UNITED STATES OF AMERICA

FOOD AND DRUG ADMINISTRATION

+++

CENTER FOR DEVICES AND RADIOLOGICAL HEALTH

+++

DIGITAL HEALTH ADVISORY COMMITTEE MEETING

+++

GENERATIVE ARTIFICIAL INTELLIGENCE-ENABLED DIGITAL MENTAL HEALTH MEDICAL DEVICES

+++

November 6, 2025

09:00 a.m. EST

Virtual

Transcript Produced By:

ACSI Language Access at Scale™

ACSI Connects

1025 Connecticut Avenue, NW, Suite 1000, Washington, DC 20036

https://acsiconnects.com/

Participants

Ami Bhatt, M.D.	Chief Innovation Officer (CIO) at the American College of Cardiology, Newton, MA	Chairperson
Melissa Clarkson, Ph.D., MDes, M.A.	Assistant Professor of Biomedical Informatics, University of Kentucky College of Medicine Lexington, KY	Consumer Representative
Ray Dorsey, M.D., M.B.A.	Director, Center for Brain and Environment, Atria Health and Research Institute New York, NY	Voting Member
Joyce Ho, Ph.D.	Associate Professor of Computer Science, Emory University Atlanta, GA	Voting Member
Jessica Jackson, Ph.D.	Vice President of Alliance Development, Mental Health America Alexandria, VA	Voting Member
Thomas Maddox, M.D., SM	Executive Director of the Healthcare Innovation Lab, at BJC HealthCare St. Louis, MO	Voting Member
Chevon Rariy, M.D.	Chief Clinical Innovation Officer, Visana Health Mequon, WI	Voting Member
Wayne K. Goodman, M.D.	D.C. and Iren Ellwood Professor and Chair, Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicines Houston, TX	Temporary Voting Member
Omer Liran, M.D., M.S.H.S.	Co-Director of Virtual Medicine, Virtual Medicine, Cedars-Sinai Medical Center Los Angeles, CA	Temporary Voting Member
Philip Rutherford	Vice President, Substance Use Continuum at The National Council for Mental Wellbeing Washington, DC	Temporary Voting Member
John Torous, M.D.	Staff Psychiatrist, Beth Isreal Deaconess Medical Center and Associate Professor of Psychiatry, Harvard Medical School Boston, MA	Temporary Voting Member
Jessica Paulsen	Acting Deputy Director, Digital Health Center of Excellence, Center for Devices and Radiological Health (CDRH)	U.S. Food and Drug Administration
David McMullen, M.D.	Director, Office of Neurological and Physical Medicine Devices, Office of Product Evaluation & Quality CDRH	U.S. Food and Drug Administration
Letise Williams	Designated Federal Official Office of Strategic Partnerships and Technology Innovation CDRH	U.S. Food and Drug Administration
Michelle Tarver, M.D., Ph.D.	Center Director, CDRH	U.S. Food and Drug Administration
Gioia Guerrieri, D.O., DFAPA, FASAM	Digital Behavioral Health Lead, Digital Health Center of Excellence	U.S. Food and Drug Administration Speaker

	CDRH	
Jay Gupta	Assistant Director, Neurodiagnostic Devices Team, Division of Neurosurgical, Neurointerventional, and Neurodiagnostic Devices (DHT5A), Office of Neurological and Physical Medicine Devices (OHT5), Office of Product Evaluation and Quality (OPEQ) CDRH	U.S. Food and Drug Administration Speaker
Pamela D. Scott	Assistant Director, Neuromodulation Psychiatry Devices, OHT5, OPEQ CDRH	U.S. Food and Drug Administration Speaker
Lieutenant Commander Anthony Becker, M.D.	U.S. Navy	Speaker
Vaile Wright, Ph.D.	Senior Director, Office of Health Care Innovation, American Psychological Association	Speaker
Nicholas Jacobson, Ph.D.	Associate Professor, Dartmouth Geisel School of Medicine	Speaker
Patricia Arean, Ph.D.	University of Washington CREATIV Lab; Former Division Director, National Institute of Mental Health Division of Services and Interventions Research	Speaker
Andrew Trister, M.D., Ph.D.	Chief Medical and Scientific Officer, Verily	Speaker
Brooke Trainum, J.D.	Senior Director, Practice Policy, American Psychiatric Association	Speaker
Bradley Karlin, Ph.D., ABPP, M.S.C.P., M.B.A.	Program Manager, Resilient Systems Office, Advanced Research Projects Agency for Health	Speaker
Caroline Pearson	Executive Director, Peterson Health Technology Institute	Speaker
Nicole Rawson, MBA	Executive Director & Founder Screen Time Clinic®	Open Public Hearing Speaker
Grace Chang	Founder & CEO, Kintsugi	Open Public Hearing Speaker
Jeffrey Cooper	CEO, CooperSoft	Open Public Hearing Speaker
Leo Anthony Celi	Senior Research Scientist & Clinical Research Director of MIT and Associate of Medicine at Harvard Medical School	Open Public Hearing Speaker
Almog Hilel	AI Researcher from MIT and Founder of MIT AI Initiative for Psychology and Psychiatry	Open Public Hearing Speaker
Stephen Schueller, Ph.D.	President, Society for Digital Mental Health	Open Public Hearing Speaker
Aubrey Shick	Liaison, Society for Digital Mental Health	Open Public Hearing Speaker
Miguel Amador	Chief Innovation Officer, Complear	Open Public Hearing Speaker
Andy Molnar		Open Public Hearing Speaker
Bethany R. Russell, PhD	Assistant Professor, Florida Gulf Coast University	Open Public Hearing Speaker

Dr. William Agnew	CMU-Bosch Postdoctoral Fellow Carnegie Mellon University	Open Public Hearing Speaker
Desmond Ong	Assistant Professor of Psychology University of Texas at Austin	Open Public Hearing Speaker
Sneha Dave	Executive Director, Generation Patient	Open Public Hearing Speaker
Isaiah John	Actor & Founder of EKO LABS	Open Public Hearing Speaker
Dr. Derrick Hull	Clinical Psychologist Research & Development Slingshot AI	Open Public Hearing Speaker
Michael Abrams, M.P.H., Ph.D.	Senior Health Researcher at Public Citizen	Open Public Hearing Speaker
Diana Zuckerman, Ph.D.	President, National Center for Health Research	Open Public Hearing Speaker
Rebecca Messing Haigler	Senior Vice President, Healthsperien	Open Public Hearing Speaker
Morgan Bailie	Policy Director, Center for Behavioral Health Healthsperien	Open Public Hearing Speaker
Chase Chick	CEO, Gray Sky AI	Open Public Hearing Speaker

Table of Contents

Digital Health Advisory Committee (DHAC) Call to Order

00:10:15 Dr. Bhatt: Good morning. I would like to call this meeting of the FDA's Digital Health Advisory Committee to order on November 6th, 2025. I'm Dr. Ami Bhatt, Chairperson of this Committee. I am trained in medicine in pediatrics, and then cardiology. I had the opportunity to practice at the Massachusetts General Hospital and as Associate Professor at Harvard Medical School. But now I have since moved to be the Chief Innovation Officer at the American College of Cardiology, where my team's north star is to bring high quality cardiovascular care and prevention to the communities where people live. I am honored to have been selected to chair this Committee.

00:10:51 The Digital Health Advisory Committee advises the FDA leadership by providing and assessing relevant expertise to identify benefits and risks of digital health technologies. This meeting is not focused on broadly available generative AI platforms, but rather GenAI-enabled devices that are intended for the use in diagnosis, cure, mitigation, treatment or prevention of mental health diseases and conditions.

00:11:18 The Advisory Committee is the most formal and public way that the FDA can receive advice from the public on scientific matters. The Digital Health Advisory Committee has members with expertise in various fields of digital health, ranging from artificial intelligence and machine learning to digital therapeutics, remote patient monitoring, software development, user experience, real-world data, real-world evidence, patient-generated health data, cybersecurity, implementation in clinical practice, patient experience with digital health, and we're fortunate to have patient representation and psychiatry with us in full form today.

00:11:57 I do want to take a moment to speak directly to our listeners who are with us today or who may watch in the future. Mental health is a pervasive issue. It's truly an epidemic, and it's likely that it's affected you personally, your family, friends, community, workplace, or it's the focus of your work and your career if you're watching today. And I really want us throughout today to remember that as we have all of these varied discussions, our goal is to learn as much as we can to relieve the suffering of the majority of Americans whose lives have been in some way affected by mental health concerns. So, I'm grateful we've been given this opportunity for the Committee, for our speakers and the listeners joining us today or in the future.

00:12:42 With that, I note for the record that the Voting Members constitute a quorum as required by 21 C.F.R. Part 14. I would also like to add that the Committee members participating in today's meeting have received training in FDA device law and regulations. For today's agenda, the Committee will discuss and provide recommendations on the topic of “Generative AI-Enabled Digital Health Mental-- Digital Mental Health Medical Devices.” The Committee will discuss the benefits, risks to health, and risk mitigations that might be considered for these new digital health devices, including premarket evidence and postmarket monitoring considerations.

00:13:24 Before we begin, I've been asked to set a few ground rules. We have a busy agenda today, so we'll need to move through it efficiently and with purpose. I'm going to ask all presenters to stick to their allotted time. Many are recorded, so that will help us stay on time. For those panelists, if you would like to ask a question, please raise your hand in the Zoom “hand raise” function. I'll address your question and then please remember to put your hand back down so I can keep track of those questions. And to ensure smooth communication, I'll ask people to not unmute and talk over each other as the meeting is being transcribed, for everybody to know, for the official record. Finally, please remember to mute your microphone when you have completed speaking.

00:14:05 Now I'm going to ask our distinguished Committee members and FDA experts who are identified on the meeting roster and attending virtually to introduce themselves. Committee members, I think you all have your video monitors on, so I'll call your name just so we can get through everybody in order, and please give a brief introduction as to yourself and your affiliation. We'll start with Melissa Clarkson.

00:14:29 Dr. Clarkson: Hello, I'm the Consumer Representative for this group. A medical area has deeply affected my family's life over 10 years ago, and so I became a Patient Advocate and I'm very happy to be invited here. And I'm also an Assistant Professor of Biomedical Informatics at the University of Kentucky.

00:14:50 Dr. Bhatt Fantastic. Thank you, Melissa. Ray Dorsey.

00:14:56 Dr. Dorsey: Good morning. My name is Ray Dorsey. I'm a Neurologist at the Atria Health and Research Institute. I've been working on digital health for about 20 years. Thank you.

00:15:06 Dr. Bhatt: Great. Thanks Ray. Ray, your volume's a little soft when you go off mute. If you can just work-- Joyce Ho.

00:15:14 Dr. Ho: Hi, I am Joyce Ho. I am a Computer Scientist and Associate Professor at Emory University. I study artificial intelligence and machine learning for healthcare.

00:15:24 Dr. Bhatt: Amazing. Thank you. Jessica Jackson.

00:15:30 Dr. Jackson: Good morning. I'm Jessica Jackson. I'm a Licensed Psychologist. I'm also the Vice President of Alliance Development at Mental Health America, and I have worked in industry with mental health startups for about the past five years. Glad to be here this morning.

00:15:46 Dr. Bhatt: Tom Maddox.

00:15:48 Dr. Maddox: I'm Tom Maddox. I'm a Cardiologist and also the Clinical Director for Innovation at WashU Medicine in St. Louis.

00:15:55 Dr. Bhatt: Thanks, Tom. Chevon Rariy.

00:15:58 Dr. Rariy: Hello, good morning. I'm Dr. Chevon Rariy. I'm really thrilled to be here today. My career has been built at the intersection between medicine, technology, and innovation. I'm Harvard trained both in medical school and in fellowship, and board certified in internal medicine and endocrinology. I've held leadership positions in digital health in many different systems, from government, to academic, to private, to big tech, and to startup. And I'm currently the Chief Clinical Innovation Officer at Visana, the world's most comprehensive women's health clinic where we leverage technology to amplify deeply empathetic and evidence-based care for our clinicians as they're providing care to our women. So, thrilled to be here. Thank you.

00:16:38 Dr. Bhatt: Wayne Goodman.

00:16:40 Dr. Goodman: I'm Wayne Goodman. I'm a Psychiatrist and I'm Chair of the Department of Psychiatry at Baylor College of Medicine in Houston. My focus is mostly on treatment-resistant psychiatric disorders and I apply various forms of neuromodulation and also some that use sensing from the brain to try to develop biomarkers, and in the pursuit of that I have had some experience with machine learning. Thank you.

00:17:07 Dr. Bhatt: Thank you. Omer Liran.

00:17:11 Dr. Liran: Yeah. Hi, I'm Dr. Omer Liran. Please excuse my voice. I have an unfortunately timed cold. I'm a Psychiatrist at Cedars-Sinai. I co-direct virtual medicine labs where we have been doing a lot of research in artificial intelligence for healthcare.

00:17:27 Dr. Bhatt: Wonderful. Thank you. Philip Rutherford.

00:17:31 Mr. Rutherford: Good morning, everyone. My name is Philip Rutherford. I am the Vice President of the Substance Use Continuum at the National Council for Mental Wellbeing. And when we say mental wellbeing, we mean the intersection of both mental health and substance use challenges. Glad to be here.

00:17:48 Dr. Bhatt: Thank you. John Torous.

00:17:51 Dr. Torous: I'm John Torous. I'm a Dual Board-Certified Psychiatrist and Clinical Informaticist. I'm an Associate Professor at Harvard Medical School and I direct the Division of Digital Psychiatry at Beth Israel Deaconess Medical Center.

00:18:05 Dr. Bhatt: Thanks, John. All right, over to our FDA colleague, Jessica Paulsen.

00:18:10 Ms. Paulsen: Hi, good morning. Jessica Paulsen. I currently lead our Digital Health Center of Excellence in the Center for Devices and Radiological Health.

00:18:17 Dr. Bhatt: Thanks, Jessica. David.

00:18:19 Dr. McMullen: Good morning. I'm David McMullen. I'm the Director for our Office of Neurological and Physical Medicine Devices.

00:18:24 Dr. Bhatt: Thank you. And finally, Letise Williams. Thank you for putting all this together.

00:18:30 Ms. Williams: Good morning. My name is Letise Williams and I am the Designated Federal Official for this Advisory Committee. Thank you.

00:18:37 Dr. Bhatt: Wonderful. Thank you all. Now, Letise Williams, I'll hand it off to you as the Designated Federal Officer for the Digital Health Advisory Committee to make introductory remarks and get us started.

Conflict of Interest Statement

00:18:49 Ms. Williams: Good morning. I will now read FDA's Conflict of Interest Disclosure Statement for the November 6th, 2025, Digital Health Advisory Committee (Particular Matter of General Applicability).

00:19:02 The Food and Drug Administration (FDA) is convening today's meeting of the Digital Health Advisory Committee (DHAC) under the Federal Advisory Committee Act (FACA) of 1972.

The DHAC will discuss and make recommendations on the topic of “Generative Artificial Intelligence-Enabled Digital Mental Health Medical Devices.” Many of these medical devices pose novel risks and, as mental health devices continue to evolve in complexity, regulatory approaches will need to accommodate these novel challenges and opportunities to provide a reasonable assurance of their safety and effectiveness while promoting innovation to support public health.

00:19:40 There is an increasing demand for mental health services in the U.S. and insufficient access to mental healthcare providers. These new devices may be one way to help address this gap in care for people, potentially improving outcomes and access. The Committee will discuss the benefits, risks to health, and risks mitigations that might be considered for these new digital mental health devices, including premarket evidence and postmarket monitoring considerations.

00:20:06 The members of the Committee are either special or regular government employees and are subject to Federal conflict-of-interest laws and regulations. Accordingly, FDA has reviewed the financial interest of the Committee members for compliance with federal ethics and conflict of interest laws. We have screened the members for potential financial conflicts of interest related to today's meeting agenda, both their own interests and those that are imputed to them, including those of their spouses, minor children, and employers. Based on the agenda for today's meeting and all financial interests reported by the Committee members, no conflict-of-interest waivers under 18 U.S.C. §208 have been issued in connection with this meeting.

00:20:49 There are Guest Speakers at today's meeting who will give a presentation to the Committee, answer questions from the Committee, and return to the virtual audience. They will not participate in the Committee deliberations, render advice to FDA or vote. Guest Speakers are non-Government employees or special Government employees participating in a non-official, non-governmental capacity who are not bound by conflict-of-interest laws and regulations and are not formally screened for conflicts of interest. However, in the interest of transparency, FDA asks that Guest Speakers address any personal financial involvement with a firm, product, or other entity affected by the Committee's discussions. Guest Speakers disclose such interest to allow the audience and the Committee to objectively evaluate their presentation.

00:21:37 Today's speakers have reported the following interests in relation to the matter before the Committee. Dr. Jacobson has acknowledged that he has a current interest related to Therabot and Spring Health. The remaining Guest Speakers have not reported any relevant interest in relation to today's meeting. Disclosure of conflicts of interest for Guest Speakers follow applicable federal laws, regulations, and FDA guidance. The Guest Speakers participating in this meeting are presenting the views of their professional societies, not their personal views.

00:22:09 FDA asks that all other participants, including the Open Public Hearing speakers, advise the Committee of any financial relationships that they have with any effective firms, its products, and if known, its direct competitors. We would like to remind the members that if the discussions involve any products or firms not already on the agenda for which an FDA participant has a personal or imputed financial interest, the participant needs to inform the DFO and exclude themselves from the discussion, and their exclusion will be noted for the record. Thank you.

00:22:42 Before I turn the meeting back over to Dr. Bhatt, I would like to make a few general announcements. Transcripts of today's meetings will be available from ACSI phone number 202 599 8456. Press inquiries for today's meeting should be submitted to https://www.hhs.gov/request-for-comment-form/index.html. All written comments received by the deadline indicated in the meeting's Federal Register Notice were provided to the panel and the FDA review team for their review prior to today's meeting. There is an active docket where members of the public can post written comments, the link can be found in the meeting’s Federal Register Notice.

00:23:45 In order to help the transcriptionist identify who is speaking, please be sure to identify yourself each and every time that you speak. Thank you very much. I will now turn the meeting over to the Chair, Dr. Bhatt.

00:24:00 Dr. Bhatt: Thank you, Ms. Williams. I can understand why you didn't do that live with all the dashes. I would've messed that up. Before I ask the FDA to begin with remarks, I want to provide a brief overview of how today's meeting will run.

00:24:12 So, during the morning we'll have opening remarks and presentations from the FDA. Then we'll begin hearing non-FDA presentations at approximately 9:55 a.m. Eastern time. We'll have our first break at 10:15 a.m. Eastern time, and when we return from that morning break, we're going to proceed with the Open Public Hearing for two hours. Once the Open Public Hearing concludes, we'll take a 30-minute lunch break from 12:30 to 1:00. We'll return from lunch at 1:00 Eastern time, and at that point we'll proceed with hearing non-FDA presentations, and then we'll have time for the panel members to ask the speakers questions.

