if you're going to -- I'm just assuming let's talk about layperson having, say, an app. They come in and it

says I do agree it should not probably say biopsy or not. This is a concerning lesion, seek care with a dermatologist or however you want to word that. But I think if someone comes in and they have that information in their hand, then that is a whole other conversation to say they may not need a biopsy or they may doctor shop until they get a biopsy, so that may create a whole other can of worms, so to speak. I think you have to look at this from who you are marketing which device to and if you are taking SLA's or the whole group, are there SLA's that are laypeople, SLA's and for non-dermatology people and dermatology people. Because very often someone will come and even from referral from primary care they are just convinced they have melanoma and the lesion is a sub K, but they may have a melanoma on their back, so it is somewhat of an opportunity. I just think you have to be very, very careful on your labeling just not as to hone that person into being… their hands tied on doing a biopsy. Or the person may doctor shop.

DR. ASHER: One of the most avoidant things that FDA does is help characterize risk, cause companies to help characterize risk, and communication. I think what you're getting at is, how do we know that those risk communication efforts are effective or not? So, if this group has recommendations along those lines, about how in the pre-market we may anticipate places where risk communication may be an issue, and embed either in the pre-market or post-market evaluations whether our communication tools are effective and are causing the appropriate actions by both providers and patients. That would be immensely helpful. This is one of our biggest challenges. In a complicated area like this, we would love to hear your thoughts.

DR. HARRIS: Next comment is from Dr. Skates.

DR. BUSH: Can I respond to what Dr. Asher said real quick? Am I good? What about if you had not only information on whatever device of what this possibly… sort of an algorithm, or information sheet that spit out that they can actually take to their doctor, or some sort of informative device that would hone them toward that, so that to guide them in the correct decision, and maybe have an algorithm that was set so that they followed the correct path?

DR. HARRIS: Okay. Dr. Skates.

DR. SKATES: So, I'd like to respond to some of the questions that FDA proposed to us, which was one of them was should we effectiveness —

DR. HARRIS: Brief interruption. So, I think we are going to be talking about the questions next.

DR. SKATES: Okay. All right.

DR. HARRIS: So, this is really our time. The next session we really won't have an interaction with the FDA. We want their feedback. It will just be a discussion amongst ourselves regarding the 3 questions. So, this is a great time to get any feedback or responses from FDA.

DR. SKATES: I see. Well, I guess I was hoping to hear from Dr. Kaminski about the risk-benefit ratio. I want to share, again, the screen, because the Nevisense link was sent by Ms. Knowles, I think, to everyone. I've been trying to go through that while having this discussion as well. I will just share Table 34 with everyone. And again, I've highlighted the false negatives. There were 9 false negatives out of 265 positive. And I

believe this is – unfortunately I haven't had time to go through it – but I believe this is melanoma. This gives you a sensitivity of over 96.5%, but it's not 100%. Dr. Burke said we need to aim for 100%. With that rule, is that not sufficient? My guess is that Nevisense is sufficient with 96.5% sensitivity. And then the cost ratio between the false negatives… These are 9 false negatives, and I think these are melanoma. Melanoma is missed by Nevisense to 956 false positives. These are negative for melanoma, but Nevisense said they were positive. So, that's a ratio of 1 to 100. Instead of this MelaFind with 700 fold, this is a 100 fold.

You know, the con, the way that the FDA had framed the questions was trying to cut them up into quantitative benchmarks. I think throwing out numbers in the hope that people start to either say, “it's too high, too low,” and come up with their own numbers, but I haven't heard any numbers from anyone else as to: for each false-negative, for each melanoma that you miss; how many biopsies would you be willing to do? Here, the Nevisense is 100 biopsies and they miss one melanoma. In the MelaFind it was one melanoma. This gets difficult because you are trying to balance patient anxiety here, going to the doctor and resources with missing melanoma. It's a subjective judgment that ultimately there needs to be some judgment quantitatively for what that ratio is to then do a benefit risk analysis and come up with: in dermatologist's hands, this is what happens, and if we do this for primary care, can we, with the addition of SLA, get them to that same level of one in 100 false positives to one false negative? I would like to hear from other panel members what their judgment is on that ratio of very bad false information from the device, to not so bad, because it's a false positive and

they go into biopsy when they didn't need it. And, you know, saying 100% sensitivity, I think there's no device in the world or no doctor in the world that's going to give answers correct all the time. So, I think that's making the perfect the enemy of the good, and what I'm trying to do is get a good value judgment, a quantitative judgment, on the good here. So, that's it.

MS. HESSER: Dr. Rotemberg had made that some of these perspectives patients say formative to us there may be a skeleton of patient preference… [Indiscernible]

DR. HARRIS: Unfortunately, Ms. Hesser, we are having a little difficulty I think with your Internet connection. You are breaking up. Okay. We'll need you to log in and log back on. Dr. Block?

MS. BLOCK: Actually, it's Renata Block and I'm a dermatology physician assistant. I'm very honored to be amongst all of you today in this Panel. I just wanted to — I've been digesting that everything everyone has been saying and looking at this from 14,000 feet and trying not to be in the weeds, as I always go into detail, this is really a discussion of future devices and, as a dermatology physician assistant, I work closely with my collaborating dermatologists. I feel that this isn't really — the SLA's are not fit for laypeople at all at this time. Like Dr. Alam said, self-driving vehicles is a great concept, and I think we will eventually be able to get there, but I think it would really do more harm than good at the end.

I'm also concerned about — Dr. Bourelly made a good point. Not only will the patient be coming and demanding biopsies or demanding other things, but it also poses

a liability for the practice in regards to these results. So, I think the focus is to really have the FDA have strict regulatory guidelines for these devices. Unfortunately, apps are going to happen. Google is going to happen. The patients are going to be utilizing those tools. I wish we could have more guidelines with those as well, but I think they already are happening, and the patients are going to rely on them. When it comes to things like Nevisense and the other one, these are things that are very, very expensive. I believe they are $10,000. I can't see many practices investing in these, at least now. I think down the line, once more technology is available and they are more streamlined or maybe more affordable… and then we can re-assess whether they are able to be utilized by PCP's and non-dermatologists. Just wanted to add that in there.

DR. HARRIS: Thank you. I think we have Ms. Hesser back. Could you make your comments again, please?

MS. HESSER: Yes, thank you. I apologize for that. I was following up on Dr. Rotemberg's comment on being able to capture some perspective patient information in advance. The FDA does have a number of patient preference information tools already developed. There is nothing that I found specific to skin cancer, but perhaps this is an opportunity to be able to develop something that could inform us ahead of time as to some of the psychological impact, the effectiveness of our communication tools, of risk tolerance. Just wanted to offer that up having gone through some of those tools already. Thank you.

DR. HARRIS: Thank you. Dr. Roth?

DR. ROTH: I just want to mention economic disparities is another kind of disparity we have not discussed, and there are a lot of people who cannot afford to go to a dermatologist if the dermatologists do not participate in insurance. It's sometimes very difficult for people who rely on their insurance plans for a referral to access a dermatologist. I would like — I understand the concerns about patient access, although the direction is in favor of greater patient access, and most patients will be able to read the doctor's notes after the doctor has written them on current electronic records. But I have several friends and patients who are at high risk for melanoma, either because of the close family member having had melanoma and/or having had a personal history of melanoma, and I think everyone I know is Fitzpatrick 1 or 2. I think there might be a role for patients training on a device, or select groups of patients training. I understand that it's very difficult to diagnose, even for professional dermatologists. But there are a lot of barriers right now, and I don't see that getting better for some people to access dermatologic care and a dermatologic assessment. And there are rural areas where dermatologists may simply not be easily available. So I think we should just consider that as well.

DR. HARRIS: Any comments from FDA? In that case, we will go to Dr. Ballman.

DR. BALLMAN: I'm sorry. I thank you. I have been waffling putting my hand up and hand down and not sure if my comments are meant for now or for the questions later. All I want to say I – and especially for the layperson thing – I think for sure has been mentioned, there needs to be prospective trials. I think it has to be shown that there is an increased benefit-risk over what is currently available. Whatever and

however that gets measured – and there's been different measurements sort of tossed out there – and it has to be above, and it has to be a randomized trial, I think. What makes it really difficult in the layperson thing is, people that are going to participate in those types of trials, we know for sure, are going to be very different from people who are going to use these things in the real world once it gets released. So, I don't know what sort of real-world studies the FDA has done in such situations, but I think that's going to be very difficult. I think that whoever is going for the indication has to have — it has to be clear to them that they need to do a really rigorous and detailed sort real-world thing once it lands into the hands of the people. I think, in some respect, that's going to be a burden and just high for these people, to the sponsors, to meet. But I think it is necessary because once it gets out to everyone, they are going to use it differently. There was some discussion about the technical sort of performance. I don't know if it should be a two-step thing. Obviously the manufacturer has to have good technical performance in the hands of experienced users, but this isn't going to be used by experienced users. I think it should just go into the population, the clinical trial, and whoever the intended is, for the laypeople. And then the real world is going to be crucial in that group, because I think it's going to be — the results from the clinical trial are going to differ much more in a layperson population than it would if dermatologists are using it in a clinical trial, versus dermatologists using it after the clinical trial.

DR. HARRIS: Thank you. Dr. McGrath.

DR. McGRATH: Thank you. I have two specific questions for the FDA. We've been talking all this time pretty much in the abstract, generally about future devices. My

question is: how close is this to actually being actualized? Are any manufacturers currently actually asking you for review of devices that would be marketed to patients? And what sort of thing are they asking? And number two, second question. Has anyone at the FDA talked to primary care physicians about this? We saw the numbers were all over the place in terms of their diagnostic accuracy on the earlier tables, but I would be curious since most patients, as Dr. Roth pointed out, go to their primary care physician with their skin lesion. What would that community, what is that community telling you about this type of a device?

DR. ASHER: This is Bonita Asher. To give you an indication, you've heard many of these are marketed outside U.S., so I think it's fair to say there's interest in marketing inside the US as well. But we at FDA want to do our due diligence in making sure we are asking the right questions, getting the right answers, and being thoughtful in the level of evidence pertaining to these technologies. In talking to specific groups, it is incumbent on a manufacturer to create a picture of how we might be able to justify a favorable benefit-risk profile. Many of our manufacturers for any sort of device may conduct focus groups or check with experts in the field to better understand how the device might be used in their hands. What they see is, is in this niche market, a favorable benefit risk profile in order to allow the manufacturer to identify an indication that makes sense for them from a business perspective, and also design a study that will address the needs from lead investigators and other thought leaders that are familiar with the space. I'm unable to talk to you about specific interactions that FDA has been having with manufacturers, but I can tell you that this is a very timely

discussion, and the questions that we are posing to you are the questions that we are encountering or that we have been grappling with. Any detailed advice on how we may address these questions, in turn, with the manufacturers would be sincerely appreciated.

DR. HARRIS: Okay.

DR. ASHER: Let me turn to Dr. Chen. Is there anything additional on that topic you wish you communicate?

DR. CHEN: No, you covered it. Thank you.

DR. HARRIS: Next comment, Dr. Alam.

DR. ALAM: Thank you. Trying to be brief just a couple of thoughts on what others have said already. Again, I want to focus on the case of laypeople using this device for primary prevention. To the questions that were raised about what specific numbers for sensitivity by Dr. Skates, I will just throw some out. These aren't just my numbers. There has been a small study done that's currently in submission for publication looking at – and I know this because I was one of the investigators – looking at family doctors and primary care physicians, dermatologists, and oncologists or non-dermatologists, and asking them what levels of sensitivity and specificity they might want in such a situation. They were given a lot of background data, such as, what are the levels of tentative video, and specificity in the hands of dermatologists with or without the currently approved devices, which as we know are around 90% or so. Basically, they wanted more, which is what we were saying. And the numbers were 95% or greater across the board. I don't want to get into the minutia subgroup, but

that's kind of where it was. I can't think into the black box of why people said that; I think the expectation is that a free-standing device should be better than a device that is supplementing a dermatologists' inherent judgment. I think that's a number to consider.

Another thing to consider, I think you said, is how many things should be biopsied. That turned out not… 100 or 700 – or I don't really know if that's the data we have, somewhere in that range – I think would be fine.

I think one other consideration that we haven't discussed, which I think is important for lay devices, is that we are talking about looking at specific lesions where the patient is going to be – with their software or camera, maybe with their phone – and they will be imaging potentially in some cases a specific lesion of concern. And then the device will spit out something out, something whether it's benign or malignant, or goes through a dermatologist. But there is evidence in the literature; specifically, a paper from a couple of years ago showing that when patients go to see dermatologists concerned about a certain device, more often than not there's a concerning lesion on the patient, but it's not the one they thought was concerning. And, in fact, when you do a complete skin exam, in the same study, they showed you often find many additional concerning lesions that the patient was not aware of, which were not really on their radar. I think there is going to be a lay device, they'll either have to have a disclaimer, or better yet, some way of scanning much of their body or many of these lesions. Because otherwise, sitting at home, they might scan something that is not concerning and then be very confident that they don't have skin cancer, because they didn't scan what was really problematic. So, I think that is another consideration.

And I think the last thing I think with regard to anxiety, and this is obviously probably beyond the purview of our discussion today, but I think it's a legitimate concern that if such devices proliferate, people will want to go for management of something that's been highlighted as suspicious, and that might cause a backlog in dermatologists offices. I know this has sort of not been done historically, but it might be worthwhile to ask manufacturers or software developers that are aiming for FDA approval to come up with a plan as to how they're going to facilitate that process and do they have a group of dermatologists or are they affiliated with some entity that can facilitate referrals in a timely manner. For the non-dermatologists on the panel, I also want to make a point, which is that usually it's important to be screened to detect these lesions, but there's seldom deadly urgency. So, it's not a matter of having to get in in two or 3 days. If you got it in a month or two, nothing much would be lost. The problems arise when you get it in a year or yeo. Just to get a sense of what the threshold should be. Thank you so much.

DR. HARRIS: Next, Dr. Burke.

DR. BURKE: I just wanted to again address some of the points that Dr. Skates brought up, that there is this 700 to 1, or 100 to 1, ratio of false positives. But remember, if we just use the synonym of false, positive suspicious lesion. So, and within suspicious lesion there is a great degree of suspicion, and perhaps, in the data that both of the devices used, they included many lesions of very low suspicion that probably no dermatologist would biopsy because it's so close to benign. So, we don't — I mean, it's just difficult with the statistics to judge. And then also, lesions that were

maybe in the top two-thirds, and certainly the top one-third of suspicious… Remember, some of them do show that they are, in fact, dysplastic nevi, and that is a diagnosis that carries a clinical weight for prognosis of a patient. So, it's just difficult to judge, and I really appreciate Dr. Skates doing the analysis so we could more clearly understand. The other thing I want to point out, the devices we are evaluating are very far from being for the layman. The MelaFind cost on the order of $59,000 and Nevisense costs I think on the order of $5000-$10,000. I just don't think there are very many primary care physicians that even might want this in their office, and certainly pharmacies might have people that can measure people, check people's blood pressure for their percent of oxygen, which are also at-home devices that are used commonly. But I think we are very far from the kind of home device, except in this evaluation, we know that other home devices are coming, and that Google has parameters that doctors totally don't agree with. There's this huge spectrum, but I don't think these particular devices that we are evaluating today would practically, in a real-world situation, would be available to a lot of laypeople, let alone if we are also talking about economic disparity and rural areas – I don't think in a rural area there's going to be a center that will purchase either one of these devices.

DR. HARRIS: Thank you. Dr. Rotemberg.

DR. ROTEMBERG: Just a few very efficiently-worded comments. First of all, for melanoma detection, I agree. The standard for detection should be high, but I would challenge us to be ambitious. I think the standards for specificity should be high. And I don't think that exactly the way that this ratio is presented encompasses an ideal

scenario of triage for a layperson. 90 to 95% sensitivity is reasonable, but we should probably have a specificity that's also fairly high. Again, it should be tested in a real-world setting in a prospective trial.

In terms of the numbers that we've been talking about, 100 to 1, 700 to 1 in the hands of dermatologists were primary care physician, I again do not think these are reasonable to evaluate in a vacuum. If you are improving the detection of melanoma in a primary care provider and you are measuring that in a prospective real-world setting, that number, practically, might not matter. We might just be able to say a 10% improvement in sensitivity and a 10% improvement in specificity... when the baseline comparator is the standard of care is a better way to judge this than the 100 to 1 or 700 to 1. I also think in that study that you showed, Dr. Skates, the overall biopsy ratio was something like 1400 negative 300 positive melanoma. So, that's closer to 4 to 1, which is what dermatologists were doing in that study. Again, these numbers are not taking that full picture of how to improve what the patient is currently experiencing. And that's what I think we should truly measure.

DR. SKATES: And all I'm asking is for people to put numbers on that and not expect that from people. It's a hard thing to do, but I think it's a better scale to do it on the specificity scale. I think the specificity scale is immediate result, and all we should start with is some ratio, or some judgment about how bad these risks are to each other. So, missing a melanoma or doing a biopsy that's not needed, or some other judgment about what's bad here about outcome. That's the risks and we need to combine them in a weighted way, because one is much worse than the other. And we need to compare

that with benefits of a true positive and true negative. And my reason for quantifying that is that gives us a bar for primary care physicians to try and meet with the aid of SLA and get them somewhere in the realm of a dermatologist.

DR. ROTEMBERG: Right, and a dermatologist is somewhere between 2 and 30 benign melanocytic lesions to one, not 700.

DR. SKATES: To one false negative?

DR. ROTEMBERG: You are right, to melanoma. We don't have the false negative rate; we don’t have any way to measure. That's one of the challenges we have.

DR. SKATES: The 2 studies did give false negative rates.

DR. ROTEMBERG: Of the devices, but that is not a dermatology exam. So, we don't know what is missed by the dermatologists, because those lesions may not have been photographed.

DR. SKATES: Well, okay. Of the ones that were photographed, I guess —

DR. ROTEMBERG: Right, but that's already a selected population.

DR. SKATES: Right, so, in a prospective study you’re going to have to select a population of nevi, and then you're going to have to draw a line. I’m just trying to come up with a rational way of trying to draw that line, rather than just saying 90% specificity or 80% specificity. Those are very arbitrary numbers that don't really make a good judgment as to how safe the device is. I think comparing it to how dermatologists, or at least dermatologists in the studies that are prospective studies that we publish, that's one bar, and I think it’s a reasonable bar. Thanks.

DR. ROTEMBERG: I think that makes sense. I think we could should take into account the current standard of care in the improvement we can make over what's currently happening, in addition to those numbers where you are creating the weights.

DR. BURKE: I keep wanting to make the very important point of the false positive. Some of them are diagnostically important, like dysplastic nevus and severe dysplasia. If something comes back as a dysplastic nevus with very abnormal cells, most of us re-excise it, but not always. I just want to point out that the false positives of the essay, 100 to 1 or 30 to 1, some of them give us information that is of clinical and prognostic importance.

DR. SKATES: And so, I wouldn't call those false positives. I will call those true positives. I would ask what your definition of a false positive is and see what the ratio of false positives to full negatives in the study is based on your definition. And we need to get —

DR. BURKE: as I understood it, the only positives are malignant melanoma or melanoma in situ. Not the dysplastic nevus.

DR. SKATES: Yeah, I'm using that as an example. If your definition of a false positive is different from theirs, then that's fine. I'm suggesting that this is a way to try and say safety levels a dermatologists achieves and that would be a reasonable bar for PCP with an SLA to achieve.

DR. HARRIS: Okay. I'm going to have 2 more comments by Dr. Suarez and Dr. Skelsey and then I think we can move into the discussion of actual questions, which I think is overlapping with the discussion we are having now.

DR. SUAREZ: Thank you very much. My question was also related to the false positive rate that is seen in clinical practice. I guess the 1 to 4 rate that Dr. Rotemberg mentioned is what is seen in the real world as practiced by dermatologists. Perhaps someone could tell me also how that does translate to the false positive rate. I'm just curious, because it seems the specificity that's being mentioned by the FDA of 80% — I mean, it's much higher than what's actually seen with general practice anyway. We have to pull the numbers together to get the 1 in 4 for the false positive rate, but you have those numbers. What's a false positive rate of biopsies in general practice as we see it now?

DR. HARRIS: Can anyone respond to that question?

DR. ASHER: I'm sorry. This is Binita Asher. You are asking — can you repeat your question? Is it pertaining to…?

DR. SUAREZ: In current standard of care, the biopsies that are being sent – let's talk about melanoma to make it less confusing. With a potential diagnosis of melanoma, what is the false positive rate?

DR. ROTEMBERG: I can briefly answer that just based off the number of biopsy that we have. It's not perfect, as we already discussed already with Dr. Skates; it's very difficult to know what the sensitivity of a dermatologist is, because we only know what they selected for biopsy. But approximately somewhere between 2 benign to 30 benign lesions are biopsied for every malignant melanoma, and this is only to rule out melanoma. And then for lesions for non-melanoma skin cancer, it's much more difficult to analyze that data, but dermatologists are, in practice, probably significantly more

specific, and the number is probably closer to 1 in 2 to 8 in 10. That's more approximate. But the non-melanoma skin cancer skills of a practicing dermatologists are very high. And it reflects exactly this cost-benefit analysis that dermatologists are making in practice that Dr. Skates is pointing out to. We are much more willing to biopsy a lesion that is suspicious for melanoma than non-melanoma skin cancer.

DR. LEE: This is Henry Lee. Also to answer your question, not from a number need to biopsy perspective, but from sensitivity and specificity… With regards to my presentation on the Cochrane systematic review, they found that providers with experience with thermoscopy have a sensitivity of 92% and specificity of 95% for melanoma. That specificity of 95% would correlate with 5% false negative rate – positive rate, excuse me.

DR. HARRIS: Okay. Dr. Skelsey.

DR. SKELSEY: Thank you, it's Maral Skelsey. One discussion I want to make with companies is inquiring whether or not the technology can assess evolution of a lesion. I think it's important to recognize we know from the literature there's a body of melanoma that is featureless and just having a discussion with these companies about whether or not some lesion can be monitored I think is important because it would increase potentially both the utility and safety of that particular device. The other issue, unrelated, is to making sure the data is representative of a robust set of intermediate lesions. Looking at some of the prior data, they are very heavy on definitely malignant lesions and unequivocally benign lesion. Having a robust selection of that intermediate morphology I think is critical for a useful technology. Thank you.

FDA QUESTIONS

DR. HARRIS: Okay. Excellent discussion. A lot of nuance and complexity here, especially for a non-dermatologist. We are going to move on now, if you will, to the questions session. We are going to focus on the FDA questions, and, Panel members, copies of the questions are in your Panel packs. I asked that each of you identify yourselves each time you speak to facilitate transcription. Can we please show the first question and have it read by FDA?

MR. ANDRIANI: Good afternoon. My name is Rudy Andriani. I'm a mechanical engineer and Lead Reviewer at the Office and General and Plastic Surgery devices on the Cancer Diagnostics and Treatment Team. The agency has 3 questions to the panel covering potential metrics for group truth, user accuracy, and generalizability to the full U.S. population. For the purpose of this meeting, we define ground truth as the gold standard that will be used to determine the diagnosis of a lesion, and re-define accuracy as the measured sensitivity and specificity of a device compared to ground truth.

In clinical trials for diagnostic devices, accuracy is assessed by comparing the device output to the ground truth. For skin lesions, particularly when reeling out malignancy, clinical practice has traditionally relied on histology for ground truth.