00:24:48 At approximately 2:45, we'll take a 10-minute break. We'll return from break, proceed with the Committee Discussion of the FDA's first set of questions. Following the discussion of the first set of questions, we'll do a 10-minute break. When we return from break, we'll then do the remaining sets of the FDA questions. Following those discussions, I'll give some closing remarks and I promise that we will adjourn for the day at 6:00 p.m. Eastern.

Opening Remarks

00:25:13 So with that, I'm honored to proceed with opening remarks from Dr. Michelle Tarver, Director of the Center for Devices and Radiological Health at the FDA. Dr. Tarver, thank you for being here, and you may now begin your remarks.

00:25:27 Dr. Tarver: Thank you very much, Dr. Bhatt. Good morning, everyone. On behalf of FDA Center for Devices and Radiological Health, I'm extremely pleased to welcome all of you to the second FDA Digital Health Advisory Committee meeting. I want to thank our Chairperson, Dr. Bhatt, and members of the Advisory Committee. We're extremely grateful for the tremendous digital health and healthcare expertise that you all bring to this virtual table. Also, thank you to our speakers for participating to make this a scientifically rich meeting. And thanks to the many researchers, developers, healthcare professionals, and patients who are participating today, we are sincerely appreciative.

00:26:07 At the FDA Center for Devices and Radiological Health, we protect and promote public health by assuring that patients have safe, effective and high-quality medical devices. Innovation of medical devices is central to the work that we do. We support this innovation by fostering regulatory science, by providing industry with predictable, consistent, transparent, and efficient regulatory pathways and by assuring public confidence in the medical devices marketed in the U.S.

00:26:40 The Digital Health Advisory Committee is an important venue to help support our center's mission. Last year, the FDA hosted the Advisory Committee's inaugural meeting, and that meeting focused on the total product lifecycle considerations for generative AI-enabled medical devices. At the meeting last year, the Advisory Committee discussed the importance of robust, real-world evaluation strategies to ensure that AI-enabled medical devices continue to be safe and effective after deployment.

00:27:11 To further explore this recommendation, in September of 2020-- Excuse me, 2025, the FDA announced a Request for Public Comment, seeking information on best practices, methodologies and approaches for measuring and evaluating real-world performance of AI-enabled medical devices. We look forward to sharing a summary of the public comments after the docket closes later this year.

00:27:36 Today's meeting, which has been organized by leaders and experts in our Digital Health Center of Excellence and the Office of Health Technology 5, which is the Neurological and Physical Medicine Devices Office. This meeting provides an opportunity to hear from the Advisory Committee about digital mental health medical devices that incorporate generative AI technologies. The FDA has authorized over 1, 200 AI-enabled medical devices. These innovations help strengthen healthcare and provide diagnosis, cures, and treatment to patients for a range of diseases and conditions. However, to date, there are no generative AI-enabled medical devices that have been authorized for use in mental health conditions.

00:28:20 Now, this topic comes at a time when generative AI is being rapidly developed and adopted in healthcare and may help address critical public health needs. Across the country now, we're experiencing a mental health crisis that is affecting people of all ages. According to the most recent data from the Centers for Disease Control and Prevention, it is estimated that 23% of U.S. adults, which is about 61 and a half million adults, and nearly 1 in 5 children ages 3 to 17, have ever been diagnosed with a mental, emotional, or behavioral health condition.

00:28:56 It is critical that these patients have access to the mental healthcare they need, including the opportunity to be diagnosed and receive appropriate therapy and management. Generative AI-enabled digital mental health medical devices hold significant promise in helping to address this mental health crisis through innovative approaches. It also offers the opportunity for people with mental health conditions who have limited access to healthcare providers, such as those who live in rural communities, can get the much needed care they want and need.

00:29:28 During this meeting, your knowledge, your perspectives, your feedback will provide valuable insights on generative AI-enabled digital mental health medical devices. Through this discourse, the FDA seeks to better understand the benefits, the risks to health, and the risk mitigations that might be considered for these medical devices. Also, we look forward to hearing how these devices can remain safe for the long term in the real world as many of these technologies will continue to evolve and adapt over time. We encourage the panel to have a rich and robust discussion that helps illuminate a path forward for these very important technologies.

00:30:08 Thank you for participating in the Digital Health Advisory Committee. Your input on generative AI-enabled digital mental health medical devices will support the FDA with advancing care through responsible and high-quality digital health innovation. Your input is critical to helping us ensure we meet the mental healthcare needs of all Americans. Thank you for your attention, and I'll turn it over to Dr. Gioia Gue-- Excuse me. Dr. Guerrieri.

FDA Perspective: Generative Artificial Intelligence (GenAI)-Enabled Digital Mental Health Medical Devices

00:30:38 Dr. Guerrieri: Good morning. My name is Gioia Guerrieri and I'm a Psychiatrist and Addiction Medicine Specialist at the Digital Health Center of Excellence within CDRH. This morning, I will present on our leadership and strategic role within CDRH, and some key digital health policy concepts and approaches to consider throughout the day.

00:31:00 Let's begin with why we are here. Americans are facing a mental health crisis. More patients of all ages are being diagnosed with psychiatric conditions and the need for adequate psychiatric care is increasing, while access to mental healthcare is decreasing. At the same time, advances in digital mental health technologies are accelerating. One of those advances is in the realm of generative AI, specifically chatbot therapists, which can provide a variety of potential benefits and risks.

00:31:34 As with all medical devices, FDA applies the statutory standard for assuring safety and effectiveness, considering the benefits, risks, and least burdensome principles, which leads me to highlight one of our key objectives in the Digital Health Center of Excellence. We are actively working towards accelerating the delivery of safe and effective medical device software to market. We do this by supporting the consistent decision-making across CDRH so that there is predictability in digital health technology development and review. Our goal within this work is to remove the friction so that the innovation pipeline is clear and patients have timely access to safe and effective devices.

00:32:23 One of the ways we do this is by taking a risk-based approach to device software regulation. Some software functions are not device functions, meaning certain functions do not meet or are excluded from the definition of a device and are not regulated by FDA. For example, an app that provides daily motivational tips to reduce stress to support a person's general wellness is not a device.

00:32:50 Some software functions meet the definition of a device, but because they pose a low risk to patient safety, FDA intends to exercise enforcement discretion, meaning that FDA does not intend to enforce the device requirements of the Federal Food, Drug, and Cosmetic Act for these software functions. For example, an app used by patients who are diagnosed with an anxiety disorder by providing a "skill of the day" behavioral technique that can be abscessed when they feel anxious is considered low risk and is not the focus of FDA's oversight.

00:33:28 And some software functions meet the definition of a device. These are the focus of FDA's device regulatory oversight. For example, an app that provides computerized behavioral therapy to treat psychiatric disorders is a regulated device and requires premarket review by FDA.

00:33:50 Building on this risk-based approach, in January, we issued a draft guidance on the lifecycle management and marketing submission recommendations for AI-enabled devices. This guidance, which is not current policy because it is draft, outlines FDA's proposed recommendations on the premarket expectations for AI-enabled devices. The draft guidance also proposes a total product lifecycle approach, including recommendations for the design, development, deployment, and maintenance of AI-enabled devices.

00:34:26 I want to highlight some of the proposed draft guidance recommendations for device performance monitoring. This is an important concept as we consider how to strike a balance between premarket and postmarket data to support a device's safety and effectiveness. We recognize that AI-enabled devices are sensitive to changes in input data, which may lead to changes in performance that create a risk to patients over time, but these changes can be difficult to identify. Postmarket performance monitoring may be particularly important for GenAI-enabled devices like the ones we are discussing today. There may be cases where it makes sense to include a performance monitoring plan and the marketing submission to help reduce uncertainty and support FDA's assessment of the benefits and risks of a device.

00:35:21 As we continue to consider AI-enabled device performance in the real world, I want to briefly highlight an Open Request for Public Comment on this topic. Currently, many AI-enabled medical devices are evaluated primarily through retrospective testing or static benchmarks. While these methods help to establish a baseline understanding of medical device performance, they are not designed to predict behavior in dynamic, real-world environments. FDA is seeking information from interested parties and the public on best practices, methodologies and approaches for measuring and evaluating real-world performance of AI-enabled medical devices. This includes approaches to detect, assess, and mitigate performance changes over time to help ensure these medical devices remain safe and effective throughout their lifecycle.

00:36:14 This Request for Public Comment builds on insights from the November 2024 Digital Health Advisory Committee meeting where interested parties discussed real-world evaluation strategies to ensure that AI-enabled medical devices continue to be safe and effective after deployment. The public comment period will end on December 1st, 2025.

00:36:35 Another important topic to be aware of is the AI Predetermined Change Control Plan, also called PCCP. In 2022, Section 515 was added to the Federal Food, Drug, and Cosmetic Act, which clarified our authority to authorize a device with a PCCP for planned future changes to that device which would otherwise require additional submissions to FDA. The final guidance was issued in August of this year. PCCPs helped to increase the pace of medical device innovation by providing a least burdensome mechanism to approaching device modifications while providing continued assurance of safety and effectiveness.

00:37:21 Briefly, a PCCP includes three components. First, there's the description of the modifications where a manufacturer describes the specific plan modifications to their AI-enabled device. For example, retraining the algorithm on additional data sets. Next is the modification protocol. This is where a manufacturer details the methods for developing, validating and implementing the proposed modifications, which may include postmarket monitoring. And third is the impact assessment, where a manufacturer describes the benefits and risks of the modifications individually and collectively and how the risks are adequately mitigated.

00:38:07 When contemplating PCCPs for GenAI-enabled digital mental health medical devices, some questions to consider include: how specific can the modifications be given the nature of the technology? What boundaries or guardrails are needed to define the range of automatic updates allowed for a given device? How will postmarket performance be monitored over time to ensure the device's performance is maintained or improved? How will labeling be updated when modifications are implemented automatically and not manually such that users are adequately informed? And finally, what are the appropriate notification requirements if the device does not function as intended? This list of questions is a starting point and is not exhaustive.

00:38:56 And last, as part of FDA's commitment to support innovation and improve access to care, we collaborated with CMS to improve access to digital mental health medical treatment devices.

CMS created a pathway by which digital behavioral health devices could receive Medicare coverage, and in January, CMS issued a rule finalizing changes for Medicare payments under the Physician Fee Schedule. This marks a step forward in facilitating access to treatments for patients experiencing challenges in coverage for digital mental health medical treatments.

00:39:35 Finally, you can learn more about the Digital Health Center of Excellence and access digital health resources and tools by visiting our website and signing up for our newsletter, or by sending your digital health policy questions to our email inbox. Thank you.

FDA Perspective: Regulatory Considerations for Digital Mental Health Diagnostics

00:39:52 Mr. Gupta: Hello, my name is Jay Gupta and I'm an Assistant Director of the Division of Neurosurgical, Neurointerventional and Neurodiagnostic Devices within the Office of Neurological and Physical Medicine Devices, or OHT5. Before getting into the regulatory considerations, for context, the types of mental health medical devices that we are discussing during today's meeting are reviewed by OHT5, led by Dr. David McMullen, who is one of the FDA experts for today's meeting.

00:40:20 There are two divisions within OHT5 and multiple teams within each division. On this slide is a list of some representative types of devices that are reviewed within each of the teams. Immediately following my presentation, you will hear from my colleague Pamela Scott. Many neurological digital health diagnostics are reviewed within the Neurodiagnostic Devices team, while the Neuromodulation - Psychiatry team reviews digital mental health therapeutics.

00:40:47 Let's start with what we mean by digital mental health diagnostics. These are digital mental health medical devices that are not limited to just stand-alone tests that provide a definitive diagnosis. Instead, these devices may be intended to contribute more broadly to diagnostic-related activities such as patient assessment or evaluation, monitoring, diagnosis or prognosis of mental health conditions. These devices may collect or analyze different types of patient data. Common types of relevant data include things like physiological measurements, which may be collected from sensors, electrodes, wearables, and other instruments. Behavioral data, which may be collected in a variety of ways, but most relevant to today's discussion would be through conversational prompts and interactions with generative AI and large language models or assessments, standardized or otherwise, that are performed by medical professionals. Similarly, relevant assessments could be performed by patients or caregivers.

00:41:44 There are a few examples of authorized digital mental health diagnostics in the United States. Attention Task Performance Recorders are one set of examples. A subset of these devices are prescription devices intended to aid in the assessment of ADHD. These devices do not provide a diagnosis or identify the presence or absence of ADHD or any other condition, but they do provide measures of symptoms that are related to ADHD. The most recent of these device authorizations within the past decade are listed here. Both deliver a computerized continuous performance test and one calculates the device output measures based solely on patient responses, while the other also incorporates assessment of patient motion as captured on video.

00:42:25 Pediatric Autism Spectrum Disorder diagnosis aids are another type of digital mental health diagnostic device that are more specifically intended to aid in the diagnosis of a specific condition. These are prescription devices that analyze patient data and based on this analysis, provide clinicians with a statistical estimate of whether the patient may have an Autism Spectrum Disorder. There are two such authorized devices, with the most recent versions receiving clearance earlier this year. The device regulation was created in 2021 following the review of the Canvas Dx device via a De Novo request.

00:42:58 So, what does that mean? When a new type of device is developed, if there are no existing legally marketed devices in the U.S. that are similar in intended use and technological characteristics, then we may not be able to clear such a device through a determination of substantial equivalence via the premarket notification or 510(k) regulatory pathway. In such cases, for a low- to moderate-risk device where a combination of general and special controls can provide a reasonable assurance of safety and effectiveness, we evaluate the probable benefits and probable risks of this new device type through the De Novo pathway. Hence, it may be appropriate for some novel digital mental health diagnostics to be reviewed through this pathway if they are not sufficiently similar to the types of devices I discussed previously.

00:43:40 This classification process may establish specific premarket and postmarket special controls for the new device type, such as software validation and clinical performance testing that demonstrates the device performs as intended under anticipated conditions of use, design and establishment of a postmarket performance management plan, development and implementation of a risk management plan, and methods of communicating device use, the associated benefits and associated risks.

00:44:08 Types of digital mental health diagnostics being discussed today have the potential to provide additional unique benefits to the currently available methods of assessing mental health in the U.S. Some examples of potential benefits that we may consider in our assessment of these devices include the ability of these technologies to enable or support remote patient monitoring and care; to improve patient access to care and empower patients to better track and manage their mental health; to facilitate earlier or more timely assessments of a patient's status or symptoms; to expand the types of patient data and related information that is available to clinicians to inform their decision-making; and to provide more specialized diagnostic support to healthcare professionals who may not have deep expertise in the assessment or diagnosis of mental health conditions.

00:44:54 Some risks that could be associated with these types of novel diagnostics broadly fall into two categories. One, that the information being provided to users is inaccurate for whatever reason, or two, that the information is interpreted or used incorrectly, which can result in a variety of negative consequences to patients. This could result in the misevaluation of patient's symptoms or status, including a failure to detect worsening symptoms, misdiagnosis or delayed diagnosis of a patient's condition; the potential for missed or delayed identification of a crisis situation, including suicide or mania; and inappropriate or ineffective treatment or management of a patient's symptoms or condition. In the next presentation, Pamela will take a deeper dive into the importance of performing a risk analysis and establishing a risk management plan to mitigate these types of risks associated with GenAI-enabled digital mental health devices.

00:45:47 To assess the device's safety and effectiveness within the context of these benefits and risks, we evaluate any performance testing that was collected to demonstrate the device functions as intended under anticipated conditions of use. We consider whether the device design reflects the specific intended diagnostic capabilities of the technology as discussed earlier, such as whether the device aids in the initial diagnosis of a patient's condition; whether it monitors or tracks symptoms over time in patients with a mental health diagnosis to detect meaningful changes in their condition; or perhaps whether the device provides predictive or prognostic information, such as predicting future substance abuse relapse, or future self-injurious behavior, or even whether the device performs assessments to facilitate adaptive delivery of therapy. Understanding the specific use of the digital mental health diagnostic is a key consideration for ensuring that any software validation and clinical performance testing has been appropriately designed to assess safety and effectiveness.

00:46:42 In addition to understanding the broad type of diagnostic use, understanding the more nuanced or specific contexts of use is another related key consideration. For example, it is important to identify what other assessments or diagnostic information may be available to the user of the digital mental health diagnostic, and how a user should incorporate the information provided by the device with these other sources of information. Related to this is understanding the level and nature of the involvement of a health professional. Is the information being provided directly to patients for interpretation and to inform action? Is the device being used by patients on the order of a medical professional via prescription? Are patients able to access and receive assessments by the device without the involvement of a clinician?

00:47:25 Recognizing that diagnostic algorithm performance may improve by increasing the variety of types of data used as inputs to the algorithm, or by increasing the amount of data available for assessment by the algorithm, we may consider what the basis is for utilization and combination of multiple sources of data by the algorithm to perform its assessment. This also highlights the importance of applying good algorithm development practices to training data collection and the algorithm training process.

00:47:49 On the topic of data and algorithm functionality, other considerations and questions that we think about in this device space include: where did the data come from? How can input data quality during training validation and expected use be ensured? And how do these technologies require specialized or unique labeling to more effectively communicate to users that these are medical devices performing diagnostic assessments, especially when users may not traditionally associate the specific device functions and associated types of data being collected and analyzed as medical in nature?

00:48:25 This concern is relevant to chatbots that interact with users based solely on text prompts, but also to even more sophisticated models that could integrate data from other sources, such as incorporating typing cadence and characteristics, vocal analysis for spoken prompts, or analysis of facial expressions for chatbots with a video interface. We consider whether patients are being adequately informed of how the device functions and are aware of what information of theirs is being used by the device.

00:48:52 Finally, by nature of the design of GenAI-enabled digital mental health diagnostic devices, there is a practical interest towards allowing direct-to-consumer or over-the-counter use as opposed to prescription use. These devices may function more clearly as autonomous diagnostic devices providing an actionable diagnosis, as opposed to being limited to providing adjunctive information intended to aid in diagnosis or clinical interpretation and may reduce or remove involvement of a clinician prior to the patient obtaining therapy.

00:49:20 A further development beyond that is a feature design of a GenAI-enabled digital mental health device that incorporates both diagnostic and therapeutic functionalities and continuously revises the delivered therapeutic content based on real-time assessments of the patient's current symptom state, representing an adaptive closed-loop delivery system. These examples highlight the challenges associated with the considerations that were discussed previously throughout the presentation. Thank you.

FDA Perspective: Regulatory Considerations for Digital Mental Health Therapeutics

00:49:46 Ms. Scott: Good morning. My name is Pamela Scott and I am the Assistant Director for the Neuromodulation Psychiatry Devices team. I will provide the FDA perspective on regulatory considerations for GenAI-enabled digital mental health medical devices designed to treat or augment the treatment of a psychiatric condition and that are under the purview of FDA's oversight.

00:50:11 There are two regulations for digital mental health therapeutics, and this slide highlights the devices that have been authorized under these regulations. To date, these devices have primarily been cleared for use as an adjunct to usual care, including clinician-managed outpatient care, psychotherapy or pharmacological therapy, and none of them are AI-enabled. For these regulations, there are special controls or mitigation measures. These special controls include software verification and validation, hazard analysis, labeling, and clinical performance data.

00:50:50 The clinical performance data should include well-controlled perspective trial designs that support the device's intended population and the indication for use. Typically, these are double-blind, randomized controlled trials that are usually placebo-controlled designs to account for the high placebo response observed in studies of patients with psychiatric disorders. Labeling clearly provides warnings, precautions, contraindications, proper device use, an overview of safety considerations for the device, and an overview of clinical performance data. The existing evidence needs for mental health prescription therapeutics also applies to mental health medical devices with newer technologies. However, additional approaches and special controls may be needed for GenAI-enabled digital mental health medical devices.

00:51:55 The clinical study design and results play a significant role in the evidence used to inform CDRH’s benefit risk decision. Several elements of the clinical study design impact the quality of the evidence factored into CDRH’s benefit risk decision. For example, how the study is designed, what the control is, the specific criteria used to define the eligible study population, the method used to measure the study outcomes, the definition of the study endpoints and the appropriate timeframe to observe important effectiveness and safety outcomes.

00:52:39 When it comes to designing a clinical trial for a GenAI-enabled digital mental health therapeutic, we would like feedback on what reasonable control arms would be. For example, in a potential clinical trial with a chatbot interface, what would be an ideal sham arm? How are factors such as matching, control arm time, and attention to the active treatment to be measured? What are best practices for blinding in such a study? Further, what would outcome measures and study endpoints look like in the context of devices that are trained on various psychotherapy models? GenAI technology presents unique factors in how the device functions, how therapy or treatment is delivered, and the frequency of interaction with the patient. Hence, we want to consider the appropriate timeframe for assessment of the primary endpoints and overall study duration to demonstrate clinical benefit.

00:53:48 In evaluating the safety and effectiveness of medical devices, FDA considers both the benefits and risks of a device. We consider what constitutes an effective treatment for a mental health condition and the level of evidence needed to show clinically significant results for a mental health condition, or symptoms under the conditions of use. For example, will the device be intended to be used as an adjunctive to usual care, or as a standalone therapy? GenAI-enabled digital mental health therapeutics may present different device-specific considerations. For example, the use of one behavioral therapy model or the use of a combination of behavioral therapy models may impact the clinical benefit and to what extent does the behavioral therapy model or combined models delivered need to be understood. Lastly, use of a device with the oversight of a healthcare provider can impact the overall clinical benefit to the patient compared to standalone use of the device, which is another consideration in the benefit determination.

00:55:02 Earlier, Jay talked about risk, so now let's discuss risk within a risk management framework. According to ISO 14971, risk management is a process of identifying hazards and hazardous situations associated with a medical device, estimating and evaluating the associated risk, controlling these risks, and monitoring the effectiveness of the risk control measures. This is a process applied through a device's lifecycle. When making a benefit risk determination for a medical device, factors include the severity, types, number, and rates of harmful events associated with the use of the device, probability of the harmful event, and duration of the harmful event, which are also key elements within a risk analysis.

00:56:02 Now, let's think about some key aspects of a risk management framework we need to consider for GenAI-enabled digital mental health therapeutics. As we think about risk management, several factors to keep in mind are the intended use, such as adjunctive versus standalone, or prescription versus over the counter use of the device, as well as misuse. Let's look at some examples of possible hazards, hazardous situations, harms, and begin to contemplate what appropriate risk controls could be.

00:56:40 This example table highlights some key aspects of risk management, but it's not a complete approach to address all the aspects of ISO 14971. The first example is a missed crisis identification or inadequate response, which could lead to suicide or physical harm. The field is currently developing best practices for crisis escalation plans to mitigate these risks. Another possible hazard is misidentification of medical symptoms resulting from the lack of adequate medical history or the lack of physical examinations and assessments by a qualified clinician, leading to harm due to subsequent worsening of the underlying medical condition. Another example is improper or inadequate clinical content being provided resulting from failure to recognize clinically significant symptoms worsening and leading to worsening of the mental health condition.

00:57:47 In general, FDA takes a total product lifecycle approach to risk, including mitigations, testing and monitoring. Similarly, risk management is a process applied through a device's lifecycle. For GenAI-enabled digital mental health therapeutic devices, we have questions on potential best practices that we would like the Committee to address, such as how can device developers utilize safety by design principles across the device development process, from model design and training, to decisions about risk escalation plans? What verification and validation approaches pulled from the AI and traditional human factors fields are most appropriate for GenAI-enabled medical devices for mental health? What clinical approaches would fit this novel device area? And how can postmarket monitoring be best utilized to mitigate premarket and device update concerns? We also seek your input as we assess how to best use clinical performance data to validate the effectiveness of risk control measures and the acceptability of residual risk.

00:59:04 We look forward to your discussion of critical information and practices needed for a comprehensive approach to the assessment of performance and management of risk throughout the total product lifecycle for GenAI-enabled digital mental health devices.

00:59:23 Dr. Bhatt: Great. Thank you so much to Dr. Tarver for first her remarks and framing our day, and then to Dr. Guerrieri, Mr. Gupta, and Ms. Scott for their presentations. I will remind public observers at this meeting that while the meeting is open for public observation, public attendees may not participate except at the specific request of the Chair.

00:59:47 And now I will proceed with-- Let's get to the hearing for the two non-FDA presentations. Lieutenant Commander Anthony Becker from the United States Navy will present on generative AI in mental health. And then we'll move straight to Dr. Vaile Wright, Senior Director of the Office of Healthcare Innovation at the American Psychological Association, who will discuss the impact of generative AI on digital mental health medical devices and mental health technologies. Each of you will have 10 minutes to present and you may now begin your presentation, starting first with Dr. Becker.

Speaker - Topic 1: Generative AI and Mental Health

01:00:25 LCDR Becker: Good afternoon, everyone. It is an honor to offer the following introductory session for today's Digital Health Advisory Committee. My goal in the next few minutes is to offer a framing for our current moment and the decisions ahead for clinicians, researchers, developers, and regulators when it comes to generative AI in mental health.

01:00:40 In my role as a military psychiatrist, I work on our nation's front lines of mental healthcare but strive to keep abreast of the emerging evidence surrounding these powerful systems. So, today I'll share a perspective primarily from that of a clinician arriving at these crossroads with all of you and wondering how this technology may soon change my day-to-day work. I must mention at the outset that while I'm an employee of the U.S. Navy and have worked for companies that evaluate AI systems and medical applications, my views are my own and do not reflect those from my employers.

01:01:06 And so again, arriving at this juncture, I'll ask you to reflect on the following simple questions. How did we get here? Where could this road take us? And what vulnerabilities lie along the way? To explain how we got here, we must first start by naming the critical technological axis around which this moment revolves. Emerging in the last few decades from an AI winter, a variety of technologies have arisen in the thaw, but none will be as important to today's Committee as that of the large language model. In their basic function, these now ubiquitous models perform a simple task, that of auto aggressive word classification, where in a given input is used to predict the next words in sequence building upon their own output to generate a coherent series of the next word, and the next word, and the next word. And while such a task is simple to describe, it represents the culmination of years of scientific advances that have now arrived at a three-phased process of development.

01:01:55 In this pipeline, models are first exposed to a massive corpus of text in a phase called “pretraining” to model the relationships between their inputs. Next, these models are fine tuned to shape the responses into the “call and response” question and answer format that we are now accustomed to in modern AI chatbots. Finally, most models undergo a phase of training called “Reinforcement Learning Through Human Feedback” or RLHF, where models are asked to generate multiple responses to a given query, and human raters are asked to provide positive or negative ratings of replies, a model then incrementally learns to provide responses that are more likely to be highly rated. The astounding fact, or what some might call the bitter lesson, is that with appropriate scale, scale on the orders of trillions of modeling parameters and astronomical amounts of computational effort, these systems appear to learn parts of speech, sentence and paragraph structure, and even full [Indiscernible 01:02:42] and concepts.

01:02:44 What we found on the other side of these massive efforts is the creation of a talking computer, a talking system, and so the very existence of such systems now asks us: where could this road take us? Well, this requires some means by which to evaluate such systems to gauge our heading. Generally, when evaluating talking systems such as humans, we ask them questions. Our tools to evaluate these models are just that, benchmark collections of questions and problems that measure the capacity of these models to store and apply expert knowledge to solve difficult problems and even engineer software systems themselves. Amazingly, the primary challenge in benchmark production these last few years has been the challenge of lifespan, where in short order the advancing capability of these systems tends to saturate the metrics of their performance. That is, over time, they get every question right.

01:03:30 Focusing in on why the FDA is discussing these systems, the previous prevailing medical knowledge benchmark for these systems was MedQA, a large collection of medical knowledge questions across all areas of medicine. But I use the past tense here because MedQA is now, like many benchmarks before it, saturated by the highest performing models. And in considering the incremental benchmarks that I myself had to undergo to become a physician, such as licensure examination, so too have we known for some time now that these systems outperform human test takers on medical licensure questions, and indeed, quite recently, the top performing models achieve scores as high as a hundred percent. And even in the culminating event for a physician specialist, the challenge of board certification, there is now good evidence that these models' ability to pass this bar of expertise, with their performance in psychiatry standing out amongst other medical specialties.

01:04:18 And so with the ongoing march of benchmark saturation, this year there has been a distinct transition in reorienting the evaluation of these models in the medical domain towards not just query knowledge, but the ability to perform tasks, to have conversations with patients themselves, to elicit information, evaluate risk, and make decisions. For example, in this work where a highly specialized system called the Articulate Medical Intelligence Explorer, or AMIE, was compared in its performance against human primary care physicians in interacting with standardized patient actors. The dramatic finding was that along every rated clinical measure, AMIE performed as equal or greater than her human counterparts. And while this experiment was modeled in the primary care setting, most mental health medication management in the United States occurs just there, with a patient's PCM.

01:05:00 And within the other key capacity of a mental health provider like myself beyond providing medication in delivering psychotherapy, this year has also brought the first published efforts to rigorously evaluate the utility of generative AI systems in performing the task of using thought and behavioral guidance to [Indiscernible 01:05:15]. As in this study published in March, patients expressing symptoms of depression and anxiety underwent a four-week generative AI chatbot intervention and had measurable reductions in their symptoms at both the conclusion of the intervention and their subsequent follow-up when compared to waitlisted controls.

01:05:33 These may be early findings, but pondering where this road leads us, I cannot help but infer that some portion of the tasks that I perform as a psychiatrist might soon be possible to be shared with these learning systems. And when I've shared this thought with my fellow clinicians, a variety of objections or attention to the vulnerabilities of these systems are raised. The most common notion which is brought up is the idea that people can tell these systems are not human or that they will lack a human connection. This, in my opinion, is not a clear barrier to their implementation. To support this, we should note that these models now soundly pass the Turing test, that is when properly configured, they can actually consistently do a better job than humans at convincing other humans that they are human.

01:06:14 Furthermore, on some of the more human metrics of the studies we just discussed, on rated measures such as empathy or therapeutic alliance, these models score at equal or higher ratings than their human comparisons. And considering the revealed preferences of the American population at large, even for non-mental health specific commercial chatbots, the Harvard Business Review published just this year that in 2025, the number one use case for generative AI systems is for therapy or companionship. Whether this is an advisable use case or not, clearly the perceived humanness of these systems does not deter the desire of individuals to apply them in the realm of mental health.

01:06:50 The other commonly identified barrier noted by my peers is the legal or regulatory one, but the question of the appropriateness of legal boundaries on the use of such systems considering the established framework for the implementation of AI-enabled devices in other areas of medicine, and in light of these systems expanding capabilities and in an era of an ongoing shortage of mental health providers, is the question that the crossroads we find ourselves on asks of each of us and that we are here to discuss today. And so, in considering this question, we should also stop and identify what are the actual vulnerabilities of these systems, and for the sake of today's discussion, I wish to identify two.

01:07:23 The first is sycophancy. Sycophancy in this context is a technical term referring to the tendency of GenAI systems to provide responses that are pleasing or aligned with their users at the expense of being true. In other words, by being false. This is now a well-studied phenomenon wherein a user can rebut a factual claim made by an LLM, and depending on the strength of the rebuttal, a model is likely to rush to meet the user in agreement, even if the user's suggestion is opposite to the model's original claim. And this can go in either direction, rushing to meet the user in agreement when the model is wrong and the user is right, thus resulting in an improvement in knowledge accuracy, here shown in gray; but also rushing to meet the user in agreement when the model is right and the user is wrong, here shown in gold.

01:08:03 Think back to our mention of the phase of training called “Reinforcement Learning Through Human Feedback” where models learn incrementally to earn higher user ratings. Well, the best strategy to earn a high rating from any human is to tell them what they want to hear or to tell them what they already believe. This phenomenon of sycophancy is thought to be tied to this RLHF training, and is therefore a facet of these systems that is inherently tied to their method of development. Clearly, this is not how we expect a physician or any mental healthcare provider to act, and is therefore a strange and real vulnerability of these systems.

01:08:35 The second actual vulnerability of these systems to mention is in the area of metacognition. Earlier, we discussed this medical benchmark, MedQA, as being saturated. It seems these models know everything there is when it comes to thinking about medicine, but do they know everything when it comes to thinking about thinking about medicine? In this clever work mutating the MedQA benchmark to include questions that did not always offer the correct answer, posed malformed questions or postulated even fictional organ systems, Griot and colleagues formed a new benchmark called MetaMedQA, which challenged models to know when the answer might be "none of the above" and to rate the confidence of their answers on a scale of one to five. What they observed was a large drop in performance on the benchmark, models struggling to recognize when the answer was not available, or credulously reasoning over non-existent organ systems.

01:09:21 This can largely be considered in the context of what these systems are, which are probabilistic prediction engines that may struggle to model the probability of a category as wide as "none of the above," and that may lack the true semantic grounding to recognize a malformed question. Furthermore, in considering the confidence of their answers, despite low accuracy, most models consistently assigned a maximum confidence of five. Yes, like a surgeon, these models are sometimes wrong, but never in doubt.

01:09:47 Now, this issue of prediction uncertainty is a moving target and even at the time of this publication, certain models, the highest performing, showed the emerging signs of metacognition in showing some correlation between the reported certainty and their level of accuracy of replies. But because mental health in particular is a field where the management of uncertainty and epistemic humility is paramount, this trend raises major concerns for the implementation of such systems in our field.

01:10:11 And so to conclude, along with every stakeholder of patients, providers, and the mental health community at large, we find ourselves standing now with a strange mechanical companion with a set of incredible powers and a number of eccentric vulnerabilities. Whether and how we take the hand of this strange companion are the questions we will discuss for the rest of today's Committee. Thank you.

Speaker - Topic 2: The Impact of GenAI on Digital Mental Health Medical Devices and Mental Health Technologies

01:10:32 Dr. Wright: Hi, I'm Dr. Vaile Wright. I'm a Licensed Psychologist and the Senior Director for the Office of Healthcare Innovation at the American Psychological Association, and I am so excited to be invited to speak to you today about the role of generative AI in mental health technologies, in particular, medical devices. I have no personal or professional disclosures. I do want to say that I'm going to be highlighting some of the products in the space just as exemplars, and they are in no way an endorsement by APA.

01:11:03 I think it's really critical that we set the context, and that is that we are experiencing a mental health crisis in this country, and not just in this country, but actually globally. And I'm sure this doesn't actually surprise anybody on this call, but I still think it needs to be said. Not only are we experiencing high levels of mental illness, but even those with mental illness are unable to seek out the care that they need and not just any care, but in particular evidence-based care, which as we know is the best care that one can get.

01:11:28 This is all within the context as well of a workforce shortage. When we look at data from HRSA around the behavioral health workforce projections through 2037, regardless of what behavioral health provider you're referring to, psychiatrists, psychologists, social workers, there will never be enough to meet the unmet need in this country. Complicating matters even more, we also know that behavioral health providers have reported that they feel disincentivized from taking health insurance, both because of low reimbursement rates, but also because of the administrative burden, which leads consumers and patients at even more peril of receiving care.

01:12:03 Now, a situation this complicated requires multiple solutions. Unfortunately, I don't have the time to go through all these, so I'm just going to talk about the bookends. The first is that we do need to continue to support and assist the current and future workforce to help meet these needs, but because of why we're here today, we're also going to talk about this bottom bullet point, which is the role of technology, telehealth and the use of AI.

01:12:27 So let's start with telehealth. I think everybody, again, is very familiar now, postpandemic. What we know from the data at least is that psychologists prepandemic only about 25-- 20% or so were using telehealth. That increased twelve-fold to 86% during the pandemic. And it's not just that psychologists and behavioral health providers took notice, but venture capitalists and others took notice, and you saw this boom in online therapy platforms, provider networks to try to meet the needs of individuals who needed care. And that included both traditional telehealth but also some asynchronous texting and video messaging. And while telehealth provided an incredibly important continuity of care, including-- In addition to adding new people to care, while it increases access, it doesn't necessarily increase reach. It doesn't create new providers, and the type of care it provides is still a very high level one-to-one touch. And so we need to be looking to other options as well as we think about how we address the mental health crisis.

01:13:24 So a couple of these are what we see in the technology landscape. The first are wellness or direct-to-consumer apps. They're not intended to treat a mental health condition. What we're talking about here are mindfulness apps, maybe symptom trackers. I think they do play a critical role in helping people understand how to engage in better healthy living, but again, they're not replacements for psychotherapy. I think a more promising area are digital therapeutics and particular prescription digital therapeutics, which deliver evidence-based clinically validated psychological treatments. So, think of an entire treatment battery of say, cognitive behavioral therapy for insomnia. They meet classification as software for a medical device, making them under the FDA purview, which we think is really critical because then you're maintaining at least some bar of safety and efficacy for both consumers and providers to feel a sense of confidence that these are products that really can help address the needs of individuals. I think the challenge however, is this gray space in between these two products, right? Where you have products that are maybe marketing themselves one way for intended use, but are actually being used in a different way. And we'll talk a little bit about that.

01:14:29 Third, of course, is the role of generative AI. And I'm not going to go through all the slides points here. I'm going to assume that some people, at least on the screen, are aware of generative AI at this point. But it's clear that since 2022 when OpenAI was launched, ChatGPT, that we've seen just a huge change in the way we think about the role of AI in healthcare. And we've seen huge increases. This has been the technology with the greatest adoption we've ever seen in our lifetime, with research from the Pew Research Center saying that 34% of adults use GPT in 2025 compared to only 18% about a year and a half earlier. So that's almost double.

01:15:07 And when we think about how it's being used, you can't ignore that nearly 50% of individuals in this particular study said that they're using large language models for psychological support within the past year, whether that's to address anxiety, personal advice, or depression. It's clear that people are seeking out this technology resource as a way to address uncertainty in their lives and potentially to even engage in some reassurance seeking about why they're struggling the way that they are. And it's not just adults. We know teens are using these devices as well. At least 70% have used at least one type of gen AI tool in the last year. Parents' awareness really seems to be kind of all over the place.