FDA has requested that histological diagnosis (core specimen processing with a consensus diagnosis from an expert dermatopathologist panel) be used for ground truth because it provides the greatest certainty in diagnosis.

Device developers, however, cite concerns, both practical and ethical, in requiring biopsy of all lesions, particularly those that appear benign. They have proposed alternate means of defining ground truth, including consensus opinion of experts (of visual or dermoscopic examination of the lesion(s)), opinion of one expert visual or dermoscopic examination), or other methods.

One, should histological diagnosis be required for obtaining ground truth diagnoses and all lesions of skin analyzer clinical trials? Two, are there scenarios for which alternate means or a combination (for example, histopathology for suspected malignant lesions and consensus opinion of experts for suspected benign lesions) of ground truth that would be acceptable?

DR. HARRIS: Okay. So, I will open the floor for discussion as we try and formulate an answer for the FDA to ask question. Dr. Rotemberg.

DR. ROTEMBERG: Thank you. I would say I agree with histopathologic diagnosis of malignant lesions. I think there's a big risk for over-diagnosis and other issues if we do not use pathology for malignancy. However, I agree, and I think this point was made in the earlier session, that, in order to have devices that perform well, we need a large amount of benign lesions for training and validation. And so, there are — I would favor a hybrid approach. For melanocytic lesions, habitually, we have done 6 to 12 months of lesion monitoring with no change is a potential gold standard, in addition to three-expert consensus. It would be important for trials to a priori decide what to do for lesions where there is not consensus, and there I would also consider histopathology.

DR. HARRIS: Okay. Dr. Alam.

DR. ALAM: I would agree that histopathology is necessary — sorry, Dr. Alam. I would agree it's necessary in a study format. I would go a little further – and I understand, historically, studies have just tracked benign lesions to make sure they are not malignant – but my strong preference would be for those to be biopsy as well. As we've heard before, there are melanomas that are clearly obvious and there are ones that are clearly benign. The problem, at least in the context of a dermatologist-assisted diagnosis, is to detect the ones that are not obvious. So really, the biggest threat we have is the benign-appearing lesions, some of which are actually malignant. Now, if we look at the benign-appearing lesions and we say, “Yep, they look benign to all of us,” and they still looked benign to us 3 or 6 months later, I think that's sort of us proving that we agree with ourselves. It doesn't prove much else unless it's a nodular melanoma, and it's not – God forbid – going to kill anyone in 3 to 6 months. It might not do much else in 3 to 6 months, but it still could still be meeting criteria for being malignant. So, my preference would be to biopsy those as well, or if it's not feasible to biopsy every single lesion designated as benign, at least biopsy a certain proportion of them. If it's a large enough sample, that would give you an indication as to how many you might be missing by not biopsying them. If you think about it, the dermatologist is the gold standard, then by definition, every dermatologist is perfectly accurate. Then, you are comparing apples to oranges, because you are comparing dermatologists to histopathology, and you're comparing this app to dermatologist. I don't think this is a reasonable bar. I think histopathology is a well-defined standard, and I think deviating

from that is a terrible idea. I think that's a standard of care issue and, quite frankly, is well beyond the purview of this Panel, or for that matter, FDA, to weigh in on the standard of care. That is practice of medicine issue determined by practitioners and regulated by States. That's just the way that is, and until that changes, that's the gold standard we have to work with. Thank you.

DR. HARRIS: Okay. Dr. Farber.

DR. FARBER: Thank you very much. Neil Farber. I would separate this out. It's clear that any kind of melanocytic lesion has to be biopsied because, not only in terms of the idea that you want to have enough biopsies to be able to standardize the SLA, but also because of the fact that I think it wouldn't be ethical for a patient to be watched for a period of time when there was a possibility of a melanoma being present. However, on the other side, if we are talking about lesions that are suspected of being either BCC or SCC, I have concerns about those patients being biopsied, simply because of the fact they are in this study. Certainly, there would be informed consent, but many of the patients might not opt, then, to be in the study because of that, and/or they would still be in the study but be biopsied perhaps just because they are in the study. With those, I think perhaps some portion of them – and I would leave to the dermatologist to decide who might need to be biopsied – but the others could be watched, or an expert panel deciding what to do with those patients.

DR. HARRIS: Okay. Next, Dr. Ballman.

DR. BALLMAN: This is Carla Ballman. Going along the lines of what was indicated before, I really think it depends upon what the purpose is for what they're

going for indication. If we take just the melanoma case and the study is intended for lesion suspicious for melanoma to determine whether to biopsy or not, I think it's essential that all lesions be biopsied or go through histopathology. I think it needs to be driven as to what the intended indication ultimately is. And we are talking about the pivotal clinical trials, not the training and test sets that went beforehand. That's all separate; I'm talking about the pivotal clinical trial and for lesions suspicious. For melanoma, I think they do need to be biopsied and undergo histopathology.

DR. HARRIS: Thank you. Dr. Bourelly.

DR. BOURELLY: Thank you. Paula Bourelly, private practice. Really quickly, I really like what Dr. Alam said: you really are comparing yourself in January to yourself in May in terms of your criteria for biopsy. I would feel more comfortable if you decide to monitor – if you had a separate dermatologist check it in four months, but that's beside the point. I'm also in favor of histopathology. This is my point from earlier in the morning when I was talking about including brown skinned people is not to look for the needle in the haystack melanoma. It's to make your algorithm smarter. I think when you are biopsy benign lesions you are making your algorithm smarter. I know that's bleeding into another question, but I vote for histopath, even in the things that look very, very benign. I assume we are not going to be biopsying a bunch of seborrheic keratosis; that would probably be the only exception I would say to having a panel of derms say, this is clearly subderm that does not need to be biopsied. Thank you.

DR. HARRIS: Dr. Skates.

DR. SKATES: Steven Skates. With histopathology, you've got a gold standard accepted by everyone. When you implement some undergoing histopathology and some ongoing consensus, how you divide that into is going to be likely variable from one study to the next and it's going to be rather difficult to compare between studies or between devices under that sort of study design. So, I vote strongly in favor of doing histopathology on everyone.

DR. HARRIS: Okay. Dr. Rotemberg.

DR. ROTEMBERG: Thanks, everyone, for their comments. I think I should clarify that I agree anything that is suspicious in any way clinically should, of course, be biopsied. I think what I would argue in terms of the nuance of what Dr. Ballman was saying, is the intended use setting is very important. If there are truly benign lesions – seborrheic keratosis is a perfect example, Dr. Bourelly – we are talking about reviewing a gold standard for all sorts of use cases. The use case of the layperson with lots of seborrheic keratosis, I really think should be considered separately, and I do not think those lesions need histopathologic review if there is monitoring without change, or with only benign change, and a consensus panel of dermatologists that agree the lesion is totally benign and does not need to be biopsied. I would consider that to be sufficient for anything where there is any kind of question about whether it's cancer, or specifically and especially melanoma. Of course, that needs to be considered differently. But realistically, especially for very easily benign cases, I think we need to consider as a Panel an alternative to histopathology, especially because we want to have a lot of those lesions in a clinical trial to validate the specificity of the algorithm. So, it further

increases the amount of biopsies and benign lesions we would be suggesting to do, if we keep that criteria fixed.

DR. HARRIS: Dr. Bush.

DR. BUSH: Thank you, Dr. Rotemberg. That's kind of what I was thinking. If it's something that you are on the fence for melanoma, and you are truly considering that diagnosis, I'm full in favor of histopathological diagnosis. But I take into account what you said regarding extremely benign lesions. If the true intended use for this is, we are thinking this is a melanoma, then I'm all in favor of histopathology.

DR. HARRIS: Dr. McGrath.

DR. MCGRATH: I also favor a hybrid model with histopathology, but also, with 2 codicils to this that would be very specific. One would be if the lesion appeared to be benign, there should be a clear delineation of who decides that. Is the investigator dermatologist or does it necessarily in the clinical trial have to be a panel or whatever, but that would have to be really carefully spelled out about these lesions that appear benign. And secondly, to add for those that are not biopsied and appear to be benign, that there be some defined follow-up required as part of the clinical trial.

DR. HARRIS: Okay. Dr. Skelsey.

DR. SKELSEY: Thank you. Maral Skelsey. As said, it's a minimally invasive procedure. I think it's important for us not simply to have consensus amongst ourselves within the field of dermatology. We are not infallible, and I don't think there's any point in putting… the data coming out is only going to be as good as the data that is put in. As Dr. Bourelly said, we would like to make these algorithms better. So I'm strongly in

favor of the initial studies obtaining a biopsy of even frankly benign lesions because, as we all know clinical experience there are times when a lesion is biopsied and there's very low suspicion. It's just removed, for instance, and I think for purposes of the study is important to get histopathology, and all that remains the standard.

DR. HARRIS: Going to disrupt the sequence for a minute and ask, I believe it was you, Dr. Farber, who were voicing concerns about the biopsy of benign lesions. Could you talk a little further about what the downside of that element of the protocol would be?

DR. FARBER: Sure. It was me. There's perhaps an ethical concern, as well as practical concern, and it's a minor risk, granted, in terms of biopsy. But in fairly certain benign lesions, the only reason you are biopsying the lesion is because you want the data to standardize the protocol for the SLA. Some patients may not then be interested in participating in the study. Others would participate and there could be a challenge in terms of the fact that you are doing things to a patient where it's unnecessary. At the very least, the informed consent needs to be much more rigorous for those patients.

DR. HARRIS: Thank you. Dr. Alam.

DR. ALAM: Thanks, everyone, for their comments. My concerns about not having everyone biopsy… First of all, to address the Dr. Farber's risk: he's quite right. People would get biopsies who don't need it, and I agree that would have to be in the consent form. I also think it's a relatively minor risk. It's not unlike someone being in a study who gets a blood draw every Tuesday. Are all of those blood draws necessary to track the change, and whatever parameters are being tracked? Probably not, but I think

it's on the same level as a blood draw. I don't think it's a major hazard. I think it would be possible to explain to a reasonable patient what this entails, and it would be possible for them to understand what it entailed. I mean, it wouldn't be exposing them to a risk that was more than they anticipated and could cause them inadvertent grave harms. I think that can be managed, and one thing we are getting at here is inconvenience of the study. Yes, it would be inconvenient to explain that to everybody. Yes, some people might say, "I don't want to be in the study," but I don't think inconvenience to the study sponsor is a compelling argument to change the study design.

In the same way, I have a concern about not biopsying every lesion. Theoretically, I do understand and obviously agree that something that is obviously seborrheic keratosis, and a whole panel of dermatologists agrees, maybe we don't need to biopsy. But – and this is a big but – this is clearly benign, this is clearly malignant… There's all this stuff in the middle. My concern is we create this concern where certain "benign" lesions don't need to be biopsied, that bar will be moved to accommodate the convenience of the sponsor to make the study more feasible. And pretty soon we will be missing a lot of lesions that really we should have biopsied. I think the dangers are twofold. On the one hand, we might biopsy some benign lesions that really probably didn't need biopsy. Okay, it's a little inconvenient. People got a tiny little ouchy due to that… not great. Still a bad thing, but I think that's less bad than not by biopsying benign lesions that were melanomas that were then missed. It's very, very difficult for us to know without histopathology which camp they fall into. We've all been deceived in our practices. Even very experienced dermatologists cannot tell this with perfect

accuracy. So, while theoretically I agree there could be some group that we don't biopsy, I think for practical purposes, it's better to do it, because there's less risk of setting the threshold incorrectly and getting bad data.

DR. FARBER: If I can very, very briefly respond to that. I concur that in able to get a better look at the performance of the SLA that you may need to biopsy everybody. However, I think it needs to be very clear to the manufacturers that they have to be very careful with the ethics of doing that in terms of assuring the patient knows they are being biopsied for the purposes of this study. Not necessarily for the purpose of their diagnosis.

DR. HARRIS: Dr. Roth.