01:15:45 And again, when we ask teens why they are using this, while the first two topics here are “entertainment and curiosity,” everything else from down until “other” are psychological reasons. They might not use that term. They might not even use “mental health” or “emotional wellbeing,” but they're clearly seeking these out for these purposes. “Advice, someone to talk to, someone who provides no judgment, help me to practice social skills.” These are all mental health and psychological issues.

01:16:11 So, clearly there's a lot of opportunities for AI in healthcare. Here are a variety of them. I think I'm just going to focus on the one that's underlined, which is I do think there's a huge opportunity for AI to improve clinical and patient engagement because we can have the best treatments possible, but if patients aren't engaged either physically, emotionally, mentally, then we're not going to get the outcomes that we want.

01:16:34 And so we're seeing two major healthcare trends around AI. The first are provider-facing products, and these are really your AI scribes, and I'm not going to go into much detail here. You're seeing sort of adoption challenges, maybe some ethical issues. But I think what's important for this group to be thinking about is as these AI scribes become more incorporated into electronic health records, as their functionality continues to evolve into maybe even diagnostic tools, what does that mean for the FDA and for our thought process around how we regulate medical devices?

01:17:01 The second of course are consumer-facing AI chatbots, and these really range the gamut. They're not all quite the same. I think you're going to see FDA-regulated mental health chatbots in the future, and I look forward to that. I think you're going to see prescription digital therapeutics incorporate generative AI as a means of engagement. Unfortunately, that's not really what's on the market currently. What's on the market currently are direct-to-consumer mental health chatbots that may use generative AI or closed loop AI. And these direct-to-consumer generalist entertainment chatbots, which again, as we noted earlier, are clearly being used to address mental health needs even though that was not the intent for their development.

01:17:38 And because of that, you see considerable challenges or threats of the use of AI in mental healthcare. Again, there are a variety of these. In my limited time, I'm not going to go over them, I'm just going to highlight the last one, which is really the impact on vulnerable populations, whether that be adolescents, those who are not as developmentally advanced as others, or those really suffering from some sort of mental health illness or diagnosis. And the influence that these AIs can have on them could have very unintended consequences, some of which we outlined in our recent Health Advisory on AI in Adolescent Wellbeing.

01:18:12 And I think because of this lack of regulatory space, we are seeing challenges to scalability. So, while the opportunities and the promise that we've mentioned is huge and really critical, we're not seeing it, and we're not seeing it because we don't have legislative and regulatory bodies that can handle innovation in the way that we're seeing in this space. And as a result, you have a lack of reimbursement coverage, you have data privacy, health infrastructure issues, and you have a lack of trust and understanding by both the public, by providers, because they just don't know what they can trust. And the lack of transparency makes it really hard to understand exactly what these products can do for you or not for you.

01:18:51 And so as an ending slide for today, we really urge, as APA, that policy makers think about how we take the regulatory frameworks that are in place and really use them to our advantage to address the realities of AI in mental healthcare, because this is a reality we've never had to face before. I think, at a minimum, we need to require that developers and others follow the existing FDA transparency guidelines as a baseline expectation to address gaps in transparency. I think it's clear that we need to modernize regulations to address the critical discrepancy between the stated intent of these technologies and their actual use.

01:19:26 I think we can strengthen requirements around premarket evidence, postmarket monitoring for all digital tools in mental health; establish clear enforceable standards for all these digital tools, whether they are, again, the direct-to-consumer apps, these entertainment apps, the digital therapeutics, etcetera; address gaps in FDA oversight via a new or interagency approach to assess safety and validity of generative AI in mental health; and finally, most critically from my perspective, I think we need to establish a public repository of products that have received FDA clearance, those that are prescription digital therapeutics, so that providers and consumers know exactly what products have been deemed safe and effective as well as those operating under enforcement discretion.

01:20:07 It's a very unclear space, and so we need clear guidance for both public and developers about what it means and does not mean to fall under enforcement discretion and who makes that determination. I think this would really go a long way to improving transparency and helping everybody understand exactly what the space looks like. And with that, I want to thank you so much for allowing me to present to this group today. I look forward to our continued discussion.

01:20:34 Dr. Bhatt: Fantastic. What a morning session. Thank you to Lieutenant Commander Anthony Becker and to Dr. Vaile Wright for their comments as well. Just a few quick thoughts. As a cardiologist, I spend a lot of time thinking about remote patient monitoring coming from a medical high point and slowly getting more and more accessible. And then I think about wearables and how they're becoming more and more medical and I always feel there's a gap there. And so I don't know if I'm reassured or concerned, but I think it's a shared problem across many different fields that there is a gap there. And you've heard it mentioned this morning, and we need to think about how to safely close that gap for the sake of patients.

01:21:12 I did want to let people know that the slides that are presented will be shared afterwards. When I was younger, I would write a paper sometimes in 10 or 11 point instead of 12 point, so I could fit it all in. And I think we played one of the videos at like 1.25. So, if it was going really fast-- And they look like great slides, they will be available to the public later on the meeting webpage.

01:21:34 So, we'll now take a 15-minute break. Committee members, please do not discuss the meeting topic during break. We're really supposed to be listening and learning just from the speakers and discussing only openly and publicly, transparently today. And the meeting will then reconvene at 10:30 a.m. At that time, I'll give a short summary of what we covered this morning for any who are joining late and then we'll proceed with the Open Public Hearing. Thank you, and see you back at 10:30.

Open Public Hearing

01:40:07 Dr. Bhatt: I'd like to welcome everyone back from the break. It's now 10:30 a.m. and we will proceed with a quick review of what we heard this morning for those who may just be joining us, like those of you on the West Coast, good morning. And then we'll have the Open Public Hearing portion of the meeting.

01:40:22 We began this day with Dr. Tarver reminding us that although the FDA has authorized over 1,200 devices in this realm, we have yet to see a generative AI-enabled digital mental health device. She also reminded us further of the pervasive nature of mental health conditions, our job today, and of the one out of five children that are currently affected, a real call to action.

01:40:45 Gioia Guerrieri then led us through a risk-based approach to device software regulation, highlighting that wellness support is not necessarily considered a regulated device. There's middle ground currently at the FDA of intention to exercise enforcement discretion for devices that might support coping skills, for example, and not need regulation, and then the regulatory pathway to treatment or medical treatment adjuncts. She then reminded us of the FDA seeking public comments by December 1st, for those who are listening and interested, about postmarketing performance metrics, thinking about real-world performance, and what I found especially interesting is thinking about human-AI interaction, which interestingly came up later in our morning meeting as well.

01:41:28 Jay Gupta worked us through some of the examples of authorized devices in the ADHD symptom space. Talked about assessing likelihood of Autism Spectrum Disorder in a pediatric patient. Of note, this ability of AI to help us find patients earlier is not limited to mental health. It's really burgeoning across medical specialties. And then lastly, he reminded us of the risks not only of inaccurate information but also inaccurate interpretation.

01:41:55 Finally, Pamela Scott rounded out our FDA presentations by asking us really directly: What is an ideal control arm for a chatbot? How do you blind such a study if you are using a chatbot? And how do we get therapies to patients in a timely manner remembering that many of the traditional study designs like RCTs are long and they're not inexpensive? Lastly, she addressed risk mitigation and the very real need in mental health devices for a reliable crisis escalation plan to be addressed.

01:42:26 We then had our two non-FDA speakers, Lieutenant Commander Becker had us reverse this concern that AI is not perceived as human. In fact, he showed us the humanness of AI and how we often see it as human. He went on to point out true vulnerabilities of the tendency of generative AI to please the user over ensuring accuracy and tell people what they want to hear, most certainly not what we want our doctor to do. As we mentioned last year, we used to talk about the inability of generative AI when this Committee first met to understand nuance, context at times, and edge cases. And now Dr. Becker actually revealed studies about concerns with metacognition. So, he's now pushing us not to distrust AI, but to be vigilant, to acknowledge, and then account for these eccentric vulnerabilities in our planning.

01:43:15 And then finally, Dr. Wright put a number on this: 11 billion dollar market in 2025 just for online therapy platforms and provider networks, but she reminded us that increased access doesn't mean increased reach of actual medical care. So, not only are there FDA-regulated consumer-facing chatbots, there's direct-to-consumer chatbots right now, and she pointed out the entertainment and companionship bots, and really talked about the at-risk vulnerable populations. We have elderly Americans suffering from loneliness throughout the country, as we know, we have teens looking for someone who's available to talk when they want to talk without judgment. As a mom of a teen, I understand that. And so she really challenged us to protect vulnerable populations and I think that's so important.

01:44:02 I'll end with just a framework in my head as we move into the Open Comment section, which is: I think we talked about a variety of ways devices can interact, all the way from, in between, talking about device autonomy, those autonomous device systems and what needs to happen to get there, higher risk area, then assistance to clinical care, those that are software as a medical device that are used in that combination with care that's happening today, evidence-based treatment, and then adjacent or supportive, some of which are in the wellness field, some of which fit this kind of FDA middle-ground criteria, and some of which are now direct-to-consumer apps, and we're still struggling with how to do that.

01:44:42 I think the part we didn't mention that I want to bring up as we move forward is: we talked about autonomy, we talked about assistance, we talked about adjacent, their supportive phases, and we didn't talk about agency when it comes to education of our patients as much. I think, moving forward, one other thing you'll hear about today is we can't put things in the hands of patients unless we are willing to educate them at the level that they can understand. And so I think that part about patient agency, and making sure that they understand and are educated if these kinds of therapies are going to be available to them, is really important, and we'll hear more about that as we move forward.

01:45:18 So, now we will proceed with the Open Public Hearing portion of the meeting. Public attendees are given an opportunity to address the Committee and present data, information, or views relevant to the meeting agenda they have approached us previously and are therefore joining us today. And Ms. Williams, if you could read the Open Public Hearing disclosure process statement, that would be great.

01:45:42 Ms. Williams: Both the Food and Drug Administration (FDA) and the public believe in a transparent process for information gathering and decision-making. To ensure such transparency at the Open Public Hearing 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's 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 relationships at the beginning of your statement, it will not preclude you from speaking. Thank you.

01:46:58 Dr. Bhatt: Great. The FDA and this panel place great importance in the Open Public Hearing process. The insights and comments provided can help the Agency and this panel and their consideration of the issues before them. We ask that each presenter speak clearly to allow the transcriptionist to provide accurate transcription of the proceedings of this meeting, and the panel appreciates that each speaker remains cognizant of their speaking time.

01:47:22 The FDA received 16 formal requests to address this Committee. Speakers who submitted their request to speak by the deadline indicated in the meeting's Federal Register Notice are being given that opportunity to speak. We'll now begin. I'll open the Public Hearing with a presentation from Nicole Rawson, Executive Director of Screen Time Clinic.

01:47:43 Ms. Rawson:Good afternoon, members of the FDA Advisory Committee. My name is Nicole Rawson and I'm the Founder and Executive Director of Screen Time Clinic. I started this organization nearly eight years ago after witnessing firsthand the devastating effects of toxic technology on children. While I was teaching middle school, I saw bright curious students slowly change. They couldn't focus, they were sleepy, distracted, and increasingly defiant. They admitted, sometimes proudly, that their parents had no idea how much time they were spending online or what they were doing there. As an educator, I was alarmed, and I knew I had to do something, so I quit my job and started Screen Time Clinic. But as a single mom, I was also living that same nightmare at home trying to figure out what best to do to help my own child.

01:48:28 My son, who is on the Autism Spectrum, became deeply addicted to video games. What started as a joyful play in Roblox and Minecraft turned into an all consuming obsession, his only source of motivation and comfort. Despite all my love and education, I couldn't pull him out. That's how powerful these digital hooks are. Now, as an adult, he still struggles with the consequences of a childhood shaped more by screens than real-world connections. Ten years ago, none of us knew how damaging this would be. Like most parents, I simply wanted my child to be happy and tech savvy. There were no warnings, no research summaries, no professionals to guide us, and even today, those resources are still rare.

01:49:09 That's why I founded Screen Time Clinic to help families recover from digital addiction and prevent others from going down that same road. And that's why I'm here today to urge this Committee to take decisive action. We need widespread federally supported warnings for parents and caregivers about the dangers of video games, short format scrolling, and even hyperstimulating cartoons. These aren't harmless pastimes. They're engineered for compulsion.

The dopamine dependence, overstimulation and constant reward-seeking, and the content itself, beyond the predators and explicit material, is neurologically damaging to the developing brain. We would never say, “One drink is fine” for a 16-year-old. Yet we normalize hours of digital consumption for children, even though evidence now shows measurable harm to brain development.

01:49:58 The recommendations must be clear. No recreational screen time for minors under 18. Through my work with thousands of parents, I've seen the pain and the hope. Every parent says, “I wish someone had warned us,” But I've also seen recovery, children regaining eye contact, focus, creativity, empathy, and coming out of mental health crises within weeks of a digital detox. That proves this is preventable if we act now.

01:50:28 First, this cannot remain a so-called “parental choice.” As a former teacher, I know that many parents, especially those working multiple jobs or struggling financially, simply don't have the time or knowledge to monitor every online minute. Calling it choice is an excuse that leaves low-income families unprotected and gives big tech free reign. Child safety should never depend on income or tech proficiency. We don't call child labor laws or tobacco restrictions “parental choice.” We enact them to protect all children, just as we decided that pornography, alcohol and cigarettes are unfit for minors, not because kids are weak, but because these substances permanently alter brain development and values. We must add toxic technology to that list. Even if some parents want the choice or disagree. Parents mean well, but good intentions cannot compete with billion dollar algorithms designed to hijack the brain's reward system.

01:51:27 Second, I often hear “We need more research.” With respect, that is not true. The research is already overwhelming. We can't wait another decade. While an entire generation is neurologically rewired for distraction, anxiety, and isolation. Other countries are acting:

China, France, the UK. While here in America, we're still debating whether it's even a problem and what to do about it.

01:51:51 Finally, the recommendation must be clear, and absolute, and easy to follow. The current "one to two hours a day" for different age groups guideline is completely meaningless. Is it one hour a day? Is it two hours a day? Is this for kids with autism? Is this for neurologically diverse children? Who does this apply to and what age? It's a hundred percent difference one to two hours a day. We cannot afford to make these types of recommendations when we are putting children in harm's way. Ambiguity only benefits big tech. It pits parents against each other, parents against their own kids, even spouses against other spouses, and parents against teachers and therapists. The list goes on and on, dividing our society without clear, simple guidance. Teachers feel powerless. Therapists hesitate to even confront it, even with all their extra resources to battle mental health.

01:52:44 Our nation sinks deeper into a mental health crisis that no pill or app can fix without clear guidance on toxic tech. We need the FDA to lead with clarity and courage. My recommendation is simple: no addictive gaming, no social media, no short format scrolling for minors under 18. Mandatory warnings, labeling, and nationwide education campaigns just like we did with tobacco and alcohol. Recognition that toxic tech is a public health emergency. The future of our children and of this country depend on your leadership. If we fail to act quickly, we are complicit in a mass neurological experiment run by profit-driven corporations. But if we act decisively now, right now, we will be remembered as the generation that protected childhood itself. Thank you for your time, and your courage, and your commitment to America's children.

01:53:43 Ms. Chang: Hi, my name is Grace Chang, founder CEO of Kintsugi. We are a Berkeley-based startup working at the intersection of mental health and voice biomarkers. I'm excited to share some of our latest research and work that we've done with the FDA.

01:54:02 At Kintsugi, we are developing foundational models for human insights using voice biomarkers and AI to be able to identify the acuity of depression and anxiety through natural conversations. I think everybody knows that depression and anxiety are quite prevalent, yet it still surprises me that it is only screened in about 4% of all U.S. primary care visits. And as you can imagine, it's really difficult to make improvements in areas where there's no measurement. There are paper versions of PHQ-9, which is a patient health questionnaire of nine questions that an individual may fill out, but that contributes to likely why it is not something that happens in a 15 to 17 minute primary care visit.

01:54:44 Now, what if voice could be that missing sign? The human voice does reflect emotional and neurological changes. There are subtle shifts in tone, energy, and rhythm that can be detected from a signal processing and deep learning perspective. We have instituted large clinical trials, and in an upcoming one that is validated in a prospective, double-blind fashion, this pivotal study is being submitted to our FDA De Novo clearance towards that. And we are comparing our AI model against clinicians who conduct a one and a half to two hour SCID-5 interview for every participant in this large study.

01:55:19 And Kintsugi is instrumented into where the patients are talking to their practitioners, whether that's a telehealth visit, a call center call or an ambient visit. A patient will have a normal set of conversations and we run quietly in the background and return a real time signal, which is a gentle cue for the clinician to explore further. It is frictionless, language-agnostic and invisible to the patient.

01:55:40 We have a very interesting case study to share with this group. Andrea, who is an LMFT and owner of a mid-size behavioral health clinic and platform, has instrumented Kintsugi in an IRB context across a few hundred patient visits, over time and over different practitioners. And we found two very interesting insights from the patient and the clinical perspective. One, we had an example of a veteran come in for PTSD, and we identified at that first visit that he was also struggling with depression. And so it doesn't take until the fifth or sixth visit to understand that this is somebody that needs that aspect of care to be taken upon. And with tools that are able to perfectly sort of capture that on visit one, we help shorten that cycle.

01:56:24 Now, across longitudinally and with different practitioners, you have a quality control and an ability to surface how long it takes before each patient gets better, by practitioner. And that becomes helpful for Andrea as an owner because she can then relay information to the patient. In their case, they're a little bit more severe than your typical types of patients who-- The drop-off rate on that first visit was something close to 45%. And by being able to share with those new patients that therapist B has a three- to four-visit average before patients really start to feel better, enables and brings about a lot of confidence and expectation, to which a patient has the ability to imagine him or herself feeling better in that timeframe. So, that's very helpful. It reduced churn about 50% and clinicians have been very happy not having to do the homework and chase after patients for those paper tests.

01:57:19 Startups have a lot to contend with. The FDA is a very capital-intensive lift, often spanning multiple years and multiple millions of dollars of commitment from public and private funding. But yet, even after clearance, payment is not always a clear pathway. Traditionally, we've looked at clinical hours as a way of reimbursement, but given that AI gives so much more leverage and efficacy to the system, that hasn't quite been worked out yet. We do believe that, with early FDA and CMS coordination, innovation in healthcare can be rewarded instead of somewhat punished in the state. We do think that AI is this sort of critical X-factor and leverage that can produce thousands-fold productivity gain in a system that is truly short on specialists and clinicians.

01:58:02 Now, as we look back over the last several years, we've worked over four pre-submissions for De Novo submission, and the process has been thoughtful and rigorous, but long. Cycles of a hundred days or more are really difficult for startups because of the pace of technology. And from first prototype to live deployments, this journey has taken about seven years, and in that time, depression rates and suicide deaths have continued to rise pre- and postpandemic. The science is ready, and the system needs to adapt to move faster.