DR. ROTH: This is Carolyn Roth. Actually, I think it might be more complicated to write a consent form in which you had to explain why some people would not be biopsied I think the consent form goes both ways. If we are trying to develop an AI tool, I think the gold standard is histopathology. So, I favor for the reasons the last speakers said: just biopsy, doing biopsies on everyone, and that will improve your data set and hopefully improve the accuracy of the device.

DR. HARRIS: Thank you. Dr. Ballman.

DR. BALLMAN: I'm becoming convinced, in a study with the accuracy of device, that even benign should be biopsied, and eligibility criteria can be set so that certain lesions are excluded upfront, that you for sure know are benign. And my question is: for a matter of inconvenience, at least for the patient, is it more inconvenient to have that

biopsy – to keep coming back as part of the trial to be reassessed – if you are not going to biopsy?

DR. HARRIS: Ms. Block.

MS. BLOCK: Okay. We have all done biopsies. We have all worked with dermatopathologists, other pathologists. We all have biopsied benign and malignant lesions. Why cannot we create a bank that these companies can use? Because, in the end, we are benefiting the patients, and this is to obviously diagnose skin cancers. That being said, I've been doing this for 20 years. I have 20 years of biopsies, probably at a lab, that are obtainable. Is there any way to obtain these biopsies, obviously contacting the patients that have had them? I get patients coming in and saying, “I want this mole removed.” It's benign. Anything that I take off, I sent to pathology, and I let the patient know that, even though I know it's benign. It's kind of like a comfort zone, obviously for me, my practice, and my patient. So, is it bad to ask for that pathology bank? Because artificial intelligence is going to be as smart as what we put in it, and I know a bunch of you have said that. So, the more data that we have, the more data we can share with that system and be able to save lives.

DR. HARRIS: Okay. Next comment, Dr. Skates.

DR. SKATES: Hi, Steve Skates. I would just like to share the results from the MelaFind study. This is definitely melanoma and cannot be ruled out the middle panel here, and definitely not melanoma. At least, definitely not melanoma according to the dermatologists. There are 83 of them. When you went to dermatopathology, there were 2 out of the 83 that were melanoma, that the dermatologists were convinced were

not melanoma. So that's part of the reason why I very much strongly favor dermatopathology on everyone entering into the study and having that as that's part of the reason why I very much favor dermatopathology on everyone entering into the study and having that as part of the criteria for eligibility.

DR. HARRIS: Thank you. Dr. Bourelly.

DR. BOURELLY: Thank you. Paula Bourelly, private practice. I have a comment of my own, and I wanted to respond to something Ms. Block just mention. If you required his towpath banking and all the biopsies from 20 years ago, if you don't have your analysis within SLA ahead of time, it feels to me like you only get one side of that equation. You can have the path, but you haven't previously evaluated them with the device, so I just think that might be what's missing there. My own comment was, again, in favor of histopath, even on things that look benign. I think what you will end up doing if you choose not to do that is missing a real opportunity to include all skin types. Again, I'm in favor of that, because you're going to look at somebody who's a skin type 4 and say, "it's dark and it's irregular, but the odds of it being melanoma are probably low." Well, I want to know what a dark irregular nevus looks like in an SLA setting someone in someone who has a skin type 4 or 5. We already said those are the patients who are hardest to get into the office, because they are not traditionally coming in for mole evaluation. They are coming in for sebderm; they're coming in for eczema. I think if we choose not to include all folks, we are really going to miss an opportunity to increase our pool from skin type 1 to 6, even though obviously, 6 is not going have as much cancer, but 6 is going to keep your algorithm how to look at a 6. Thank you.

DR. HARRIS: Dr. Asher.

DR. ASHER: Thank you. I appreciate the discussion. This is been helpful. My apologies in advance for pressing on this issue, but I want to make sure I capture your thoughts on the record. If I'm understanding correctly, your comments essentially are, for the most part, to rely on histopathology under almost all circumstances, except for benign lesions. So, if manufacturers in a hypothetical situation come to us and say, "Our device is not intended to diagnose; our device is simply a resource or app or something that's easily attainable my patients to tell them whether or not to see their provider. It's additional information in addition to their own judgment; in addition there's many downstream steps that happen in that process before you even think about biopsy, no biopsy, melanoma, not melanoma. We are just telling people whether or not to go see their doctor." In those circumstances, please understand, when we are talking about skin lesion analyzers, it includes the entire spectrum. This group at the end of the spectrum, what is your advice there? Am I to assume that this conversation pertains to that category in that situation as well? Thank you.

DR. HARRIS: Dr. Rotemberg.

DR. ROTEMBERG: I think that's exactly what I was going to address. I think there's a very big difference between skin lesion analyzer in the hands of a provider and one in the hands of a layperson, and we've discussed that at length already. In the hands of a provider, especially a dermatology provider, I agree that biopsying everything is probably appropriate, because this is where we don't know the answer and where we want to skin lesion analyzer to help us.

Now, when we talk about a layperson and a lay user, I think it becomes a lot more complicated, because the scope of the trial is going to be much broader, and biopsying all those lesions… I would like to disagree with Dr. Alam, especially if we want to include uncommon anatomical sites, feet, faces, patients who might have a higher risk of scarring… This is not the same as a blood draw. We should really be careful in lesions where everyone on this Panel would agree that something is benign, that we are ethical in our approach to whether or not we are biopsying those lesions.

Dr. Asher, when we talk about apps that are going to be patient-facing, we have two questions: one of them is sensitivity, which we discussed, and the other one is specificity and overwhelming referrals to sub-specialty care. We have to balance those, and if we allow consensus review for suspicious lesions, we run the risk of overwhelming dermatologists and performing studies where not even one of those lesions is melanoma. I've seen multiple of those studies, and they actually include zero melanomas, and they only include suspicious lesions. That would completely overwhelm sub-specialty care. So even for triage, I would argue that malignant lesion histopathology should be the gold standard, but we should be more flexible in our clear benign cases.

DR. HARRIS: Okay. Dr. Burke.

DR. BURKE: I absolutely agree with what was just said. Practical sense says we shouldn't biopsy things that are just so clearly junctional or compound nevi, or some cases that are just so absolutely certainly that. Patients come in with many, many compound, or older patients come in with hundreds of seborrheic keratoses. Where do

you stop in the biopsy of a benign lesion? It isn't quite equal to a blood draw. I think all the points that were just made is what I was just making. If there's a minuscule suspicion, then, of course, it should be biopsied and, again, gold standard should be histopathology. But I think it has to be hybrid or not biopsying things that are absolutely clearly benign.

DR. HARRIS: Okay. Dr. Suarez-Almazor?

DR. SUAREZ-ALMAZOR: Thank you. I was also in favor of a hybrid approach, because it seems that the accuracy studies may be running different populations. We all understand that histology is the gold standard, so if it is high risk with a high positive predicted value population, it should be required for that study. But if it's a study that includes larger numbers of people with benign lesions, because it's going to be for a device for general use by lay populations or whatnot – in that case, I think that this study could be done with the hybrid approach I suggested before.

DR. HARRIS: Thank you. Dr. Alam, do you have another comment?

DR. ALAM: Yeah. I just wanted to address the issue that Dr. Asher had raised. I can see we're all struggling with what to biopsy and what to not. Some considerations for study design are that we are concerned that by biopsying everything, we are potentially causing some patient harm, but that can obviously be adjusted. We don't have to skin patients alive or take every mole off of them; there can be some way to mitigate how many biopsies at maximum a particular individual could receive, which would sort of compensate for that risk. But I'm still in favor of more histopath than less. To your issue or question that… would a study or device that wasn't purporting to make

a strong claim to identify malignancy, but rather just a screening device of sorts for the lay public to know if something is suspicious or not… I do not think I would lower the bar in that case at all, because I think what you're going to end up with, from a practical standpoint, is an end run around the FDA regulations. Everyone is going to get that sort of device and prove whatever has the lowest possible bar. And, in their advertising to the patient and the public-facing, will sort of indicate it was FDA approved, and nobody will figure out the nuance between the fact that it wasn't really designed to detect a melanoma, it was really just design to detect a suspicious lesion. I think you should be very cautious in having this bifurcated approach, because I suspect, from a practical standpoint, everyone is going to want this 'weaker' indication, but it's not going to publicize that indication as weak when it comes to marketing it. You're going to have a lot of difficulty communicating to patients, even to physicians, that some apps are better than others, and some apps are only doing this, but some apps are really, really telling that something is a melanoma… I think that's nuanced; that's very, very subtle, indeed.

DR. HARRIS: Thank you. Dr. Ballman.

DR. BALLMAN: Yes, this is Carla Ballman. Again, I've been wavering on whether or not to raise my hand. In respect to the case of the app at home, and it's just to tell someone whether or not they should go to the dermatologist… I think it all depends upon what the performance metrics should be for that trial, and I think that should be the manufacturers need to come up with that to show. I think it needs to show some sort of benefit, if it be this app increased the number of melanoma diagnosis over some sort of established baseline, or something like that. I think something needs to be done. In

that case, there will be a lot of sort of people that are sort of just nervous and going to the dermatologist because the app said go see one. I agree. I don't think that histopathology should sort of be the gold standard for everyone on such a study.

DR. HARRIS: So, we had a very robust discussion. I would ask if we could project the question again so we can summarize it for Dr. Chen and FDA. Can we project question number one? So, I am going to, at the risk of failing to represent everyone – and by all means, people, please speak up if I do not include an important point in summarizing for Dr. Chen – in response to the question should histological diagnosis be required for obtaining ground diagnosis in all lesions in SLA clinical trials? I would say the Panel generally believes that to be true, but did identify important caveats and reasons to adopt, perhaps, a more hybrid approach. And that would involve and included: what would be the indications for the device, what is the clinical setting which would be used, who would be using it, is this for a layperson or a dermatologist or a skilled practitioner? There were also questions regarding the trial design and the ethical nature of biopsying lesions that are clearly benign, with the benefit being the ability to perhaps better educate or develop specific algorithms, and making sure that in some ways, we may potentially augment the input of a dermatologist since obviously no one is infallible. So, is that a reasonable summary? Anyone from panel one to add something for Dr. Chen's benefit?

And then the second part of the question, "Are there scenarios in which alternate means or combination, both histopathology for suspected malignant lesions and consensus opinion." I think I kind of already addressed that in my early summary. The

important points there were it really depends upon the indications for the device and the setting for the study and the intended user – when these devices are going to be used by dermatologists, heavily weighted toward lesions that are suspicious... There needs to be more reliance on histopathology as the ground source of ground truth, but consensus opinion might be of value in reference to develop devices that are, perhaps, simply highlighting a lesion should be referred for evaluation by a professional, and perhaps used in rural or low access environments. Any other comments about summary of that answer for Dr. Chen?

DR. SKATES: This is Steven Skates. I heard a divided opinion on the Panel. I heard some people would like to see histopathology in pretty much all trials. Other people felt there were clear exceptions, where histopathology could be replaced with consensus opinion. So, I'm not so sure that it's a clear answer. I think we've got a divided Panel here. My preference would be to the default being histopathology, unless the sponsor has a compelling argument as to why some fraction of the patients in this study should not have undergo. Maybe that's one way to split the difference here.

DR. HARRIS: Well, we are not going to be voting. I just want to make sure that all of the opinions that have been expressed have been summarized for Dr. Chen to evaluate, along with the other members of his team. Any other opinions you feel have not been included in any of these final comments?

DR. ALAM: I would agree with Dr. Skates and to do his towpath in all cases I understand that some others feel there are exceptions and histopath is not necessary. Thank you.

DR. HARRIS: Sure. Any other comments? And, Dr. Chen, is that sufficient for you?

DR. CHEN: Yes. Thank you for the information and discussion.

DR. HARRIS: So, we can move on now for question number 2, if that can be projected and read to us.

SPEAKER: Question 2A: Performance Thresholds For Adjunctive Use.