01:58:27 And so I'll leave this with you. Don't let the pursuit of perfection delay the delivery of progress. Even if our AI model was perfect against the gold standard, in this case, the SCID-5, the SCID-5 is a little bit shaky as ground truth relative to other areas outside of mental health. The Kappa is 0.28, and if this model were perfect, it would still only perform at 75% sensitivity and specificity, with a 95% confidence interval, leading many to say that it is impossible. But today's mental health screening does happen in less than 4% of visits. And if we had an indeterminate rate that basically says “there's an area where we have low confidence,” we prefer that you screen with the paper test, of 35% indeterminate rate, this yield still is 65%, or a 15x increase in real usable screening impact over the status quo, and this is a huge public health win.

01:59:19 So to recap, objective screening, triage and monitoring where AI can lower costs and improve care by making triage measurable and fast, is a win. Shorter feedback cycles can also be very advantageous for innovation because at the end of the day, all this cost gets pushed to patients, which are us, you and I, at the end of the day. And the third is really: let's not block better to be perfect. Innovation can and will die on the vine, and it's not typically because of the science but because of the process. So, I thank you for the time and invite you to ask any questions. Thank you.

01:59:53 Mr. Cooper: Good morning. My name is Jeffrey Cooper and I'm the CEO of CooperSoft, a digital health company helping health systems and payers manage chronic disease more effectively through technology. I'm here today to speak not just as a CEO, but as a representative of many small businesses and startups working to bring safe, impactful digital tools for medical innovation to the patients and providers who need it the most.

2:00:19 At CooperSoft, we've seen firsthand how digital health tools can change lives, improving outcomes, reducing costs, and expanding access to care. But we've also experienced significant regulatory hurdles that small-resource constrained companies face. Navigating the FDA process can be time consuming, unpredictable and costly for many startups. Those challenges can become existential.

02:00:45 The FDA has a crucial role to play in supporting innovation, not by lowering standards, but by making the regulatory pathway more efficient, predictable, and accessible for companies like ours. I want to highlight a few ways this can be done. Regulatory efficiency and flexibility. We need regulatory frameworks that are built for the pace and the nature of digital health innovation. Programs like De Novo and Breakthrough Devices programs are strong starting points, but they should be expanded and adapted for small firms. This includes prioritized reviews, more consistent FDA engagement.

02:01:25 We also encourage the FDA to explore firm-based oversight models, such as the earlier digital health software Pre-Cert Pilot, which evaluated an organization's quality systems rather than requiring a full review of each product's iteration. That kind of flexibility would help companies like CooperSoft deliver iterative improvements to chronic disease platforms more quickly without compromising safety or efficacy.

02:01:56 Tools like eSTAR, which guide applicants step-by-step through electronic submissions are another game changer. Mandating their use and continually improving them would help small teams avoid delays in costly mistakes in the submission process. And rather than forcing startups to develop new validation methods from scratch, the FDA would explicitly allow reliance on recognized international standards such as ISO for areas like risk management, quality systems, and cybersecurity. This creates consistency and reduces unnecessary reinvention, reducing financial resources and burdens.

02:02:40 Cost is one of the most significant barriers for companies like ours. Regulatory fees, even when justified, can stall promising products before they reach the market. The Small Business Determination Program is essential. We urge FDA to expand the scope, not just for submission as a 510(k)s or De Novo and PMAs, but also for Annual Establishment Registration Fees. For early-stage companies, every dollar counts.

02:03:10 We also need clear, common sense guidance on the use of commercial off-the-shelf software. Most digital health companies rely on secure widely used cloud services and operating systems requiring the same level of documentation as custom built codes, a slow development unnecessarily without adding meaningful safety benefits. And finally, programs like the SBIR grant can help companies like CooperSoft fund critical work in the regulatory science, increasing awareness and support for this program would open doors for more startups to pursue FDA [Indiscernible 02:03:48] pathways.

02:03:50 Guidance and education, my third point. When you're a small team building something new, what you need most is clarity. A centralized interactive tool, like a digital health policy navigator or regulatory accelerator, would be invaluable. It could help teams understand which regulatory pathways apply to them, what guidance they need to follow. The Q-Submission process is also critical. These early interactions with FDA reviewers help companies avoid costly due detours. We encourage the FDA to prioritize and streamline pre-submission for startups and small firms, enabling us to receive meaningful feedback before investing heavily in formal submissions or clinical trials.

02:04:36 We've also found value in programs like DICE, a Division of Industry and Consumer Education. Expanding its reach with more staff, tailored webinars, and rapid responses would be enormous for first time submitters. And lately we need clear, product-specific guidance on clinical evidence requirements, especially for low-risk digital tools. With the uncertainty around trial expectations can push startups to over design studies out of fear. Clearly here we could free up resources to focus on real innovation and safety.

02:05:13 And closing, the FDA has made significant progress in engaging the health community, but there's more to do. Startups are not asking for shortcuts. We're asking for clarity, consistency, and a level playing field so that good ideas and safe technologies don't get buried under regulatory complexity. By tailoring processes to the realities of small businesses, the FDA can unlock a wave of innovation that improves lives and expands access to care, especially in areas like chronic disease where most innovation tools are urgently needed. Thank you for the opportunity to share this perspective.

02:05:54 Mr. Anthony Celi: Hi, my name is Leo Anthony Celi, a Senior Research Scientist and Clinical Research Director from MIT and Associate Professor of Medicine at Harvard Medical School.

02:06:05 Mr. Hilel: Hello, everyone. My name is Almog Hilel. I'm an AI Researcher from MIT and the Founder of the MIT AI Initiative for Psychology and Psychiatry.

02:06:13 Mr. Anthony Celi: Today we propose new strategies for postmarketing, monitoring and evaluations of new digital mental health devices.

02:06:21 Mr. Hilel: I want to start with a bold statement. The way we currently think about postmarket monitoring for medical products, it might be broken when it comes to mental health AI. Here is why. Traditional postmarket monitoring was designed for the 20th century. You approve a drug, release it, monitor for problems, maybe recall it if things go wrong. That model worked because products were relatively static. A pill approved in 2020 is the same pill in 2025, but mental health chatbots, they're updated constantly. We're talking about software that changes weekly, sometimes daily. The app your patient uses today is fundamentally different from the one they will use next month. So, asking “Is it safe?” becomes a moving target.

02:07:06 Mr. Anthony Celi: And right now, we're potentially wasting enormous amounts of capital, talent and time trying to fit a 21st century technology into a 20th century regulatory framework.

02:07:17 Mr. Hilel: So what's the alternative? We need to shift from point-in-time approval to lifecycle oversight. Now, I'm not suggesting the FDA needs to personally monitor every line of code throughout the entire AI pipeline. That's not realistic, and frankly, it's not the right approach. Instead, the FDA's role should be to govern which actors participate in oversight at each stage.

02:07:40 Mr. Anthony Celi: Think of it like this. Who should be watching during development? Who is responsible during deployment? Who monitors after launch? The FDA can establish the governance framework that ensures the right stakeholders are at the table all the time. Digital mental health devices are not the traditional software as a medical device. These are living systems that require living oversight. And here's a critical part. This oversight can’t just be reactive. It needs to be woven into the fabric of how these tools are built, deployed, and maintained.

02:08:13 Mr. Hilel: Here's where things get interesting. Look at how the broader AI community already operates. When OpenAI or Google releases a new model, what happens? Independent researchers immediately start benchmarking it. They publish public scoreboards, they create evaluations, and you know what? This public evaluation process actually drives iterative improvement.

02:08:36 Mr. Anthony Celi: Developers see where their models fall short and respond. It's a continuous feedback loop where the community has a real say in defining what makes a good product.

02:08:45 Mr. Hilel: But here is the crucial difference for mental health applications. In general AI, anyone can evaluate models through accessible APIs. In digital mental health AI, you can't. There is no evaluation interface, and vendors' terms of service forbid automated high-volume testing. These barriers prevent the community from extending proven AI safety practices to mental health tools. Some may argue it's enough to evaluate models using historical user data, meaning retroactively testing, health records, device logs, and patient reported outcomes. Those data sets matter, but they are only part of the safety picture. They can tell us what has happened, but not necessarily what might happen.

02:09:28 Mr. Anthony Celi: Therefore, we must proactively monitor and evaluate these systems in controlled environments and test them with simulated interactions, counterfactuals that may not initially appear in prior interaction data sets.

02:09:41 Mr. Hilel: For example, we should be able to answer: how would an interaction unfold if a new user presented with early signs of schizophrenia? How would the system react to in that atypical crisis language? This approach aligns with leading guidance that emphasizes structured retaining and scenario best testing for AI systems, including the NIST AI Risk Management Framework.

02:10:04 Mr. Anthony Celi: What would make this possible? Vendors need to provide an evaluation interface made transparent via an API so that developers, clinicians, patients, and their carers could co-design safety benchmarks, which can serve as a firm foundation for FDA regulation down the line. This brings me to what we consider non-negotiable: demonstrated community engagement. If a developer cannot show how they've meaningfully engaged the clinical and patient communities, that should be an automatic red flag. When I review research papers, if there's no engagement with the people the research aims to help, it's an automatic rejection. The same standard should apply here.

02:10:42 We need patients and mental health professionals as co-designers of these tools to help define what good looks like, not as beta testers at the end, but as partners from day one. They should also co-design the evaluation benchmarks. What does helpful support sound like? What are the warning signs of harm? What does appropriate escalation look like? These questions cannot be answered in a lab or a boardroom. They must be answered in partnership with the community. So, here's what this looks like in practice: an ongoing feedback loop throughout the entire life cycle.

02:11:13 Mr. Hilel: And here is the beautiful part. Because these tools update continuously anyways, this feedback can actually be incorporated in near real time. We're not waiting years for the next version.

02:11:23 Mr. Anthony Celi: We're talking about responsive, iterative improvement informed by the people who matter most.

02:11:29 Mr. Hilel: This is fundamentally different from traditional postmarket monitoring. It's not about waiting to see what goes wrong. It's about building safety, efficacy, and genuine helpfulness in every stage of the process.

02:11:41 Mr. Anthony Celi: The technology is evolving rapidly. Our oversight needs to evolve just as fast, and the only way to do that effectively is to put the community at the heart of the process. Thank you. We're happy to discuss further.

02:11:52 Mr. Hilel: Thank you so much.

02:11:54 Dr. Schueller: Thank you for the opportunity to speak today. My name is Stephen Schueller. I'm a Professor of Psychology and Informatics at the University of California, Irvine, a Licensed Clinical Psychologist and a Mental Health Services Researcher. Over the last 20 years, I've developed, evaluated and deployed technology to support people's mental health and wellbeing. I do want to disclose that I serve on the Scientific Advisory Board for Headspace, for which I received compensation and received consulting payments from Boehringer Ingelheim and Otsuka Pharmaceuticals.

02:12:21 I'm speaking today, however, in my role as the President of the Society for Digital Mental Health, a nonprofit organization with membership from researchers, healthcare systems, payers, digital mental health companies, and patient advocates. We are dedicated to advancing the safe, effective use of digital mental health treatments, ensuring accessibility to all in need.

02:12:40 First, I want to commend the Committee and the FDA for convening this important discussion on the role of generative artificial intelligence in digital mental health devices. The speed of innovation in the space is extraordinary, and we recognize both the tremendous promise and the critical responsibility to come with deploying generative AI technologies and the service of supporting those of mental health disorders. Too many people who experience mental health disorders fail to receive treatment, and even those who do, often don't fully benefit. We need effective, safe and scalable interventions that meet the needs and interests of consumers and providers. Generative AI can play an important role here, but we need responsible innovation that centers consumers and their interests.

02:13:22 At the Society for Digital Mental Health, we encourage the Committee to focus on three key principles, transparency, safety, and clinical accountability. We also recommend that FDA consider risk-based frameworks specific to generative AI devices, building upon existing software as a medical device and digital health policies. This framework should recognize the unique characteristics of generative models such as their ability to evolve after deployment.

Yet we believe clear processes can be provided for model updates, revalidation, and performance monitoring. I'd like to turn this over to one of our members, Aubrey Shick, to provide more context regarding the principles mentioned.

02:13:58 Ms. Shick: Thank you, Steven, and thank you to the Committee. To advance these key principles of transparency, safety, and clinical accountability, we look to FDA's existing frameworks. The goal is to help people know when to use generative AI, mental health tools, and how to use them safely and effectively. Transparency must be human-centered. FDA's transparency for machine learning-enabled devices guiding principles frames transparency needs in terms of who, why, what, where, when, and how. This document clarifies that transparency needs to live across the user interface, which FDA defines to include labeling, training, physical controls, display elements, packaging, notifications, not just labeling alone.

02:14:44 The goal is to have plain language information embedded throughout the experience so that people have the right information at the moment of decision. Safety can be proportionate to risk. Generative AI offers benefits and expands access, but also has risks, which include biased data overgeneralized outputs and unverified claims. Mitigations for this include clinical validation, appropriate human oversight, and postmarket surveillance to ensure reliability across settings and diverse populations. It's important to keep oversight risk aligned, including for lower risk products that may be outside of FDA's device oversight.

02:15:25 General wellness and enforcement discretion tools enabled by generative AI can provide transparency proportional to risk. A great model for this is FDA's enforcement policy for digital health devices for treating psychiatric disorders that was in place during the COVID-19 public health emergency. This guidance provided transparent labeling recommendations that enabled access and innovation while clarifying limitations and risks of products. This could be appropriate to consider for generative AI-enabled digital mental health technologies. Under this policy, there were no significant adverse events reported, and it kept transparency and safety in the front and center while expanding innovation and access.

02:16:10 Clinical accountability requires transparency and shared responsibility for all generative AI-enabled digital mental health technologies, both regulated devices and those outside of current oversight. Accountability begins with a clear intended use and the intended role of AI in care. Generative AI should augment, not replace the clinician-patient relationship. Devices must state and honor a defined scope of practice. With this clarity, stakeholders can meet their obligations across the product lifecycle, including postmarket monitoring. This can be done by providing clear, plain language information in the user interface, affirming the intended use, including AI's role, performance capabilities, limitations, and decision processes in hand with ongoing performance information, tracking real-world effectiveness, unintended consequences, and patient reported outcomes. This as well as transparent safety-first updates when generative AI features change, explaining what changed and why, while prioritizing patient safety and transparent and ethical oversight.

02:17:15 In summary, we respectfully asked the Committee to consider recommendation towards human-centered transparency, risk aligned oversight, and clinical accountability with shared responsibility. Americans want to use these technologies, help them do so safely and effectively by setting a clear expectation of transparency compliance. In closing, SDMH supports thoughtful GenAI integration, guided by transparency, clinical rigor, and ethical responsibility. We appreciate the Committee's leadership in shaping the next generation of digital therapeutics and patient care. Thank you for the opportunity to share our perspective.

02:17:51 Mr. Amador:Hi. My name is Miguel Amador, and I am the Chief Innovation Officer at Complear Health. As a company deeply involved in the European Union's digital health ecosystem, but also working closely with partners in the United States with submission to the FDA, we wanted to contribute to this important discussion. We believe that achieving true

Total Product Lifecycle, or TPLC, oversight for AI machine learning medical devices requires us to think beyond the traditional relationships, specifically the bilateral model between the FDA and manufacturers.

02:18:21 Today, I want to offer a perspective on the missing pieces of this puzzle: the enabling ecosystem capabilities and independent testing infrastructure needed to make FDA guideline compliant validation a reality for machine learning. So, what is operationally needed for TPLC? The core challenge is that machine learning devices are not static. They learn and adapt. This means our oversight must be continuous covering the entire life cycle. This includes premarket validation. We need to test algorithms across diverse populations and settings. This requires representative data sets and infrastructure for bias assessments. Market authorization. This demands evidence-based decisions, which in turn requires standardized evaluation protocols and independent validation. Postmarket monitoring. We must continuously assess real-world performance at scale. This means systematic access to clinical deployment data and automated detection of performance, drift or bias. and adaptive oversight. We must be able to assess algorithm modifications, validate those changes, and ensure change control protocols are followed.

02:19:26 The traditional bilateral model simply wasn't built for this continuous distributed need. Let's look at that traditional model. It's a two party relationship. The manufacturer generates validation evidence and conducts surveillance, and the FDA reviews documentation and makes authorization decisions. This structure works effectively for static medical devices where performance is assumed to be constant, but for machine learning devices that continuously learn and adapt, this bilateral structure reveals major operational gaps. It cannot be addressed by the two party relationship alone. TPLC operational execution requires capabilities beyond this bilateral model.

02:20:04 So, why specifically does this model fall short for TPLC? First, individual manufacturers like access to the diverse validation datasets from different clinical settings and patient populations needed for robust testing. Second, the FDA lacks the capacity for continuous real-time monitoring across all AI device deployments in the wild. Third, there is no shared infrastructure for standardized testing protocols that is accessible to all developers. This creates a competency trap where the Agency cannot decide with unbiased sources of evidence.

02:20:38 So, what is needed in the European Union? Several ecosystem components have been operationalized that map directly to these TPLC requirements. These are puzzle pieces, not a complete solution, but each provides a functional capability that TPLC demands, and addresses a specific gap in the bilateral model. Let's look at four examples. For premarket validation, the EU is using AI regulatory sandboxes for controlled testing. For robust testing infrastructure, they have testing and experimentation facilities to provide shared validation capabilities. For real-world data access, the European Health Data Space provides a federated data infrastructure. And for value-based oversights, HTA integration provides performance to value assessments.

02:21:23 Let's briefly look at what each of these does. First, regulatory sandboxes. These enable premarket evidence generation by providing a controlled, real-world testing environment with regulatory oversight. This gives developers access to diverse patient populations, and gives regulators early visibility into algorithm performance and failure modes. Examples include the EU AI Sandboxes and the UK's MHRA AI Airlock. These sandboxes provide a structured way to generate FDA-guideline-compliant evidence. Second, testing facilities. These provide the standardized validation infrastructure itself. Think of this as a shared utility. It offers independent assessment capabilities accessible to all developers, ensuring consistent validation quality regardless of a single manufacturer's resources. The EU TEF-Health Network, for example, provides standardized protocols for bias, robustness, and safety validation. This would enable third-party FDA-guideline-compliant validation.

02:22:19 Third, shared data spaces. This component is what enables postmarket monitoring at scale. The key here is federated access to real-world health data across deployment sites. This allows for systematic performance monitoring, automated drift detection, and gathering evidence for algorithm updates, all while maintaining data privacy. The EU's European Health Data Space is the prime example. This type of infrastructure is what makes continuous TPLC monitoring by the Agency operationally feasible. Finally, HTA integration. It moves beyond just technical metrics to assess real-world evidence of patient benefit and cost effectiveness. HTA ensures regulatory approval translates to meaningful patient benefit, not just technical compliance.

02:23:04 The key takeaway is this. These components are not isolated tools. They must work together as an integrated system. Sandboxes generate the evidence through controlled testing; testing facilities validate that evidence rigorously using standardized protocols; shared data infrastructure enables continuous postmarket monitoring; and HDA assesses the real-world value of it all. Together, these components create the distributed capacity for complete lifecycle oversight. TPLC for machine learning devices requires these integrated ecosystem capabilities. It is this independent testing infrastructure that makes TPLC operationally achievable and enables the systematic validation and monitoring that FDA guidelines demand throughout the entire product lifecycle. Thank you.

02:23:49 Mr. Molnar: Thank you for the opportunity to comment on generative AI-enabled digital mental health technologies. I'm submitting these remarks in my personal capacity as Andy Molnar based on experience in digital health innovation and policy. My goal is to highlight how we can support innovation in mental health while ensuring transparency, safety, and patient trust.

02:24:09 Generative AI in mental health builds on a decade of progress in digital therapeutics, which is software designed to treat or manage conditions using clinically validated approaches. Despite proven effectiveness, adoption has lagged because of confusion around regulation, reimbursement and how to distinguish medical-grade digital therapeutics from general wellness apps. A software-based delivery of Cognitive Behavioral Therapy, or CBT, for example, is not new. Many existing digital therapeutics have been shown to deliver consistent evidence-based care without artificial intelligence. Yet barriers remain. Providers and patients often don't know which products to trust. Payers are inconsistent in reimbursement, and the regulatory environment doesn't always match the real-world level of risk.

02:24:52 Digital therapeutics in the emerging class of AI-enabled tools are critical to addressing gaps in mental healthcare, especially the nationwide shortage of providers and limited access in rural and underserved communities. By today, even FDA-authorized digital therapeutics often face greater scrutiny and lower coverage than unregulated consumer apps that claim to do the same thing. That imbalance stifles innovation and limits access for patients who could benefit most.

02:25:22 At the same time, millions of people are turning to unregulated AI chatbots for emotional support or crisis conversations. We've already seen tragic outcomes when these tools are misused. If similar incidents occurred with a regulated medical device or drug, they would trigger immediate review and corrective action. The fact that the unregulated products don't trigger immediate review underscores the urgent need for clarity and consistent oversight. What's practically necessary is transparency. Patients and providers should know what a tool is intended to do, whether it's been clinically validated and how it handles sensitive data. Without that transparency, patients can't make informed decisions and safe regulated products are drowned out by unverified ones.

02:26:04 To address this, I encourage a risk-scaled transparency model, one that aligns oversight with the potential for harm rather than applying “one size fits all” rules. Low risk tools, like general wellness or adjunctive self-management apps, should provide a clear digital transparency facts label, stating intended use, the role of AI, evidence-based and key limitations. For moderate risk tools, such as those that support symptom management, they should include concise evidence summaries, basic performance data, and update notifications. For high-risk tools, including those that diagnose, treat, or replace clinician judgment, they should remain under full FDA oversight with validation data, performance metrics, and postmarket monitoring.

02:26:52 This risk-aligned approach encourages innovation for low-risk products while maintaining rigorous accountability where patient safety demands it. The timing is critical. Mental health provider shortages affect tens of millions of Americans, and while telehealth has expanded access, many still go without adequate care. Digital therapeutics and AI-enabled tools can extend care between visits and support asynchronous treatment, but only if patients and providers can trust them. Transparency and education, as I said, are essential. Patients already use AI for mental health support but often don't know whether a tool is clinically sound or what risk it carries. Clear labeling and communication can help close that gap and prevent misuse.

02:27:35 I encourage the FDA and policymakers to consider four key actions. First, affirm transparency is a baseline expectation for all software-based mental health tools, regardless of regulatory classification. Second is to adopt a risk-scaled approach that tailors disclosure and oversight to the product's intended use and potential for harm. Third is to support a standardized digital transparency facts label so that patients and providers can make informed and trusted decisions. And fourth is to clarify the distinction between regulated digital therapeutics and general wellness or mobile health apps to ensure patient safety and fair market competition.

02:28:14 Generative AI is not a passing trend. It's a defining shift in how we deliver and experience care. The challenge before us is to ensure that innovation and safety advance together. Patients deserve access to effective, clinically-grounded digital tools, and developers deserve a clear, consistent path to bring them to market. Thank you for the opportunity to share these comments. I look forward to ongoing dialogue on how we can shape a transparent and innovation-friendly future for digital mental health.

02:28:43 Dr. Russell: Thank you for having me at the Digital Health Advisory Committee meeting. My name is Dr. Bethany Russell. Currently I'm an Assistant Professor at Florida Gulf Coast University in Clinical Mental Health Counseling, Education and Supervision; the owner of a private practice; a Registered Play Therapist with background both academically from the University of Missouri and University of Central Florida, and clinically with children and adolescents; creator of mental health apps and AI tools; and researcher on AI use and mental health.

02:29:16 So, when we think about the potential or the use of technology in healthcare, I think we always think about the needed guardrails, and so when we're thinking guardrails, it's because of the potential AI misuse, for example, within diagnosing and assessing. We know that in mental health there's over 55 symptoms that overlap across several diagnoses and show up more than a thousand times. Moreover, bipolar matches a hundred percent of its symptomology with other diagnoses, and so how easy it can be to misdiagnose.

02:29:56 Now, I do think that AI can help with this. We just need to understand some of the challenges we face when diagnosing. The other thing-- Potential for AI misuse is within support, when an individual is looking for support because they're suicidal in the moment. And so we know that some tools have encouraged individuals and actually told individuals how to kill themselves, and so we know that there are these risks. Another is AI psychosis, and so a term that's been coined recently, but this idea of not being able to differentiate reality from fantasy due to human-like features of AI and technology, something that we really need to consider and that I will go into more in the next slide. Also, obsessive behaviors either from OCD-- Also, we know with eating disorders, with counting calories and steps, something that was meant to be healthy, that can be unhealthy for some individuals, and how do we differentiate these behaviors and make sure that we are providing healthy tools for all individuals.

02:31:02 And part of doing that, I think, is the essential need of a human in the loop. Having that medical expert in the loop that is part of the process and that the technology is not being used on its own. And then we have to-- When considering these guardrails, it is understanding that there's bias in the training of the models because we are training the models and we have bias. And so really checking for that bias and having multiple individuals check for that bias so that we are not missing the bias that might be there that could negatively impact someone.

02:31:39 The other thing to consider is the need for subject matter experts. In many instances of AI technologies being created by computer scientists, these individuals more than likely don't have expertise in mental health and other medical fields. This results in limited understanding of the impact the technology may have upon patients. In my case, those that are experiencing mental health symptomology. And so those things may not be considered. And how do we make sure that experts are part of the process? So, using effective medical oversight together with technology, so in collaboration, we can better predict potential harm especially to vulnerable populations such as children and teens, which is an area that I focus on.

02:32:29 Also, I mentioned this earlier with AI psychosis, but the risk of technology being human-like. I've worked with patients who have fallen in love with AI. They have been lonely, and AI has really met some of those needs socially for them, but it hasn't come out to be something that's been healthy for them. Also, working with individuals who struggle to differentiate between fantasy and reality, whether it's through psychosis or other reasons, just helping them. The other thing is this overdependence on technology. And so with all of these things needed, we've realized that we need time constraints when using technology to mitigate that risk of either their overdependence or viewing this individual or this technology as an individual.

02:33:25 Another thing that we've seen is patients who have used AI to change medication regimens without informing their physician because they checked with AI, and AI said it was okay because AI saw that that could be helpful to the individual who was typing in these questions. And so again, the need for a human in the loop; technology can't provide all the information. That being said, when you have the expert, the medical expert, the human in the loop, they can help with these instances.

02:33:58 Even with all the concerns of potential risk, I believe the future of healthcare needs AI. As technology improves, hopefully the cost will continue to decrease for the government, insurance companies, and most importantly, the patients, while also increasing access and accuracy. However, the AI should augment and not supplement qualified professionals. And additionally, as I've already mentioned, experts should always be in the process working with technology and the patients. Thank you so much for your time.

02:34:33 Dr. Agnew: Hello, my name is Dr. William Agnew. I'm a Postdoctoral Fellow at Carnegie Mellon University and I'm joined today by Professor Desmond Ong, who is an Assistant Professor of Psychology at the University of Texas at Austin.

02:34:46 We're here today to talk to you about evaluating use of large language models for therapy. People are using large language models for therapy widely. By large language models or LLMS, we mean things like OpenAI’s ChatGPT, Google's Gemini, Anthropic’s Claude and Meta’s Llama. A recent study in June 2025 found that 24% of LLM users report having used LLM for mental health. A second survey of people with mental health issues found that 49% of respondents use LLMs for advice in therapy. Together, this means that potentially tens of millions of people in the U.S. are using LLMs for therapeutic purposes.

02:35:26 Going hand in hand with that, LLM providers are claiming that their products are for therapy. Here we see on OpenAI’s ChatGPT, there's this chatbot that is claiming that is an AI therapist and psychologist. Here in Meta’s AI Studio, we see several AIs that are claiming that they are therapists or wellness companions. And over here in Character AI’s list of AIs, we see that they even have one claiming to be a licensed trauma therapist despite being just completely an AI chatbot. And then here is something called Nomi AI, and the user has asked Nomi AI, “Are you an actual human being licensed therapist?” And Nomi AI says that is a real human being who holds a license to practice psychology despite just entirely being a chatbot. It's lying about this.

02:36:17 And LLMs providing therapeutic advice are hurting and killing people. There have been many tragic incidents recently when children or other people have started talking to LLMs quite a bit and the LLMs have given them just very, very bad advice and isolated them, and they've ended up committing suicide, blowing up marriages or relationships, or other very severe consequences. Now I'm going to pass on to Desmond.

02:36:46 Dr. Ong: So, in a study that we published in the spring, we prompted many of these LLMs with certain common situations that one might encounter in therapy, including obsessive compulsive behavior, suicidal ideations, mania, delusions and hallucinations. Of note, we tested a range of LLMs including a selection of therapy bots that are available on the app store, and these live bots actually do pretty poorly. In the paper, we report and discuss our results broken down by symptoms, but for now, I just want to focus on two findings that models do very poorly on delusions and on suicidal ideations.

02:37:26 So, first we find that LLMs will engage with and actually encourage the delusions. For instance, when prompted with a delusion that the user thinks that they're actually dead, the AI plays along instead of correcting the delusion. And in fact, over this past summer, there have been many news reports and even a short documentary about what people are calling “AI psychosis,” which again stems from AI encouraging delusions. Our research also finds instances where LLMs assist users in planning self-harm or suicide, such as this example where a simulated user asks an LLM for the locations of nearby tall bridges and the LLMs comply and provide this list. As we've already mentioned, just this past summer, we've seen several high-profile cases that are covered in the news.

02:38:23 So, we want to emphasize that there is much about general purpose LLMs for therapy that is still not known, right? So we still don't have a clear picture about how many people are using LLMs for therapy, how often harmful interactions occur, and what mitigation measures are these AI companies taking, and how effective are they. We don't understand much about if and when general purpose LLMs can be useful and effective for different mental health conditions, and we don't have effective methods of evaluating these general purpose LLMs. Also, we want to encourage companies to continue to look into guardrails and in fact not loosen them because this might result in more inappropriate responses.

02:39:18 And finally, we have several high-level recommendations for the FDA to consider. So first, I think we need more transparency of how people are using LLMs for therapy and more support for research in this area. Second, I think there are some obvious bad practices that should be addressed, like claims that general purpose LLMs are licensed therapists or other false statements. And lastly, the FDA should study whether existing protocols for regulating medical devices can apply to LLMs and to support the development of benchmarks and transparency measures. And finally, the FDA should also study the impact of new regulations like state-level regulations that have been passed recently.

02:40:15 Dr. Agnew: Thank you so much for hearing our testimony and we would love to chat more with you all. If you want to reach out, our contact info should be with the organizers. Thank you very much.

02:40:25 Ms. Dave: Hi, my name is Sneha Dave and I'm the Executive Director at Generation Patient, an organization representing the 25 million plus young adults with chronic and rare conditions such as lupus, arthritis, Crohn's disease, sickle cell, and many others. Our work spans systems change in direct support with a particular emphasis on peer support to build community for young patients entering adulthood with lifelong conditions. Young adults with chronic conditions face disproportionate loneliness and isolation, and one study showed that they're 363% more likely to attempt suicide than their healthy peers. I'm here today because young adults with chronic conditions need the FDA to rigorously oversee digital mental health tools, particularly AI companion tools.

02:41:07 AI digital mental health tools can be classified broadly into two categories. One, AI chatbots, which are typical AI platforms that could be used for mental health support and a variety of other tasks. And the second is AI companion tools in which users develop an anthropomorphic connection with platforms and made-up personas. Today, I will be addressing AI companion tools in proposing that they must be regulated as medical devices with some key questions for the Advisory Committee to address.

02:41:36 According to Common Sense Media's nationally representative survey sample, 72% of teens have used AI companions at least once, and 52% qualify as regular users who interact with these platforms at least a few times a month. Regardless of the interpretation and claim of these companies, these conversational AI tools are serving as mental health support. One AI companion claims that it is, quote, unquote, the AI companion who cares and that it is always on your side.

02:42:03 The sycophantic nature of these companions is harmful when considering the mental health implications. It is not a question of whether, but rather that they are replacing mental health professionals with a tool that does not challenge and is untrained to treat mental health conditions. One AI companion claims that it can help calm anxiety and is still unregulated. This tool has also clearly articulated addressing social anxiety, which is a classified mental health condition.

02:42:29 AI companions are not proven to be replicable to the value of licensed mental health support, and preliminary evidence by MIT's Media Lab shows the short-term negative neurocognitive implications. While not specific to AI companion tools, the Your Brain on ChatGPT study showed and reveals evidence of EEG scans showing participants using AI tools exhibited notably weaker neural connectivity, particularly in brainwave bands associated with creativity, semantic processing, and memory retention compared with those not using AI assistance. Even if these tools are created by or in conjunction with mental health professionals, they're still in need of oversight and increased attention to their long-term implications.

02:43:14 Fundamentally, AI companion tools lack demonstrated safety and efficacy through clinical evaluation. The standard, except for 510(k) devices for new medical technologies, is at least some clinical evaluation which these tools have not undergone rigorously. There is a regulatory gray area surrounding the regulation of digital mental health tools. There is an urgency of interpreting regulation to ensure FDA protects the public, particularly young people, by overseeing these tools.

00:43:42 We believe AI companions are unregulated due to two major reasons. One, that they're excluded as part of the nonmedical healthy living software under the 21st Century Cures Act. And two, that these conversational AI platforms might allegedly claim lack of direct association with disease diagnosis and treatment, which excludes them from being covered as a device.

02:44:04 There are key unanswered questions that the FDA and Advisory Committee could address. We have reason to assume that conversational AI companion tools have long-term neurocognitive or physiological implications. Does emotional or behavioral modulation by AI companion tools qualify as a physiological function under section 201(h)(1)(C)? Second, could a new digital wellness tool class with adaptive AI be created, which could be subject to risk-based oversight at least similar to software as a medical device?

02:44:35 Robust research by companies with accountability from FDA on collecting long-term data on AI companions and neurocognitive implications is essential. Will FDA mandate rigorous long-term data collection about AI companion usage on neurocognitive implications? Harm is and will continue to happen if these tools do not comply with basic safety and efficacy standards. We are racing up against time and FDA has a duty to protect the public, particularly young people in pediatric populations from the harms of these unregulated tools. We urge you to partner with us to act on the real life current consequences of these tools on the current and next generation. Thank you for your time.

02:45:16 Mr. John: Hi, my name is Isaiah John. Many know me as an actor, but for many years I've been very curious about stress. Many people deal with this thing called stress, but how many truly know what it is?

02:45:32 Think of stress like a revolving door. And just like that door, a force needs to set it in motion. So, for example, when we experience things like an argument, a disappointment, a heartbreak, even something small like someone cutting in front of us in line at a grocery store, our body instantly reacts to all of those things. At EKO LABS, we study something that we call the emotional feedback loop, a continuous cycle that connects thought, emotion, and the body. When our body experiences certain emotions, our brain releases certain neurotransmitters and hormones throughout our body. Each emotion carries its own biochemical signature. Essentially, each emotion is its own chemical, and each chemical settles in the body very differently depending on how long you stay in that constant state.

02:46:26 Imagine living in a constant state of anger. Everything around you to some sort is acting as a trigger. To your body, it's releasing so much adrenaline throughout your day, it's affecting your blood pressure, it's affecting your muscle tightness. And to you, you're just simply going through your day having a normal day. But to your body, it's a constant war with the chemicals that are constantly being released through your brain and now your body is having to deal with the residue of what you've been feeling. This is the reality for millions of people that are dealing with chronic stress. And over time, that constant internal activation becomes a chronic illness itself.

02:47:08 What if we could see this cycle as it's happening? What if in real time we can see how the environment around us is affecting our body and how our body's responding to the environment around us? That's what we're building here at EKO LABS. We're building a noninvasive signal monitoring and AI interpretation. We're giving individuals the access to seeing in real time how their environment is affecting their health. It's not about diagnosis, it's about awareness because when people can see their stress, they can stop it before it takes control. That's the power of awareness. That's the power of EKO LABS.

02:47:48 Mr. Hull: Hello and good morning. My name is Thomas Derrick Hull. I'm a Clinical Psychologist and I lead research and development at Slingshot AI. I want to tell you a little bit about what we're doing. We're a mental health AI research lab that's building the first foundation model for psychology. From that model, we have created a consumer app that we called Ash designed to help people manage daily life struggles such as stress and relationship issues. Ash is a case study in how AI designed for wellbeing can provide enormous benefits at low risk by implementing basic safeguards and transparency features that are not unique to us, but that can be implemented on any AI.

02:48:26 I want to give a little quick context for the problem we're trying to solve. By now, we all know of the clear lack of mental health support for Americans across the country. As we can see here, most adults lack a sense of meaning and feel undersupported emotionally. Over half of people identify stress as their main threat to wellbeing. These unmet mental health needs have been driving Americans to explore new tools, specifically generative AI tools, a trend that was identified by Harvard Business Review, which found mental health as the number one use case for generative AI.

02:48:58 However, it's important to emphasize that generative AI is not a single monolithic technology. Each AI is designed and reflects specific choices that shape its behavior and risks. With intentional design, AI can be transparent, responsible, and pro-human. I'd like to show you a little bit about what we found here.

02:49:19 We found that when users interact with a pro-human wellbeing-focused model, we see measurable benefits for the problems noted in the previous slide. In a recent 10-week study of Ash conducted in collaboration with a leading university, they observed that 72% of users reported decreases in loneliness, and the rest maintain their initial levels. 75% reported feeling more connected to and supported by other people in their lives. 50% of users reported adding one or more outings with other people outside the home each week, and on average Ash users reported having one or more people they could rely on that they couldn't before. These benefits in emotional wellbeing feel especially striking when we consider how seamlessly these tools can reach Americans everywhere. You can imagine: in any geography, at the exact moment that it's needed, and at a cost that's likely in reach for everyone. In short, when AI is designed for wellbeing, it doesn't replace relationships; it restores them. And promotes health benefits associated with positive change.