Some SLA devices may be used for adjunctive use, meaning the output will provide adjunctive information to be used: by a provider; in concert with clinical and historical information; and in reaching a management decision. The provider may be a dermatologist or non-dermatologist health care provider. The table in the following slide provides proposed performance thresholds for sensitivity and specificity for melanoma, basal cell carcinoma (BCC), and squamous cell carcinoma (SCC).

Question 2A: Performance Thresholds for Adjunctive Use

i. Should the performance thresholds of SLA advices <u>intended for adjunctive use</u> be a pre-defined sensitivity and specificity across all SLA's? For example, table 5, below. Or should performance be compared to another metric, such as the performance of the study dermatologists without the use of the SLA? Or, can adjunctive use performance be assessed by whether the SLA output improves the accuracy of the study dermatologists?

ii. If preset thresholds are preferable, are the proposed thresholds for sensitivity and specificity proposed appropriate?

If not, what sensitivity and specificity threshold do you propose?

iii. Should the performance thresholds differ if the device is intended for use by dermatologists or by non-dermatology healthcare providers?

iv. Should the performance thresholds differ based on the target diagnosis (melanoma, BCC, and SCC)?

Question 2B: Performance Thresholds for Standalone Devices.

Other SLAs may be used as standalone devices, meaning that the output will be relied upon at face value to guide care management. Devices for ley users will always be standalone.

i. First, should the performance threshold of SLA devices intended for standalone use be a pre-defined sensitivity and specificity across all SLAs, or should performance be compared to another metric, such as the performance of the study dermatologists without the use of the SLA?

ii. If preset threshold are preferable, are the proposed thresholds for sensitivity and specificity appropriate? If not, what sensitivity and specificity thresholds do you propose?

iii. Should the performance thresholds differ if the device is intended for use by lay users versus dermatologists or by non-dermatology healthcare providers? If so, what performance thresholds do you recommend for each?

iv. Should the performance thresholds differ based on the target diagnosis? If so, what sensitivity and specificity thresholds do you propose?

DR. HARRIS: Okay. My only comment before we start discussion is, in an effort to complete our discussion of the question and the next one within a reasonable period of time, try to be concise, and if someone else shares your opinion, you need not repeat it. Any comments? Dr. Rotemberg.

DR. ROTEMBERG: You know, there's been so much great discussion on this question today and I agree with a lot of what has just been said, that I don't think absolute sensitivity and specificity measures, like what has been presented in the table, are going to be sufficient for us to consider these devices. The best type of test for these devices is going to be a prospective study in the intended use setting rather than, you know, threshold on retrospective data, and in that intended use setting, improvement as defined by, exactly as Dr. Skates and others have said, value judgment, over the standard of care in that intended use setting, is going to be the best way to adjudicate these devices.

DR. HARRIS: Thank you. Dr. Alam.

DR. ALAM: Hello. I don't disagree with that at all. I think the main challenge we have today is lay devices that are used freestanding. It sounds like MelaFind with other devices is already a bit of a roadmap for how to manage the devices that are used intended to be used by dermatologists to increase their level of accuracy detection, so I'm not talking about those. If you are looking at lay devices, I think it is important to have pretty rigorous benchmarks for those, and we can review that with sensitivities, specificities, positive predictive value, risk, whatever. If we are going to go with sensitivity and specificity, I think the numbers, at least for sensitivity, should be

markedly higher than those that are expected for a device that is used by dermatologists or other practitioners. I would anticipate the number I would want for a lay device to be at least 95% for sensitivity, and something like 80% to 90% for specificity. I understand that's not easy to achieve, but I'm not the goal of FDA is to facilitate the proliferation of very bad algorithms that don't do very much. I'm sure there are enough of those already. So, thank you.

DR. HARRIS: If I can just prompt you to add a little bit of a comment on the other aspect of the question: would you want those thresholds to vary based upon what type of lesion was being diagnosed?

DR. ALAM: Yes, sir. I would like them to vary a little bit. I think there's a consensus on the panel that melanoma is the most threatening of the tumors. I would probably refer to melanoma. You know, a sensitivity of slightly lower would probably be acceptable for BCC and SCC, maybe 5 percent lower, more like 90% than 95%... but somewhat lower, I think, would be acceptable for non-melanoma or lesions not suspicious for melanoma.

DR. HARRIS: Thank you. Dr. Skates.

DR. SKATES: Steven Skates. So, on the first question about setting a pre-defined sensitivity and specificity, or can adjunctive use performance be assessed… I would like to push the adjunctive use performance being the increase in sensitivity that a provider can achieve with the use of the device, compared to what the provider can achieve without the use of that device. That's an interesting question as to how to do that. Should that have the provider assess the lesion and make a judgment as to what

should happen next, and then use the advice on that lesion and see whether that judgment changes? That's with in-person design. And then you can get the increase in performance and accuracy with the device compared to that person's performance across a population they are assessing. You can obtain that increase in performance by comparing them with and without the device. The alternative is to randomize patients to providers that use the device, compared to providers that don't use that device. That would be a prospect of randomized trial, and that would be more burdensome, but there is a cleanness about that, where no one would argue, if there's a positive result there, that that would be correct. But with in-person result, is also a feasible trial design, and I think should be considered and be less burdensome than a randomized trial.

In terms of the specificity, I want go back to making a judgment of benefits to risk, and working out, from even these two studies, the MelaFind and Nevisense studies, what the ratio is in those studies that the dermatologists achieved between the false negatives and false positives, the weight between those two as an average weighted risk. And I would suggest that that benefit to risk ratio, achieved by a dermatologist, is what devices in the hands of primary care physicians or in the hands of patients should be achieving to pass the criteria of safety. That gets away from specificity arbitrariness. It says, can we achieve the safety of dermatologists achieve in their usual practice. So, I would say that is the criteria that should be applied and can be applied across all settings here, from lay user to primary care physician to dermatologist.

DR. HARRIS: So, just to clarify so I understand your statement. In the hands of a dermatologist, an adjunctive evaluation, whereby we can demonstrate improved performance, would seemingly be an acceptable standard for a device to achieve?

DR. SKATES: Yes. Sensitivity. So, for effectiveness. Safety and effectiveness is the FDA's mantra, or criteria that they need to have studies show. So, for effectiveness, the sensitivity needs to increase; that would be across the board. And the question is, what is the specificity? How do you judge that? You can get a decrement in specificity and still have the benefits outweigh the risk, so then the question becomes how big a decrement is reasonable? And what I'm suggesting is that you use the benefit-risk ratio that you see in dermatologists' setting and apply that, and we've got two studies the FDA has reviewed previously, Nevisense and MelaFind, that could inform the benefit-risk ratio in the hands of dermatologists, and apply that to PCP's and to the lay audience.

DR. HARRIS: Just so I can clarify, you would not be in favor of a scenario whereby you might double the diagnostic specificity of the provider, non-dermatologist provider, if it fell short of the performance level of a dermatologist?

DR. SKATES: That's an excellent question. If you can increase — it would have to get close to a dermatologist. I don't know that that's meeting exactly the dermatologist criteria, but getting close to it, I think a sense of how far away we are from it… we need to set a bar for safety. That specificity and the incidence mix in a hard to fathom way, but the constant should be benefit to risk ratio; that we set a bar and achieve that across all these types of studies.

DR. HARRIS: Thank you. Dr. Ballman, I saw you vacillate.

DR. BALLMAN: I vacillated because I think others have said what I think should be done. I think trials need to show improvement in accuracy. Also, without, you know, without sort of sacrificing too much with the others. An ideal trial would show an improvement in sensitivity, and the specificity did not decrease, right? And then going to Dr. Skates, maybe we need that risk-benefit ratio because maybe you can give a little bit on specificity, but I think it needs to show for all studies that there is an improvement in accuracy with using the device over the current standard of care.

DR. HARRIS: Okay. And Dr. Alam.

DR. ALAM: I would agree in the setting where it's in adjunctive use, you want to see if it's increasing accuracy, however you define that, wherever it's currently at. I'm not sure how that would apply, though, for a layperson, because I'm not even sure what the baseline of detection for a layperson is and how you would detect if their accuracy are increasing. That's why I was just suggesting – I guess they are arbitrary, but there's some arbitrary benchmarks that other physicians have been queried specifically about that question think are appropriate. It's kind of like a Delphi process, which is sometimes the best you can do: consensus in the absence of very rich data. I think that's what my main concern is; I'm not so concerned about FDA regulation about the devices that are used for adjunctive use, because it seems like FDA has a pretty good handle on that. The area for potential problems are the devices that are going to be in the lay public and are going to be marketed very aggressively. Once they are somehow given the perimeter of FDA, no one would really think about the nuance of how they

work or what they were intended for – they will just say they're approved. So, it's sort of a high bar for us to make sure what gets out there is reasonably good. I think just being marginally better than a confused person being able to find a lesion on themselves is not good enough. I would like it to be, like I said, a sensitivity of at least wanted dermatologist would do, but ideally a little bit more. Again, it's a screening test for a layperson, so we are focusing primarily on sensitivity, but of course we would like specificity to be extremely low. Thank you.

DR. HARRIS: Dr. Farber.

DR. FARBER: Neil Farber. Thanks very much. I agree in adjunct of use simile showing an improvement in the hands of either dermatologists or non-dermatologist is perfectly acceptable. I also agree though in the lay public I think it requires some greater attention to assuring there's a baseline of sensitivity, especially when it comes to looking at melanocytic lesions. The other thing I would add is looking at the risk-benefit ratio, as Dr. Skates had pointed out, I do think the issue of the psychological impact on patients who are the lay public and using this device be looked at and there are many ways of doing that. I won't go into it here, but we would be glad to talk about them at some point in time.

DR. HARRIS: Dr. Bourelly.

DR. BOURELLY: I'm not going to repeat what was already said, but for the second part of the question, I believe, for the layperson or non-derm provider, the standard be at least that of a dermatologist's , if not higher. My reason for that is, I assume that may be the last step for the patient. Even if the patient got so much better,

or their non-derm got so much better, if it's not at least the standard of what a derm could achieve, then that person would probably never get referred. That person will probably never show up to a derm, and we are going to miss some people.

DR. HARRIS: Any additional comments either on the first part of the question or regarding the second part? As we think about evaluating -- utilizing these devices to evaluate lesions that are not melanomas and how this performance should or should not compare?

DR. ALAM: Can I make one brief comment, please?

DR. HARRIS: Please.

DR. ALAM: I think we should encourage FDA, when they decide exactly what benchmarks are going to be used for SLA devices in particular, to come up with disclaimers that are suitably clear to the public. I'm sure that's the goal that I think that will be important. Not everyone will read them, but if someone does, it will be important to understand for the average user what exactly this device can do, what it's intended to do, and what might be more than it can do. Thank you.

DR. HARRIS: MS. Hesser?

MS. HESSER: I think we need to be cognizant of what the term FDA approved means to the lay individual. To a patient who sees tools that are FDA approved, will assume that each of those is the gold standard. So, I would be very much in support of all of the FDA SLA tools meeting the same gold standard.

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

DR. ROTEMBERG: I think that was a really important point and I appreciate it. Oh, I'm Veronica Rotemberg. I think, for me, the comparator to the standard of care is so critical because we know that there are patients who don't have access to sub-specialty dermatology care. So, improving what they have access to in terms of diagnosis is critical, especially to early diagnosis of skin cancers, and especially melanoma. I also think that this discussion of lay devices is going to end up being just as nuanced as everything else that we are talking about because, exactly as Dr. Alam said, if the device says "you should see a dermatologist for this" and no other information, that's very different from "that lesion is suspicious for melanoma." That also is going to change the psychology of what the patient believes when they are in the dermatologist's office. This is why, even though I know how difficult trial in the lay population will be, these questions… we just don't know the answer to them in less we test them in real life. And we don't know how many additional biopsies we would cause with specificity of 80%. We don't know how confident a dermatologist might be to say, actually I'm not suspicious about that lesion, we can just watch it or not biopsy it… We don't know any of that until we study it in real life. So, I agree the standard for sensitivity should be high. I think the standard for specificity should also be high. But we should really think about them and we should demand prospective trials to analyze this, too.