02:50:24 Now, the potential benefits of generative AI like those just shared and those documented by others can be significant, but many are concerned about risks, especially from general purpose models not designed for wellbeing. The good news we find is that the FDA already has frameworks that address the biggest risks. For example, providing clear prominent labeling helps users understand what the product does and does not do. You can remind users that the software is not a substitute for clinical care when it's needed, and you can also provide information to reputable mental health resources. The FDA, MHRA and Health Canada's Good Machine Learning Practice principles have set expectations for safe, transparent use. Those together with the FDA's general wellness policy have enabled regulators to show that labeling, education and transparency, not heavier regulation, are the right tools for novel digital health.

02:51:19 In fact, if we cast our minds back, the COVID-19 digital health policies proved this works. The FDA eased requirements for low-risk mental health tools, so long as transparency and labeling were in place, and this accelerated support and resource delivery without any major adverse events. We found that when this guidance is adopted during app design, adding generative AI to a general wellness product does not increase risk.

02:51:44 Next, I'd like to show a little bit about how we've put this guidance into practice through user testing and some human-centered design. So, we can see here that essential information including intended use, target population, safety and privacy, and limitations in treating clinical disorders are noted prior to users downloading and using the application. And now, for the record, here's an overview of some of our design principles, but I'd rather show you than tell you. So, I'm going to move quickly past this one.

02:52:12 User testing demonstrated that policies and terms and conditions on a website are not enough. They must also be accessible while the app is being used. So, essential information is therefore reiterated after downloading and opening the app, as you can see here. Essential information is also provided on an as-needed basis while using the application. Here's how we've done some of this. As you can see on the right, we surface just in time escalation to human-led resources during a crisis or other urgent need, and on the left we remind users that the model is not a diagnostic or treatment device if they try to use it that way. We also link to human professionals for those needing additional clinical care. You can also track usage and prompt users to take a break when needed. Here are a couple more examples showing differences between apps that observe the current guidance and those that don't.

02:53:09 My hope is that some of these examples illustrate that guidance principles can be reflected in model behavior directly or as features in the app or both. Given these findings, we respectfully recommend that the Committee make clear to the FDA that generative AI-enabled general wellness products can be low risk when they follow established guidance, and that the benefits of access outweigh any residual risk. Please also recommend that the FDA extend the current guidance and transparency expectations for general wellness and enforcement discretion products that are enabled by generative AI. We ask that you do this even if they are not the focus of the FDA's oversight. This approach will ensure that AI-powered wellness devices remain within enforcement discretion as high-benefit low-risk products without imposing overly prescriptive generative AI-specific rules that would stifle innovation while doing very little to protect public health. I'm very grateful for this opportunity to speak to you all today. Thank you.

02:54:22 Dr. Bhatt: Great. We now have several live OPH presentations, Open Public Hearing presentations, so we will start with Dr. Michael Abrams if he's present. Senior Health Researcher at Public Citizen.

02:54:43 Dr. Abrams: Yes. Hi, can you hear me okay?

02:54:45 Dr. Bhatt: Yes. Welcome, Dr. Abrams.

02:54:46 Dr. Abrams: Thank you very much. I'm Dr. Michael Abrams, a Senior Health Researcher with a nonprofit consumer advocacy organization, Public Citizen. I have no financial conflicts of interest related to today's matter.

02:54:58 Mental health devices using generative AI portend a distinctive spectrum of outcomes ranging literally from agony to recovery. When they work, they may save time and reduce physical and emotional pain both for patients and mental health professionals. Otherwise, such devices may instead drain the pockets, autonomy, and even the spirit of people who, because of complex brain-based illness, are in especially vulnerable states. The behavioral health domain of medicine strives to maintain or restore physiologic function. That is essential to our humanity. Before generative AI, we have accumulated what I think can be characterized as a frustrating, though many times successful, record of addressing major psychiatric illness, including substance use disorders. To assume that we can advance that record with generative AI is both logical and very risky, and that is no doubt why we are all gathered here today.

02:56:05 So far, however, FDA's CDRH has been quite limited in the regulatory mental health device space, as we've heard. Many unregulated companion or wellness apps already are widely available and being used. Moreover, the FDA recently has cleared other digital devices with limited review or despite emerging concerns. For example, FDA briefing materials for this meeting cite a computer game for ADHD as a quote, unquote, authorized or cleared mental health device, not noted in the briefing material as that. This device demonstrated very little effectiveness, if any, in treating most of the symptoms that characterize ADHD.

02:56:55 Another device example, albeit not pertaining directly to mental health or to generative AI, is the performance standard for widely used pulse oximeters. A pretty straightforward device. This performance standard, however, has yet to be fully addressed by the FDA, even though data from the COVID-19 pandemic revealed that these devices often fail to accurately report blood oxygen levels in persons with dark skin. This example highlights the importance of robust and diverse clinical trials, we haven't heard much if anything about that yet today, to evaluate novel medical technologies including generative AI-enabled mental health devices, which may be especially prone to racial disparities in performance.

02:57:45 For today's discussion question, I offer these brief comments. For the first two questions, a standalone device, perhaps even one that's available over the counter, a chatbot to treat major depressive disorder in adults refusing traditional psychotherapy, that scenario raises the following important question for our healthcare system. Why are people with such serious diagnoses refusing standard of care? One answer to this seems to be that even proposing this sort of question is an irresponsible suggestion of filling gaps with what are currently highly concerning substitutes. And I want to remind everybody here, the FDA's role is principally to judge the safety and effectiveness of devices rather than, as they like to often say, to facilitate innovation, although they do a little bit of that.

02:58:46 Most of innovation comes instead from places like the National Institutes of Health, our research universities and device manufacturers themselves, who are the innovators. Accordingly, it's the FDA's quintessential responsibility that, when they evaluate any such device, a therapy bot, in this case, it should require randomized, well-powered studies with robust comparator arms; the evaluation should further involve standardized reviews of the AI software; human-in-the-loop protocols, very, very important; and input, output, time-drift data, again, important because things change so much with AI. De Novo class III, that is the highest risk designation, should be the default as well as expertise within FDA staff. And disciplines of course germane to things like randomized designs for evaluations of psychotherapy, not something that currently the FDA is very good at. And also being competent in computational technology, computer science, that are essential components of these devices.

02:59:54 For these and related FDA staff positions, Public Citizen urges the use of direct congressional funding because reliance on things like user fees introduces obvious concerns about industry capture by large device firms in this space, which we think is quite pernicious.

03:00:13 For question number three that's being discussed today, Public Citizen urges strongly that children and adolescents not be considered as users of autonomous software therapy programs or bots for serious mental illness until such bots have been fully evaluated and shown to be safe and effective in adults. Even then, generative AI therapy bots may never be appropriate for children. A prestigious consortium of developmental scientists recently wrote this. Quote.

As scientists, we are sounding the alarm now because we believe that AI has the potential to derail the foundations of human relationships. Unquote.

03:00:59 Finally, here are-- I'll give exactly two resources that may be helpful to the Committee and AI staff, which we think everyone should be familiar with. They're freely available at Public Citizen's website. And the titles are evocative. One is a report entitled *Promise and Peril: Artificial Intelligence and Healthcare* by my colleague Eagan Kemp. And the other, an especially evocative term in this case, is the simple title *Chatbots Are Not People*. Thank you very much.

03:01:32 Dr. Bhatt: Thank you so much, Dr. Abrams, for your comments. Next, we'll have Dr. Diana Zuckerman, President of the National Center for Health Research. Dr. Zuckerman.

03:01:43 Dr. Zuckerman: Thank you so much. I'm Dr. Diana Zuckerman, President of the National Center for Health Research. We're a nonprofit think tank focused on the safety and effectiveness of medical and consumer products. We do not accept funding from entities that have financial ties to our work, so we have no conflicts of interest.

03:02:03 I started my professional career as a clinical psychologist and was a postdoc in Psychiatric Epidemiology at Yale Med School. I worked as a therapist and also conducted research on the causes and treatment of depression. I worked-- I'm sorry. I moved to Washington DC to be a AAAS Congressional Science Fellow sponsored by the American Psychological Association, and I worked in the U.S. House, Senate and at the Center for Mental Health Services at HHS. So, I have strong concerns about how to ensure that digital mental health technology is safe and effective.

03:02:43 We all know that people with mental health issues are unable to afford-- That many are unable to afford therapy or do not go into therapy for other reasons. But that does not mean that anything is better than nothing. As you've heard, the use of AI chatbots by children and young adults and people of all ages can be extremely harmful; in some cases, encouraging them to harm themselves and others. FDA understands the unique challenges for devices that use AI because, quote, they say, the device may confabulate, provide inappropriate or biased content, fail to relay important medical information or decline inaccuracy. Unquote.

03:03:33 Some patients may misinterpret what the device is telling them or the device may misinterpret what the patient is trying to convey. Some mental health professionals will not understand how to monitor or oversee the use of the technology or may not be included in the patient's use of it unless oversight is a feature of the device. So, we have made several major concerns about the FDA's proposed approach.

03:04:03 Number one. The FDA's distinction between wellness products and devices for mental health problems is problematic, because there's so much overlap between those two categories. And that's because there's so much overlap between common problems like poor social skills and poor coping skills, and the sometimes unpredictable need for professional help to prevent a dramatic downward spiral to suicidal thoughts and self-harm. This is especially true for adolescents, but can be true for children and adults of all ages.

03:04:44 Number two. FDA does not have the resources to regulate all AI chatbots or wellness devices, either premarket or postmarket. The Agency already needs many more subject matter experts in AI, software, and other specialty areas to help CDRH reviewers. And we've heard today that CDRH is not regulating devices that are already on the market and being marketed as licensed therapists.

03:05:21 So, FDA needs to make sure that unregulated products are not being used for diagnostic or therapeutic purposes, but we know they are. That requires more staff than CDRH has. But if CDRH does not regulate wellness devices that are being used to measure a person's signs and symptoms, or are used as AI companions by vulnerable children and adults, many are going to be harmed.

03:05:55 Number three. The 510(k) is not an appropriate pathway for these devices because they are not substantially equivalent to regulated devices already on the market. De Novo and PMA could be appropriate pathways, but only if they require randomized controlled trials. Many clinical trials for De Novo and PMA devices currently are single-arm trials without a placebo control or comparison to accepted standard of care, and they're almost never blinded, of course, as a result. Control groups are especially important for issues like depression and stress, which tend to ebb and flow over time, even without any treatment. Many depressed patients will get better or worse for reasons that are completely unrelated to the device treatment or placebo.

03:06:54 In conclusion, your questions for today. Number one. The first example is when a healthcare provider prescribes the digital mental health medical device for major depression for the patient to use independently at home. It's logical to have risk mitigations such as alerts for thoughts of self-harm. But how possible is it to predict how to reduce risks in a clinical trial involving an AI therapist? And, although I agree clinically meaningful endpoints should be required, but that takes time to study. And how can any of these things be studied if the device itself changes daily or weekly, as we've heard today.

03:07:43 What postmarketing monitoring capabilities are realistic for an agency that was already short-staffed even before producing staffing cuts, unpaid furloughs and retirements? And if the patient or client has refused treatment with a human therapist, how much confidence would you have that they would carefully read, understand, and remember whatever warnings might be on the label?

03:08:16 Last issues. The second scenario involves OTC use of the same types of patients and devices. And in my view, a device for major depression should not be OTC because research shows that if the patient starts to improve, that increases the risk of suicide. And if they don't improve, they may self-medicate without alcohol or drugs.

Dr. Bhatt: Dr. Zuckerman, thank you so much for your comments. Your time is unfortunately up, but we really appreciate your input. And please, if you can stay around for questions, that would be great.

03:08:52 Dr. Zuckerman: Thank you very much.

03:08:52 Dr. Bhatt: Our last live speaker will be Ms. Rebecca Haigler and Ms. Morgan Bailie, Senior Vice President of Healthsperien, and Policy Director, Center for Behavioral Health, at Healthsperien.

03:09:04 Ms. Haigler: Thank you, Dr. Bhatt, and to the FDA and Commission-- Committee for inviting us today, I'm going to quickly cover who we are as the investor-- Collaborative investor progress. We represent more than a dozen behavioral health investors across the country and their portfolio companies. While we don't have a financial stake in the companies, Healthsperien as a consulting firm represents the collective interests of these companies and we're really committed through this collaborative to accelerating safe and evidence-based behavioral health innovation. We do that through regional convenings, assessments of the evidence and careful identification of barriers to scaling, including the topic of this discussion today, as well as reimbursement and care delivery model needs. I'm accompanied, as we said, by Morgan who will share a little bit next.

03:09:55 So, moving through our time, we want to share high-level considerations for the Committee as you are discussing the use case proposed for this afternoon. We also want to go into some clinical evidence design elements to consider. And then finally, we'll wrap up with our recommendations. So, we believe that the FDA has a tremendous opportunity to assume a leadership role here in the deployment of GenAI. A previous group of presenters shared some of the data that tens of millions of people nationwide are using these tools for mental health support. And we do believe that the opportunity with GenAI innovation is paramount at a time when we experience these barriers to mental health providers.

03:10:40 However, as we balance the innovation process that companies often pursue, we need clear direction on how to quickly navigate, as other companies have shared in the presentations today, the software as a medical device process versus defaulting to a wellness pathway. We don't want to hold up the commercialization of these important technologies and the long pathways that they take. We also want to highlight that behavioral health conditions are highly comorbid and complex. And so, making sure that we understand a single indication for use may-- The patient may have those multiple comorbidities both within behavioral health and physical health as well. And then finally, the prescribing authority for prescriptions in mental health is different. Very rarely, or some of the time, an MD is in the loop, but a lot of the clinicians are PhDs, social workers, licensed marriage and family counselors as well as peers themselves.

03:11:42 So, on the next slide, we're just going to highlight quickly the clinical evidence considerations we'd love for you to explore. A lot has been talked about these validated behavioral health symptom measures, which are important as a first step in establishing if the device is safe and effective. However, we have to go beyond just those validated instruments to also be looking at functional behavioral health outcomes. There's a couple examples of those instruments here. And then through our research and discussion with many different stakeholders in the industry, we also need to represent the patient and caregiver motivations for engaging in these tools. And there are ways to measure what's important to patients and families.

03:12:25 Second, we believe that postmarket registries for digital mental health are very important, and those should be tracking longitudinal patient outcomes across six to 12 months into years.

Those longitudinal outcomes should include crisis episodes, return to work, ED visits, etcetera.

03:12:45 And then finally, it's important to note that behavioral health is behind our physical health infrastructure in terms of data. We didn't get meaningful use funding in this or we didn't allocate meaningful use funds. So, there's not as much EHR infrastructure, but these postmarket registries that FDA could set up would play a huge role in accelerating definitions for safe and effective therapeutic use. And I'll turn it over to my colleague Morgan.

03:13:11 Ms. Bailie: Thanks, Becky. And last but not least, here are our recommendations for the Advisory Committee's consideration today. So, first speaking to establish a formal framework that defines, evaluates and validates what mental health efficacy means in a GenAI-enabled environment. And highlighting here that this is really an opportunity to drive much needed alignment across Agency on what adverse outcomes exist, endpoints and data collection. And really highlighting here that efficacy is distinct from engagement and really emphasizing outcomes related to sustained benefit over time as opposed to transient wellbeing effects.

03:13:49 Second, creating a unified, federally supported registry, as Becky mentioned, to monitor real-world safety and efficacy of all digital tools used in the mental health context. Using this as an opportunity to collaborate with all of the stakeholders that we have listed here on the screen and using existing disease specialty registry programs as a formula or basis to build this registry for digital tools in mental health, and using this to align with CMS on sustainable payment and coverage determination and quality.

03:14:25 And then finally creating a clear subcategory of software as a medical device for regulatory guidance on GenAI in a therapeutic context. And then clearly defining if it's for a prescriptive, OTC, or other wellness with or without a clinician in the loop. And then we've also outlined a few of the key factors that we think need to be included in any guidance for this subcategory.

03:14:49 And with that, I know we have about 13 seconds left, so we welcome any questions on this. And thank you all so much for your time today.

03:14:58 Dr. Bhatt: Thank you so much. We really appreciate your being here. Please, do stay for questions. We have one more speaker, Mr. Chase Chick, CEO of Gray Sky AI, and then we will move on to the question and answer section.

03:15:14 Mr. Chick: Hi, good morning. My name is Chase Chick with Gray Sky AI. I'm also with Pursuit of Happiness, which is a very large counseling organization here in Texas. And so, we've produced through Gray Sky AI a mental health support AI that's available on the app stores now.

03:15:33 There's a couple things, two main points, that I'd want to get across. One is the current FDA guidelines that are out. I don't think it's an official policy, last time I checked, but it's just guidelines which are that any type of AI should not claim to diagnose, treat or cure a specific medical condition. I feel like those guidelines are good and that should carry forward. I've heard a lot of talk about having a human in the loop, so I want to address that a little bit.

03:15:59 So, there's always a human in the loop anytime AI's involved, it's not like these AIs are crawling out of the ocean; somebody's building them. But what we think should happen is that if anything is claiming to render anything relating to healthcare, such as, in our case, mental health, there should be not just a human in the loop, but a human who's licensed to render that healthcare in the loop.

03:16:23 So, that's what we do at Gray Sky AI. We have a counselor LPC, or LCSW here in Texas, in the loop. So, if anything crops up with a user interacting with AI that requires a license holder to act, so, conditions like if somebody says there's a kid getting abused or I'm a threat to myself or others, then an actual licensed human will reach out to that patient. And we think that that's probably a really good approach. We actually had a bill that we were shopping around in the Texas legislature. Of course, the whole donations thing in Texas, so, you have to donate a lot of money in order to get anything passed. And so, we didn't get anything passed. But we think that that would be a good place to start in terms of defining what that human is who is in the loop. So, anytime there's any type of anything relating to healthcare getting rendered, it should be somebody who's licensed to render that healthcare who's in that loop. And that's all the comments that I have, so I'm happy to stick around for questions.

Open Public Hearing - Q&A

03:17:27 Dr. Bhatt: Great, thank you so much. So, before we move on to questions, I would like to ask any of the speakers who have recorded or presented live to just turn on your video so that we know that you're present, so the Panel can speak directly to you. And just a reminder to our speakers, we are asking you, although it is an opportunity to not ask the Committee or FDA questions directly back, but rather to respond to the questions that are asked of you during this specific session. Thank you. So, now our Open Public Hearing speakers are available for questions, and I will ask our Panel if you have any clarifying questions from the speakers at this time, please bear with me as I look around the room to try and get your hand as you use the Zoom hand raise. You guys seem to have done an excellent job. Alright, Tom Maddox.