DR. HARRIS: Dr. Roth.

DR. ROTH: This is Carolyn Roth. I have a couple of small comments I agree with most of what has been said. I think the standard for the primary care physician should be to bring the primary care physician up to the performance of a dermatologist. So, I

think that's very important. In terms of the layperson, I think that it would not be inappropriate for a recommendation that patients consult with their primary care doctor or dermatologist. I think we are leaving the primary care physicians out of the loop here when we talk about laypeople. But I also want to come back to the points that were made earlier today about less being known about the diagnostic accuracy of these devices in people who are in higher Fitzpatrick scores, and people of Asian and South Pacific Islander, as well as African-American descent. I don't know who I'm leaving out, I apologize. But for people who are past Fitzpatrick class II, I think we might also specifically encourage more extensive testing of these devices in these populations. Thank you.

DR. HARRIS: Thank you. I'd just like to propose a quick question and get some feedback on the panel. It sounds like we are in favor of devices that can either elevate the performance of layperson and/or private practice or PCP to that of a dermatologist. And then, pf course, further elevate the performance of a dermatologist with these devices. I'm just curious, if that were the standard we wanted these devices to meet, we are saying we would not want the device to give a particular diagnosis, or perhaps even a differential, to a layperson. Even though we are demanding or wanting these devices to perform at this level. Is that what we are saying? Anyone?

DR. ALAM: I think that's correct. If I understand, I think we do want devices to perform at the level of a dermatologist at least. But we are concerned, as has been very eloquently discussed by others, about creating anxiety for patients or otherwise

making them uncomfortable. So the goal would be to get them in a physician's office, ideally dermatologist's office for further management.

DR. HARRIS: And this is perhaps a question for Dr. Farber, since I think you have background in this issue of anxiety, is there a substantial difference in the level of anxiety being told 'you need to go see your doctor about that skin lesion' versus 'that could be an melanoma.'

DR. FARBER: First of all, yes. But I think that depends on how it's phrased. If it's phrased in, you know, you do need this lesion. It would be advisable for you to see your physician so that they can discuss it with you and if necessary do further tests to diagnose what it is. That's one thing, as opposed to saying 'you have a serious skin lesion. You have to go see or PCP.' We don't want to do that. On the other hand, saying 'you have melanoma, which is a type of cancer; you need to see your physician right away,' is very anxiety-provoking. Patients, oftentimes when they hear the word cancer, hear nothing else beyond that. And you can't obviously phrase it in carcinoma. So I think the way it's phrased is most important. But, yes, I think we don't want patients that are using the device to have a specific diagnosis. The other thing is, they go to see their dermatologist, let's say, and the dermatologist says, "No, it's not that. It's this instead," then the patient is likely either to not trust either the device or the dermatologist because there's a difference. So, I think they should just be directed to see somebody.

DR. HARRIS: Dr. Asher.

DR. ASHER: Yes, two comments. One is: the team is reminding me that the numbers the Panel has been provided in Table 5 are numbers based on dermatologist performance. And so, the question that they have is, is the recommendation for sub-studies to be done in each case with these prospective trials, characterizing dermatologist performance and proposing something equivalent or improved over that? So, that's one question. Then, the second question, and I have to take responsibility for this, but I do need to press, and that is it sounds like some of the recommendations are suggesting a higher sensitivity and specificity more because of the concern about problematic risk communication. So, I guess my question is, can the Panel envision better ways to communicate risk more accurately so that there is not this compensation by default; we are just asking for a very high threshold of sensitivity or specificity because we think people will misunderstand. Two questions. Thank you.

DR. HARRIS: Dr. Suarez.

DR. SUAREZ-ALMAZOR: I think that's a very good point about risk communication and whether we are requiring a higher sensitivity, or, for some Panelists, because of the consequences of patients not understanding SLA setting. I think that ties also with how the information should be given. From what was discussed before, my concern about just saying ‘you need to see your dermatologist,’ there's two issues. The first one is, these patients are using that tool as a screening tool for melanoma. They know that if it tells you to go to your dermatologist, it’s because it might be melanoma. Not having anything else in the descriptions as given to them might even provoke more anxiety than having something that's better explained. So, I would be in

favor of explaining a little more and not just, 'you need to see you dermatologist,' because, for some people it may create more anxiety, and others, it may not be a trigger that's sufficiently high for them to go and see a provider. Again, I think it needs to be balanced. And you never know what might be the best language; I wouldn't be too blunt, but I wouldn't be too vague either. I would try to go somewhere in between. I think going back to the question about higher level of sensitivity for certain populations… I mean, I'm not sure about that. I think that in order to judge that, more information would be required on how population react to this information before deciding what the thresholds need to be.

DR. HARRIS: Dr. Skelsey.

DR. SKELSEY: Regarding the anxiety patients experience, do others in the panel think that can be alleviated to some degree by giving some more information about the likelihood —

DR. FARBER: No.

DR. SKELSEY: — and the numbers of melanoma that are successfully in an earlier state. People are purchasing these devices, and there is that risk of having an anxiety-provoking diagnosis. I wanted to get back to the issue of follow-up analyses, looking at different users of these tests. I think it's critical that the companies look specifically at different providers in terms of their training and experience to see how these devices are used, and that means — you were going to be using it in a lay population, but for some of them and others for providers — but I think we need to look at all of these, and follow-up on how they are actually utilized for different levels of

training. So, I agree with how you summarized, Dr. Harris, but I did want to make that point that I think it would be helpful for the FDA to require that, say, users with different levels of education and experience and how they utilize it. Is it different for somebody who has a Ph.D. versus someone who went to high school?

DR. HARRIS: Thank you. Dr. Rotemberg.

DR. ROTEMBERG: First of all, I would just press that the idea of how a layperson would respond to an app has not been studied. We do not know, in this use case, whether 'just go see your dermatologist,' or 'this looks like melanoma,' or 'this has a 10% chance of melanoma,' or '20% chance of basal cell,' we do not have tests that tell us what will happen. I think that it is probably too ambitious for this Panel to say definitively what would happen in that situation, and I think we absolutely need studies to tell us the answer to that before we can understand the balance of risks versus harms of those different types of communication.

I think the answer to your question, Dr. Asher, is not that we are worried about communication, it's that it's a very low-prevalence population. So, the general population in the United States does not have a high risk of melanoma. So, if we are going to launch an application that diagnoses melanoma in 350 million people, including a 2-year-old, which has almost a 0% chance of melanoma, we want to have very, very high standards because the prevalence is low. Not because of the challenge of communication. And I think that's the difference between a dermatologist, who already has a lesion that's suspicious – the chance that's melanoma is so much higher, just your

pre-test probability, than a layperson taking a picture. We need to take that into account.

DR. HARRIS: Dr. Bush.

DR. BUSH: Laura Bush. I was just going to say that I would feel it would need to be at least equivalent in sensitivity, because we want to make sure that it does add value. And the points with it being higher than a dermatologist may be because we are worried a little bit about false sense of security for people, in addition to points Dr. Rotemberg said as well. But I feel that you might have patients that would have a false sense of like, oh, I'm okay, and they just looked at the one lesion they were worried about. The one thing we were discussing using laypeople, using communication and feeling anxious about it, possibly they could market it with descriptors, such as green light, low risk; yellow, moderate risk; and red, you need to see your doctor… kind of in a broad brush of categories. To not say you have a melanoma or you don't, but kind of maybe do the studies of guiding them to where they need to be.

DR. HARRIS: Dr. Skates.

DR. SKATES: Hi. Steve Skates. I envision these devices having multiple different — for each device probably the manufacturer is going to identify slightly different, or may be quite different, study populations for the device to be applied, and that's going to have the prevalence all over the map. And it's for that reason, and in those different populations, it is likely that dermatologist is going to have different sensitivities; that their performance is going to differ. I want to get away from having an absolute sensitivity. I want to get to a point where you can say the device increases

what would be done without the device. That's captured in question 2A part one, where you say at the end can adjunct of use basis by whether the SLA output improves the accuracy of the study dermatologists. And I would say that's what I think is the criteria that should be used to assess these SLA's, rather than saying we want greater than 90% sensitivity. Because a good all depend on population. I want to have the device improve what's currently done rather than — and that could vary all over the map. I think the better way to say it is improve what's being done.

DR. HARRIS: Dr. Farber.

DR. FARBER: I want to echo what Dr. Rotemberg said about the fact that we don't know how patients react to something like this on an app, and I fully agree with that. And actually, that study should be done. I do know there are data about patients being informed of the diagnosis of cancer. Not specifically dermatologic cancer, in which there is a lot of anxiety, and basically lack of communication once the word cancer is used. So, think it is important to phrase the information to patients, if it is to be used by patients, in a manner that would not cause anxiety, but rather would encourage to see either their PCP or dermatologist.

DR. HARRIS: Great. Dr. Alam.

DR. ALAM: I wanted to agree with Dr. Rotemberg. I think she brings up a very important point. And to the point that some of these lay apps may be intended for particular source of lay populations, while that may be the case, the reality is once the app gets out, people are going to use it willy-nilly, and whether or not they are going to read the disclaimer, ‘this is only good for people with a particular skin type within a

particular age'… They are going to use it regardless. So, those apps have to be sufficiently robust that they are providing reasonable information. Of course, exactly as Dr. Rotemberg said, while the prevalence of skin cancer is relatively high in the U.S. over all because there are a lot of white people, it's obviously going to be much lower in the general population than in a selected high-risk population in a dermatologist's office. So, to get an equivalent likelihood of detecting something bad, you would need a higher sensitivity. I also agree with Dr. Skates that other metrics may be necessary, but the bottom line point that the test probably needs to be better than a test use in a doctor office or dermatologist office for high-risk patients. Thank you.

DR. HARRIS: A couple of more quick comments, and then see if we can summarize and move to our 3rd question. So, Dr. Ballman.

DR. BALLMAN: I mean, I think I am agreeing about the high sensitivity in a layperson population. Someone just using it at home, but I think as was brought out, and all of us know that screening in a general population is a very, very high bar to me, right? We don't screen for all sorts of things, you know, just because they are low prevalence, right? I think that automatically would make the sensitivity go up. But I really think the bar has to be that the manufacturers have to show a benefit of using their device. If it's a low prevalence population that's going to have to mean that that sensitivity is going to be up there, but I would prefer to focus on showing sort of improvement over some metric.

DR. HARRIS: Dr. Rotemberg, your hand was up and now it's down.

DR. ROTEMBERG: I just want to agree with Dr. Ballman. So, I agree with what's been said.

DR. HARRIS: Thank you. Dr. Burke, final word. Well, not final word, but next word.

DR. BURKE: Yes. I just think I agree with Dr. Skates that the SLA output should improve the output of dermatologists and should be showing and improves the output of primary care physicians approximately that of dermatologist. And I think I prefer that paradigm to just numbers of sensitivity and specificity. And also, I just don't think this is for the lay public yet.

DR. HARRIS: So, in an effort to summarize for Dr. Chen, can we re-project the question?

SPEAKER: Do you want me to advance the slide?

DR. HARRIS: Please advance the slide. I think it has that. Correct. Once again, before we provide a summary for Dr. Chen, we are saying we want — I think I'm hearing the Committee say that the device needs to improve the performance of whomever is using it, whether it be a dermatologist or a primary care physician or the lay public. Are we able to provide any – and I know that we heard in great detail from Dr. Skates the preference for using that versus relatively or pre-identified specificity and sensitivity cutoffs. Are we unable to suggest any numbers in that regard for any of those users in any of those settings?

DR. ROTEMBERG: I would suggest that we have as a minimum 10% improvement in performance, but I agree that it would be better for it to be even higher,

but I think at a minimum we should expect a 10% improvement. And that's been in many type of perspective clinical trials. That can be an achievable benchmark.

DR. HARRIS: And so, these benchmarks that are on this table are currently projected, represent the performance of dermatologists, if I understand this correctly. So, are we saying that we would want these devices to meet, and/or exceed these, or establish independent criteria in the study itself?

DR. ROTEMBERG: Yeah. It needs to be tested against its intended use setting because these numbers are an approximation over many different intended use settings and many populations and shouldn't be generalized in my opinion.

DR. HARRIS: Any other comments on that regard?

DR. SKATES: I completely agree with Dr. Rotemberg on that. It's the change we want. So it's that second paragraph of that one that I think is what's needed.

DR. HARRIS: So, hypothetically, if a manufacturer were to conduct a trial and it ended up that the sensitivity of the providers was 50% and the device was 60%, or 65, we would be in favor of that device being approved.

DR. SKATES: So, I'm probably not the right person to speak on that, but that would satisfy the effectiveness criteria. The safety criteria would also have to be judged, and that is this benefit-risk ratio.

DR. HARRIS: All right. Any other comments before we summarize and move on to question 3?

DR. ALAM: I think teamwork example, Dr. Harris… that was an interesting example. I would be concerned if the increment of baseline was uncharacteristic for

that patient population. We certainly wouldn't want studies done where, I don't know, dermatologists were blindfolded, and then the computer was a little better than the blindfolded. We would want them to be at least as good as they routinely are in the device could be better. Thank you.

DR. HARRIS: So, Dr. Chen, this has been pretty nicely summarized. The attitude is that these devices, the criteria they need to meet in terms of their performance, is they are better than what is currently available in the practice setting, whether it be a dermatologist office versus comparisons in a primary care physician's office, or what the lay public I would have access to otherwise. And that the issue of safety, which is kind of what Dr. Alam was speaking on, there would be a separate criteria so that there wouldn't be an unusually poor performing device that, just by the lack of trial design or that particular study, somehow achieved approval. Does that help you? Does that satisfy?

DR. CHEN: Yes. That information is enough for us to think about and move forward for the next step. Thank you.

DR. HARRIS: I think, importantly, no one seems to be comfortable with providing or pre-ordaining kind of across-the-board sensitivity/specificity. That is also a big part.

DR. CHEN: Yeah. We heard that.

DR. HARRIS: All right. We can now move on to question number 3 and have that objective read for us.

SPEAKER: Question 3: Performance in the U.S. Population. Panelists should consider whether the skin lesion analyzer devices must be able to analyze skin lesions

with an acceptable sensitivity and specificity in all patients prior to FDA clearance, or whether proof of performance data in higher-prevalence populations, for example, non-Hispanic white individuals, can be provided to allow these high-prevalence populations access to this technology, followed by clinical studies in low-prevalence populations.

The potential benefit of a stepwise approach is that it may allow for earlier access to this technology for populations at high-risk, but it may increase the risk of false positive and false negative results in lower prevalence populations in whom the device has not been inadequately trained and tested

However, requiring SLA to be tested in patients with lower incidence before entering the market could delay the time to market due to extended enrollment times for statistically relevant numbers of darker skin individuals with skin cancer.

Should FDA allow skin lesion analyzers to be marketed based on study data from a limited U.S. demographic – for example, in higher incidence populations – with subsequent data collection in lower incidence populations to explain the indications for use?

Or, should the FDA require the training of AI/machine learning (ML)-based skin lesion analyzer technologies in all populations regardless of specific cancer incidence?

Although the previous questions have focused on skin cancer, skin lesion analyzers may also be used for other lesions other lesions that have similar prevalence across all US demographics but look different in different Fitzpatrick skin types.

To ensure generalizability across the entire US population, should FDA require all AI/ML-based skin lesion analyzers indicated for use beyond cancerous lesions to be trained and tested in a representative U.S. population?

DR. HARRIS: Okay. Comments? Dr. Bourelly?

DR. BOURELLY: I think it should include all skin types. I think there is some room to decide what proportion needs to be skin type III, skin type IV, skin type V and VI. And that can be reflective of what we expect the incidents to be in that population, but I think we are missing an opportunity if we don't include all skin types. And if we haven't learned from the last two years how healthcare disparities impact our country… We should know right in this moment that we have an opportunity to set an example for other studies unrelated to skin cancer by including all skin types. Thank you.

DR. HARRIS: Thank you. Dr. Alam.

DR. ALAM: I would agree with Dr. Bourelly that would be a good call. This is kind of the tip of the spear. And if we don't do it now, it might just not get done for other diseases as well. It's very inconvenient. For skin cancer it's quite difficult. And maybe, even with the best efforts, it won't be as good as it is for skin types I and II, to detect a skin cancer using an algorithm of this nature, especially for laypeople. I think it would be a very, very important effort to undertake and I think it would be to the credit of FDA to insist on that and facilitate on that and find a way to make it work.

DR. HARRIS: Thank you. Dr. Ballman?

DR. BALLMAN: I agree, all skin types. There could a possibility, if there's enough power and one group and sort of report out and act upon that while you are still — and

in the meantime still continue accruing to the other groups to make sure that you can confirm and/or tweak the algorithm if necessary.

DR. HARRIS: Thank you. Dr. Rotemberg.

DR. ROTEMBERG: I think these are all really important points, and I'm so grateful to the FDA for having Dr. Adamson come talk to us about his work. I think that the distinction between diseases of similar prevalence and the diseases of disparate prevalence is really important. The value of recruiting from all skin tones cannot be understated and it needs to be emphasized by the FDA, but most algorithms that are trained to detect melanoma are trained on thousands of melanoma examples. Something like 20,000 melanoma examples. Given that there are only a couple hundred melanomas and non-white patients in the United States, it's not going to be possible to create an algorithm that has equivalents in that population. So, I think there is going to be a need for balance. The thing that I think is super important is going to be, of course, the diseases where the prevalence is more similar, and the safeguards that we ask of the manufacturers when algorithms are deployed. Because as Dr. Cohen said, people don't read the warning labels. They don't read the disclaimers, and then all of a sudden these algorithms are being deployed in populations when they haven't been tested, and have every expectation that will perform badly. I would say we should ask for automated safeguards for algorithms to say, I don't think this is within my scope. Very clear warnings and communications and other things like that if we are going to allow these devices to move forward without increasing the prevalence of diverse skin tones.

DR. HARRIS: Dr. Farber.

DR. FARBER: Neil Farber. The other issue I would chime in on is approaches from an inclusivity point of view rather than exclusivity point of view, and therefore, I said absolutely no benefit of not including patients of different skin colors.

DR. HARRIS: Thank you. Dr. Roth?

DR. ROTH: Dr. Roth. I agree with what everyone has said that we should be as inclusive as possible. I do believe that it is possible to design the algorithms in such a way that people have to input information about their age, other aspects of their life, maybe even photos of their skin color, before you actually look at the lesion. So the analysis could be mitigated by the kind of data the algorithm receives. I think that it is possible to create an app that would, in fact, be respectful of those differences, but I don't think we should hold this up… But we should continue to test and we should continue to accrue data, but I agree it should be as broadly applied as possible.

DR. HARRIS: Dr. Skelsey.

DR. SKELSEY: Thank you. I think it's the most ethical path to take into consideration all skin types. Does that mean, however, devices that can't access areas – like as Renata Block pointed out, acral skin, genitalia; neither Melafind nor Nevisense for instance can do that – should those be excluded because they are going to not be able to assess areas where we are more likely to see tumors and skin of color?

DR. HARRIS: Okay. Dr. Burke, comment?

DR. BURKE: I was going to make exactly that point. These measurements really at this time can't quite measure palm souls and lesions under the nail. So, we just have

to be aware of that, but absolutely acceptable to treat all skin types. Also, since we are increasingly a mixed race population, I think the genetic data to some degree, one or two generations of data of ethnicity, should be included because there are so many mixed races now. So, somehow you have to include all races very much in the study, but also have input data for, let's say one or two generations of ethnicity.

DR. HARRIS: Thank you. Dr. Skates.

DR. SKATES: Steven Skates. I maybe misunderstood this question, but I had the sense that the assumption would be that all races and skin colors are to be included in the study, but should there be a result with the higher incidence earlier on? Is it okay for the FDA to allow the company to proceed with marketing in that population with the provision that we keep on enrolling in the lower incidence populations until you have sufficient numbers, and then roll out the algorithm in a phased way? That was my interpretation of the first part of this question. Or, the other part is, should the FDA wait until every skin type has been assessed and sufficient numbers and all types ranging from low instance to high instance be achieved before the company is allowed to put it on the market? I want to see if that was what the FDA was asking because that was my understanding of this question. And it's more nuanced. Absolutely, all skin colors and types be involved in the study. But can the company roll outs on particular skin types because it's higher incidence before it rolls it out in lower incidence skin types… I would like to hear the Panel's views on that. First of all, is at the FDA's question? Second, if that is, I would like to hear what other people on the Panel feel about that.

DR. HARRIS: Can we get a response from the FDA answering that question? Is that an accurate interpretation of the question you are asking us to address?

DR. ASHER: Yes, this is Bonita Asher. That is accurate. We want all medical devices to be created and applicable to all patients, consistent with the U.S. population, but given the epidemiology and some of the challenges, what are your thoughts around that? There is benefits and risks to both approaches. One is delayed market entry to the population most at risk. The other is that perhaps there's going to be significant delay in conducting the studies in populations that have lower risk. So, we were asking the Panel to contemplate that and to give us recommendations on how we can achieve what we understand to be the ideal, is that, we are studying the population representative of the diverse U.S. population. Thank you.

DR. HARRIS: Thank you. Dr. Alam.

DR. ALAM: I agree with what previous speakers have said, and I thank FDA for the clarification. I think, as Dr. Rotemberg has said, it's reasonable to have some kind of disclaimer. It's reasonable to continue to accrue patients. Even with all of that, and even with significant time elapsed, it is probably going to be the case that the apps will not be as good for types 5 and 6 patients as it will be for other patients, just because of the inherent limitations in how many such patients will be enrolled.

To answer the FDA's question, on the one hand, it would be nice to get the access sooner rather than later, whatever these devices are, because the very high at-risk population, that generally is not skin of color. However, I think one of the concerns that you highlighted, Dr. Asher, is that if that occurs, the incentive for companies

marketing these products to actually follow through and enroll skin of color patients in a timely manner, or at all, can be somewhat diminished because their primary consumer is quite happy and… things happen. So, I don't know how that is balanced, but I would suggest FDA consider not only a carrot, but some kind of stick. So, if you feel like these devices should go to market even as they're enrolling or while they haven't finished enrolling accurate numbers of skin of color patients, you might want to have some provision whereby, if they haven't succeeded in doing that in a timely manner, however many years that is, they lose approval of their device or its rescinded or mitigated in some manner. Thank you.

DR. SKATES: Steven Skates. I was going to ask, what are the sticks that FDA has to make sure that, if there is an early rollout, that the studies continue in 4, 5, 6 Fitzpatrick scale patients? I don't understand the sticks. Perhaps if we were reassured that there were sufficient sticks, that might be reasonable approach.

DR. ASHER: This is Bonita Asher. I think, first and foremost, we take advice that you provide to us and your recommendations very seriously. If you have strong recommendations in this regard, I think that is helpful. We have mechanisms by which we may consider the need of post-approval studies if there is a safety concern potentially regarding the use of device in certain populations. Beyond that, I think you can be creative and imaginative in suggesting what you think the right thing is to do, and we'll take that and try to put that into the regulatory framework, and see how your scientific recommendations can result in the regulatory stance supporting the science.

DR. HARRIS: Ms. Hesser.

MS. HESSER: I believe the FDA is in a position to be able to encourage industry to engage patient advocates of all skin colors in the development of their clinical trial protocol and bring people to the table early on in that clinical trial development. That will help offset some of this post-approval conversation. Strong advocates can help steer in the correct direction and be representative of a lot more populations than we are seeing now in SLA.

DR. HARRIS: Thank you. Dr. Burke?

DR. BURKE: Yes. I absolutely agree with everything. It says we must encourage further studies now and in the future. I think this is relevant to our discussion tomorrow about having this classified as a Class III, as opposed to Class II, because perhaps class III makes it more possible to have post studies, increasing studies after approval.

DR. HARRIS: Thank you. Dr. Rotemberg.

DR. ROTEMBERG: Two things I would say. I wonder if there's a framework that compares the prevalence of the training data used to develop a model, as compared to the disease prevalence in the U.S. population. So, for certain diseases, like psoriasis, I might not think that we should even allow initial launch of a study that does not have diverse skin tone representation, whereas in a disease like melanoma, where we know it would take 10 years to recruit 4000 melanoma patients of skin tone 5 and 6, that might be something that could be discussed by the FDA. Setting criteria that clearly evaluates the training data that was used as compared to the specific diseases that are being evaluated, I think would be a really valuable place for the FDA to be a leader.

The other thing that I would say is around labeling. There's a lot of effort around model labeling and certification for how fairness has been incorporated into data and model development. I think this is something that the FDA could also collaborate with, in terms of transparency around data that is being used for model training, and ethnicity and skin tone labeling for that data, transparency around decisions that were made in terms of oversampling, synthetic data, other types of technical decisions that are made by the model developers… and just making sure that that type of transparency is required. I think that would go a long way. How to communicate that to patients and end users is a longer discussion we might not have time for today, but all of that is going to play a huge role in how well algorithms perform in diverse populations. And of course it's going to have to be compared to the underlying prevalence.

DR. HARRIS: Thank you. Dr. Bourelly.

DR. BOURELLY: Quickly, I think with the first thing you said in the first half of that, Dr. Rotemberg, is my definition of equitable care, not identical care. You won't get 4000 melanomas, in short. Of course, I do want to remind that basal cells and squamous cells are also seen in Fitzpatrick 3 and 4. We are not just talking about melanomas, although I know that's our killer. We're talking about all comers. Again, thank you.

DR. HARRIS: Thank you. Any other comments? If not, can we please re-project the question? I will try to summarize. So, should FDA allow these devices to be marketed based on study data from limited U.S. demographic? What I heard unanimous was all skin types should be studied, but that there seemed to be allowance

for these considerations given to allow these devices to be marketed prior to having a full repertoire data represented of all skin types. Is that accurate, Panel?

DR. ROTEMBERG: I'm so sorry, Dr. Harris. Could you repeat it again? I think I might've missed some of the detail.

DR. HARRIS: Not many details. Just saying that the panel was unanimous in its endorsement of companies and these devices engaging in studies that include patients of all skin types, but that there seemed to be an allowance for some devices to be marketed prior to having a fully robust data set that encompassed all skin types, with the acknowledged concern that may eliminate the motivation for these companies to continue collecting the more difficult or less accessible data. Some suggested that might be mitigated by mandatory post-marketing efforts, and perhaps other strategies that FDA could employ to ensure that these companies don't lose their enthusiasm for the collection and incorporation of the more difficult or less accessible data.

DR. ROTEMBERG: I would just add under certain conditions of transparency and prevalence.

DR. HARRIS: Okay. So, Dr. Chen, do you have an adequate appreciation for the Panel's deliberation on this question?

DR. CHEN: Yes, we do.

DR. HARRIS: Thank you. I would now like to ask our non-voting members, Dr. Bryant, our Industry Representative, and Ms. Hesser, our Patient Representative, if they have any additional comments. Ms. Hesser, do you have any additional comments for us?

MS. HESSER: I have no additional comments, other than thank you for listening to the patient perspective, for allowing me to represent the interested patients, and I appreciate that opportunity.

DR. BRYANT: No additional words. I would just say kudos to this Panel – very enlightening – the level of discussion was educational and inspirational because the phone on this call are really dedicated to patients. And also, kudos to the FDA for having a Panel, having the folks come in and speak that you did… Very, very informative; this is great.

CLOSING COMMENTS FROM FDA AND PANEL

DR. HARRIS: Thank you. At this time, we have the opportunity to hear a summation and comments and clarifications from the FDA. Dr. Asher, you have any additional comments for the Panel?

DR. ASHER: Dr. Chen, did you have one item you wanted to bring up related to question 2?

DR. CHEN: One. Earlier, when we the panel discussed the question number 2, specifically question 2B, related to the lay user use, can the panel clarify again the extent of sensitivity/specificity or the accuracy that we need to pay attention to?

DR. HARRIS: So, it was my impression and by all means, Panel, please correct me if I'm wrong. The general consensus was that the devices should meet essentially the same standards as the devices that would be used by non-laypersons.

DR. ASHER: And standalone use as an adjunct, your recommendations are the same. Is that accurate?

DR. ALAM: If I may, I thought for standalone, we wanted slightly more rigorous standards. And I think Dr. Rotemberg had indicated, and others had agreed, that maybe the sensitivity should be somewhat higher because we are now using these devices in a very broad population. And without any physician oversight. So, I think – personally I think the bar should be higher for devices that are used by laypeople in a freestanding environment. When they are used for adjunctive use, the standards seem to be, they should improve people's performance over where it currently is, and there was a preference for randomized controlled trials to show that, but I think it's a little different in a lay setting where the bar should be higher. Thank you.

DR. HARRIS: Any other comments?

DR. ROTEMBERG: Yes. In the lay setting, I think we all agree that currently, there is no evidence for general population screening. And so, these devices need to show a clear benefit. One of the criteria for that benefit is going to be high sensitivity and specificity, because dermatologists in a low-risk population presumably is going to do better than table 5, but also, beyond that, you know, we need to show a clear benefit to the laypeople. Sorry to take all your words out of your mouth, Dr. Ballman.

DR. BALLMAN: I want to emphasize that again we are screening in a low-prevalence population. That's naturally going to make the sensitivity be higher, but I don't think you should look at it in terms of what the sensitivity should be, but, as what

was said previously, should show a clear benefit of screening with the device in that population.

DR. HARRIS: Any other comments? Any other comments from you, Dr. Asher or Dr. Chen?

DR BOURELLY: I'm sorry to interrupt. I thought that we had said, for the laypeople, we were going to try to bring them up to the level of a dermatologist. I'm sorry. Maybe I missed that, because simply approving a layperson doesn't seem like enough. Maybe I'm misunderstanding what's been said.

DR. ALAM: I would agree with that as well; I think we have the same sentiment. I'm still a little confused about how we are going to show that it's an improvement for the layperson, because an improvement over nothing is, well, anything would be an improvement over nothing.

DR. BOURELLY: That's my point.

DR. ALAM: I would like it to be… Dr. Ballman is shaking her head, because I'm misunderstanding something.

DR. ROTEMBERG: That's a difference between improvement and benefit. We are not saying – and I believe Dr. Ballman and I are in agreement – that we're not saying patient is improved in their ability to diagnose melanoma. We are saying that, on a population level, if this is deployed, there's going to be a benefit. A decrease in mortality from melanoma would be an example. A very ambitious example. But you could also imagine earlier treatment for psoriasis based on the number of days a patient

is suffering without X treatment by an expert... Some definition of true patient benefit for deployment of the app. I think we agree.

DR. ALAM: Sure. I think if you're going to throw the intermediate outcomes out of the window, like detection and biopsy, you're going to go for true benefit. I think that's great. I think that is the gold standard. That would be a 20-year-long study though, if you're looking at mortality.

DR. ROTEMBERG: It depends on the study and it depends on the disease.

DR. ALAM: That's a concern. Looking for true benefit in a general population, that's going to be an enormous study over a very long period of time. Just a thought.

DR. HARRIS: I can't see everybody. Can we take down this slide? Okay. Dr. Scales.

DR. SKATES: Steven Skates.

DR. HARRIS: Dr. Skates, sorry.

DR. SKATES: The general population would like to get the device performance out to the level of what dermatologists could do in standard practice. So, I think that's a bar that depends on the population of the disease, and so setting up one constant bar for sensitivity or specificity is not the right way to go. But in general, that is the level we would like to get it to.

DR. HARRIS: I actually don't think it's probably accurate to say we are anything. There are differences of opinions. These are all I think part of the opinions that have been expressed. Unfortunately, I don't think that we can give a singular summary opinion, but those opinions I think are consistent.

DR. SKATES: So, Dr. Alam was talking about how long a study would take with her really in point, and I have to agree with him on that. So, that's why I think the FDA is focusing on rather immediate inputs like sensitivity and specificity, and I think we can tweak that so that we can adjust. We can use Dr. Kominsky's benefit-risk ratio and say, that's what we want to get to with the lay public using this device, in getting it to where dermatologist is doing it, on that benefit-risk ratio. That will apply across all populations, whether very low prevalence, high incidence, what have you, intermediate... But if we are going to go for mortality reduction, that's a 20 year study or even other real clinical benefit. But sensitivity and getting benefit risk performance of dermatologist is feasible in the next 3, 4, 5 years.

DR. HARRIS: Dr. Alam.

DR. ALAM: Thank you. I just want to clarify, because I think you looking for consensus. I would be very happy with Dr. Ballman and Dr. Rotemberg's bat; it’s just more ambitious than I was thinking. But I think showing benefit – that is fantastic. If you can actually do that, I’m happy to agree with that if it will help us move closer to consensus.

DR. HARRIS: Dr. Burke.

DR. BURKE: I think it's very disparate. It's a very long study to show benefits of the general public, and I think it's more of a question that this initially be FDA approved is a very strong statement. It needs efficacy and safety and definite benefit. I just wonder if, for now, that FDA approval should be for use by dermatologists and primary

care physicians, and I don't know if we should have it FDA approved as a standalone measurement for the general public.

DR. HARRIS: Okay. Dr. Ballman.

DR. BALLMAN: I think there's benefit beyond mortality, and I was not suggesting mortality. If they can that the detection rates above what the going prevalence is in the population or something like that… but I think it's up to the manufacturers to define what that benefit is and to show it. And I don't think it's fair just to say it has to be dermatologists. It's never going to be, unless you guys want to be out of a job. It's never going to be at the level of a dermatologist, unless we use Dr. Skates's metric. But, again, you've got to think about what the population is that's being served and if there's benefit, because they don't have access at all, I think that might be something meaningful. But I think it's up to the manufacturers to demonstrate benefit, however, they can demonstrate it, and how they define it. I wasn't suggesting it be an improvement in mortality.

DR. HARRIS: Thank you. I have to say that we are getting a bit circular in our discussions, because we have been here before, and not everyone is going to agree on every point. If there's any new points to be made, as I said before, I'm not too sure, Dr. Asher and Dr. Chen, that we're going to better summarize what you've heard. Dr. Rotemberg?

DR. ROTEMBERG: By the way, I agree that I use mortality as an example, but there's many ways to measure benefit. The one thing I would say is I don't think Table 5 is an accurate demonstration of a dermatologist ability in the general population.

Those measures are of dermatologists in patients who came to the dermatologist, potentially for a suspicious lesion. So, when we are thinking about what we expect of standards for the lay population and if we are going to say, we want to say it's a dermatologist level ability, then we need to do a study in the general population of patients who do not go to the dermatologist and compare an in-person dermatology assessment with that of a layperson app. Because I expect the sensitivity of specificity are going to be much higher than that Table in a low-prevalence population.

DR. HARRIS: Any other comments? Any other comments from FDA?

DR. ASHER: None from me. We really appreciate your thoughtful discussion and really considering the issues fully. Thank you.

DR. HARRIS: [4:20 p.m.] I would like to thank the Panel, the FDA and guest presenters, and all the Open Public Hearing speakers for their contributions to today's meeting. This meeting of the General and Plastic Surgery Devices Panel is now adjourned.