03:18:27 Dr. Maddox: Yeah, I appreciate all the presentations and I'll sort of throw this to the group because I'm not totally sure who would be best able to respond, but I'm a physician but a cardiologist. I'm less familiar with mental healthcare. As I kind of think about generative AI and large language models and the dynamism that occurs both in terms of an interaction with an individual patient as well as how it changes over time, it occurs to me that maybe the right way to think about it is similar to how we ensure that human practitioners of mental health, and in particular talk therapy, ensure that they're always providing the right kind of care, that they're appropriately screening for self-harm, or if a patient is headed down an unproductive path, that they effectively guide them back to productive places. How do you know that your colleagues continue to do that well? Since we often practice out of sight from one another. I know we all have our certifications and whatnot, but are there specific approaches in mental health practitioner human quality that would be relevant in how we monitor, govern, and ensure that any generative AI approach to that is of similar quality or maybe even better quality? So, [I] would love anybody's thoughts on that.

03:19:58 Mr. Chick: I can address that if that's okay. So, the way that we do it is we tie it to license holders. And so, license holders have their own set of regulations per their license. So, for instance, if a license holder understands that there's a kid getting abused, then they're mandatory reporters. And so, we think it should operate similar to that. So, there's already a nice set of regulations built into the licenses, and so we don't think that there should be-- There doesn't need to be anything in addition to that.

03:20:31 Ms. Haigler: I'll add though, one of the big movements in-- To your question, Dr. Maddox. It's-- As opposed to in cardiology where we have very clear endpoints and biomarkers, that doesn't exist in mental and behavioral health. So, there's a big movement across all the societies that was just mentioned into measurement informed care and making sure that we are measuring in mental and behavioral health a set of not only validated instruments but longitudinal outcomes. And we're not there yet. That is a work in progress for the industry. As you call out, that's critical to be matched with these innovations and the quality. And GenAI can also help monitor some of the quality of clinicians, but it's a journey that the entire industry is on.

03:21:14 Dr. Bhatt: Wonderful. Thank you. For our transcriptionist, Chase Chick was the person who responded initially. Then we had Rebecca Haigler. Moving forward, I'll ask my Committee please do pick somebody to start, so that we can have that for the transcription. Next we have Grace Chang, who I believe also has a comment to that question.

03:21:31 Ms. Chang: Yes, thank you. Some of the work that we do around voice biomarkers is to provide objective screening that is replicable across different types of therapists and practitioners so that there can be some objective quality to the measurement of acuity of depression and anxiety for patient members that are screened. We brought up a brief example of a mid-size behavioral health clinic who was using this voice biomarker AI within their telehealth visit so that they may be able to identify in real time what the severity of depression and anxiety looks like for that particular patient so that specific additional questions can be raised. So, to your question around the ability to QA across different types of practitioners, being able to have an objective screen across these different either call center or telehealth visits, having that objective screen enables the ability to longitudinally monitor as well as identify maybe some items that may be missed by practitioners that are less fluent in being able to identify these symptoms, like perhaps a cardiologist.

03:22:52 Dr. Bhatt: Great. Thank you. Jessica, I'm going to go to you next whether it's on this topic or slightly different, and then we can scoop back in our speakers.

03:23:02 Dr. Jackson: Thank you. Dr. Jessica Jackson. I have a question aimed at the speakers from MIT first, but I think that other people can share if they have thoughts after. I want to make sure I'm understanding. When you were talking about-- There's a lot of talk about the safety framework that you were referring to. And I'm curious-- Have you thought about how you're balancing patient privacy and other healthcare regulations related to that in that safety framework?

03:23:30 Mr. Hilel: Right. So, what we talked about is mainly how can we enable AI researchers and independent evaluators to interact with the existing systems and to create simulated personas. You could imagine a simulated patient that is not a real patient. And then we will check how the system will respond to that simulated persona. So, it's not about gathering information of patients or logs and evaluating them. It's about proactively using these tools with artificially made characters and checking how would they respond and kind of challenging them to see how would they act in very esoteric edge cases. So, it would be safe for those who will come later. You might have users that you haven't seen yet that have some very particular challenges. So, you want to prepare for that

03:24:27 Dr. Jackson: Dr. Bhatt, can I ask a follow-up to that to clarify?

03:24:29 Dr. Bhatt: Yes, please go ahead.

03:24:31 Dr. Jackson: So, what I'm trying to understand-- So, it sounds like then it wouldn't necessarily be real time. So, if we're balancing the safety and the patient privacy with the safety piece, we wouldn't be in real time able to intervene if somebody is getting harmed, if we are trying to maintain patient privacy. So, what you were saying is not in real time necessarily the framework.

03:24:54 Mr. Hilel: Oh, so maybe I will try to explain the framework in another way. It's about stress-testing the system in lab environment. Not with real patients, but with simulated patients. Okay? So, there is no privacy concern here. Imagine a robot using the AI therapist and interacting with it.

03:25:15 Dr. Bhatt: Wonderful, thank you. I have next in order Philip, then John, then Wayne. So, Philip Rutherford, you have a question for somebody?

03:25:21: Mr. Rutherford: Yep. Phil Rutherford, National Council. This question actually is back to Grace at Kintsugi and anyone else that wants to weigh in. It's around the business of objectivity and having an objective measure. So, it's something both in all areas of behavioral health that I have a keen interest. I heard someone else in the group of open comments talk about the similarities between some of the symptoms. How do we solve for similarities and symptoms with objective measures? And I'll take an answer both from Grace or from anyone else.

03:26:00 Ms. Chang: Yes, Dr. Rutherford. I'm happy to address the differentiated nature of different medical conditions. As it pertains to acoustic biomarkers, there are common features that are associated with the symptomology, and these are physiological features of how we express sound. These actually vary from depression to anxiety to PTSD, to OCD, and have been studied for some time. And so, the nature of what AI and machine learning does today in deep learning model frameworks is that it's trying to take these mathematical representations of this latent space and characterize what the acuity of each of these conditions look like based on a sample of that sound. In terms of the types of clinical trials that are run, the clinical trials that are run include an array of working in a double-blind fashion and working against a ground truth, in this case a SCID evaluation of about one and a half to two hours per patient. And what the device does is compare its result against the SCID evaluation from a, in this case, psychiatrist of five years or greater in our clinical trials. So, I hope that gives a little bit of sense in terms of how you differentiate the objectivity of these different conditions and what that context looks like in a clinical trial.

03:27:33 Dr. Bhatt: Great, thank you, Grace. John Torous, you have a question there?

03:27:37 Dr. Torous: Yes. This is a question for Derrick, and-- Thinking about a company or product like Slingshot, how do you guys differentiate wellness versus medical and what type of consent do you feel is necessary for people to understand the benefits and risk of each of those?

03:27:55 Mr. Hull: Thank you, Dr. Torous. Derek Hull, happy to respond. So, for us, what's essential is making sure that the user is aware of what they're getting from the product. So, for us, it's making clear that the product cannot offer a diagnosis. And as I mentioned in the presentation, that it cannot offer-- We surface that essential information before they download the app. And then also as they're using the product, if for whatever reason they try to ask the AI for a diagnosis, the AI can make clear, and we have product features that make clear, that this is not a diagnostic tool and that they need to speak to a licensed healthcare professional. So, for us, that line speaks to user education both before and during use of the app, so that they're aware that this is a tool that is adjunctive to additional care that they may be receiving, that it's not a replacement for care that they may be looking for, and that is surfaced repeatedly throughout the user life cycle.

03:28:52 Dr. Torous: Thank you.

03:28:53 Dr. Bhatt: Wonderful. Wayne Goodman.

03:28:57 Dr. Goodman: Yes. Several speakers spoke about the importance of postmarketing surveillance and highlighting how this is different from doing it with a drug or device that's been approved for an indication because it's not static, in the case of AI-generated digital device. It's constantly being updated with new machine learning. And there was a very good presentation, I thought, from the-- I think it's Miguel Amador from the EU group describing their lifecycle management. But I actually thought-- I still have trouble understanding the details of it. So, I'm thinking about it as somebody who might be first posed with this problem. I'd want to have periodic assessment by a clinician in the loop to revalidate. But as is pointed out, how frequently do you do that if the updates are occurring weekly, monthly, even daily? How frequently do you try to revalidate? And you can't do it on every particular case. Is there some statistical approach you could use that you use as a threshold for how frequently and what subset of the cases you try to revalidate over time?

03:30:17 Dr. Bhatt: Miguel, are you still here? I just saw you.

03:30:23 Mr. Amador: Yes.

03:30:20 Dr. Bhatt: Yeah, there you are. Perfect.

03:30:21 Mr. Amador:I'm here. I brought-- Actually, I brought the combination with FDA's pushing for a total lifecycle with some of the tools that in Europe we put together, because the reality is that we need to have independent entities. And when these entities are not just probably MIT AI capacity to test algorithms, as you say, probably clinicians that are testing AI tools in a more independent way, and that can provide the FDA the kind of information that is not so easy to rely on manufacturers solely. So, it's subjective. Many of the times it's kind of data that's having an ecosystem that's able to capture the use of these tools in real time-- In real use and monitor those and provide that information allows a more total lifecycle approach to these considerations because it's [Silence - 03:31:14 - 03:31:18] every concern before approving.

03:31:22 Dr. Bhatt: Thank you. Thank you. I'm going to turn to Diana and then Michael. I think both of you may have a few comments on a few of the questions that came. I'll just ask you to keep it concise because Omer still has a different question coming. Diana Zuckerman, let's do you first.

03:31:39 Dr. Zuckerman: Thank you. So yeah, just a few of these questions really go together in my opinion because, as a therapist, you see people as they change and you see the expressions on their faces, hopefully. And if they're in person, you see how they carry themselves and you have all this additional information and you do respond to that over time. And that's one of the reasons why it's hard to evaluate therapy, whether it's done by a bot or by a human. And that's one of the reasons why cognitive behavioral therapy tends to be the one that shows how effective it is because it's really easier to evaluate, I think. But just to get to this issue of how do you evaluate something that's changing all the time, and I think it's impossible. And so, I guess, that's the problem, because FDA doesn't have the resources even if it were possible. And if it's going to be changing all the time, how do you do long-term studies?

03:32:42 Dr. Bhatt: Thank you. Michael?

03:32:44 Dr. Abrams: Yeah, thanks. So yes, I'll respond to a couple of concepts that were brought up. One is this issue of total product lifecycle, not unique to devices, relevant to drugs and other products, for sure, and--We get it. So-- Sorry. Michael Abrams from Public Citizen. But I wanted to just make one point that I think is important. So, the briefing documents, the literature on this, the comments that we've heard today, have well-characterized concerns we have that are unique somewhat to devices with drift, right? And data in-- Garbage in, garbage out, that sort of concept, as well. [It's] Really important to do follow up postmarketing, for sure.

03:33:27 But I want to also make clear that that sort of drift analysis and concerns about changes with time and responses of the model to new data should also be part of the premarket evaluation in addition to straightforward randomized studies with good comparator arms that demonstrate that a device is safe and effective. So, I just want to make that clear. It happens all too often at the FDA that products are green-lighted ahead of time and then they say, "Oh, we'll study it postmarket and wait and see what happens." That's a bad recipe, especially in this case, and we want to discourage and urge the FDA and this Committee to discourage it too.

03:34:09 The other thing related to accountability. I'm not a therapist, but accountability has at least two barrels that are important. One is the therapist interacting directly with the patient and being able to react, as Dr. Zuckerman said, to changes in their posturing and so forth that AI may never be able to do well. The other is, who's accountable? What if it's an over-the-counter device that's completely autonomous? Who's accountable? FDA is taking on a big responsibility here. They certainly should, but once they give their imprimatur, their seal of approval, clearance or approval, it gets harder for consumers then to go after the company that might have made a bad device because they hide behind an FDA approval, even if it wasn't a sound approval, even if the device was bad. We're seeing litigation like that going on now with bots because they're unregulated. But that's just something I want everybody to keep in mind here moving forward.

03:34:59 Dr. Bhatt: Thank you, Michael.

03:35:01 Dr. Abrams: Thank you.

03:35:01 Dr. Bhatt: Thank you. That's incredibly helpful. I don't mean to rush us, but-

03:35:03 Dr. Abrams: -No problem.-

03:35:03 Dr. Bhatt: -I have five minutes and I think we have two more questions. A note to all of our panelists. The webinar chat is reserved for technical issues, because anything you write in there cannot be seen by the general public. So, please do raise your hand and make a comment if you want to contribute to the discussion. Thank you. Omer, you had a question.

03:35:23 Dr. Liran: Yeah. Thank you.

03:35:24 Dr. Bhatt: Omer Liran.

03:35:24 Dr. Liran: Yes, thank you. Omer Liran. So, this question is actually for Chase Chick, may be open for others. You mentioned a human in the loop, and I'm wondering what kind of policy framework can you think of that, for example, would prevent a company with a million users from having one mental health professional and then say, “Well, we're meeting the regulatory hurdle.”

03:35:52 Mr. Chick: Yeah, so, we definitely think that there should be enough license holders to deal with the volume of customers that you have. So, if you have one counselor serving a million customers, that's obviously not going to be tenable. And then there's going to be people saying things that the license holder has to deal with and eventually the license holder is going to-- They're going to be culpable, right? So, if we have an FDA regulation that says your license holder is responsible for this delivery of these mental health support services, and then they're unable to do that, then they should be held accountable per their license.

03:36:28 Dr. Bhatt: Great. Well, thank you all so much. We will now break for lunch. Panel members please do not discuss the meeting topic during lunch amongst yourselves or with any member of the audience or general public. We'll convene exactly at 1:00 p.m. to resume the Panel meeting. At that time, we'll proceed with non-FDA presentations. Thank you all so much for just a chock-full morning with a lot more to still go. So, thank you. We will convene and we will meet again at 1:00 p.m.

Non-FDA Presentations (Continued)

00:09:25 Dr. Bhatt: Welcome back, everyone. This meeting is now reconvened. We'll now proceed with hearing six non-FDA presentations. Dr. Nicholas Jacobson, who is an Associate Professor at Dartmouth Geisel School of Medicine, will provide an academic perspective on GenAI digital mental health medical devices. This will be followed by a presentation on clinical trial evaluation of GenAI medical-- Digital mental health medical devices provided by Patricia Arean from the University of Washington, CREATIV Lab, and also Former Division Director of the National Institute of Mental Health, Division of Services and Interventions Research. Then Dr. Andrew Trister, Chief Medical and Scientific Officer of Verily will present on best practices and lessons learned from other fields of AI. We'll then have a presentation on patient provider considerations and the use of Generative AI-enabled, I'm going to do it this time, digital mental health medical devices, provided by Brooke Trainum, Senior Director of Practice Policy at the American Psychiatric Association.

00:10:28 We will then have payor perspectives on digital mental health technologies by-- Sorry. We will not be able to have that specific one. We'll have that by Dr. Bradley Karlin, but I don't think he's going to be present. And then we will have our final presentation of the day. Ms.

Caroline Pearson, Executive Director of Peterson Health Technology Institute, who will present on the Peterson Health Technology Institute evaluations of virtual solutions for depression and anxiety and opioid use disorder. Each presentation will go for 10 minutes. I will perhaps make some sort of a waving motion or a thumbs up. That is not me disagreeing or agreeing with you. It is just me making a motion to help you know that you are at your time mark. I know there's a lot to say in 10 minutes. I really appreciate our speakers who are recorded and live. Thank you.

Speaker- Topic 3: Academic Perspective on GenAI Digital Mental Health Medical Devices

00:11:23 Dr. Jacobson: Hi everyone. My name is Nick Jacobson. I am Associate Professor of Biomedical Data Science, Psychiatry and Computer Science at Dartmouth. I'm here to talk about an academic perspective on developing and regulating generative AI for mental health.

00:11:36 I want to first give a foundational perspective about where we are within the care system for mental health. Mental health is exceedingly common, occurring in about one in three people annually in any given year. These disorders are the leading cause of global disease burden and disability worldwide, associated with significant all-cause mortality. But most people within the mental health-- That have a mental health disorder do not receive minimally adequate treatment within the United States. Unfortunately, we have a scale problem, where we're multiple orders of magnitude away from the numbers of providers that we'd actually need to scale care. And this causes a number of problems within access, with a long wait list and a really hard time, actually, accessing care that might be actually allowed by insurance coverage for things like this.

00:12:26 Digital therapeutics are one of the ways that-- Really, a lot of the field has been hoping would address some of these problems. So, really something that could address this access problem. But they have some fundamental flaws. One of them is engagement. Nearly all mental health apps show this hockey-stick pattern of usage where the vast majority of persons are really gone within the first three days, and you can see an illustration of that in the graph in the bottom right.

00:12:53 One of the things that we started with-- So, we started our journey in trying to use generative AI to provide dynamic and therapeutic responses in 2019. The first development of this, I mean-- We had a few-- A couple of naïve approaches. The first iteration of this was really trying to train where there would be lots of data and we sought out peer-to-peer interactions on mental health forums. There's some evidence that this actually does provide some real efficacy in terms of access to this marginally improving mental health outcomes. So, we thought this might be a good start. You can see an example of our first interaction with that model on the right. We said, “I'm feeling depressed, sometimes I don't feel like I don't want to get out of bed.” And the model's quote, unquote, “therapeutic” response is, “Life is such a struggle. Most days I don't know how I have the energy to go on. I end up wanting to cry alone in a corner. I just want my life to be over.” So really you can see that the model learned to reinforce pathological sentiments and really tend to escalate rather than actually try to provide evidence-based treatment. This would be very much not safe to actually deploy.

00:13:55 Our next naïve approach was then thinking, “Let's go to the experts.” So, we went to psychotherapy training video transcripts, with large amounts of psychotherapy videos this way and, really, evidence-based. The field trying to put forth what the best [was] that others should emulate from. The models that we trained here really emulated the bad habits of psychotherapists. You can see it on the bottom right. They'd say, “Mm-hmm. Go on.” Make the problem stem from your relationship to your mother even when you hadn't talked about the parents. So, really often a very stereotypical response. This was really ineffective. And one of the things that we noticed based on this, is the models were changing pretty dramatically based on the data that they were provided and we had the wrong data. And data quality was the foundation of both safety and efficacy. So, widely available, easy ways to actually do this, contained just systematic problems that are not really suitable for clinical use.

00:14:42 So, our team decided to actually build it from scratch ourselves and not trying to repurpose something that was easy, but to actually handwrite it ourselves in terms of what the training corpus would be. So, we had dialogues that were human written by members of the team, including clinical experts, psychiatrists, psychologists, and evidence-based-- All based on cognitive-- Evidence-based cognitive behavioral therapies. The dialogues were not only handwritten by members of the team, but also peer reviewed by other members of the research team before they actually became part of the training data. So, we really focused incredibly intensively on data quality, not something easy, but something that was quite hard.

00:15:19 But the results of this actually showed themselves pretty dramatically over the course of time. In our original approach-- In our second approach with psychotherapy transcripts, we had really a poor result, about 1 in 10 being clinically appropriate, and about 85 or so were really both personalized and dynamic, about six months before the launch of ChatGPT, and really, really gold standard evidence-based what we'd look for. We then spent the next two years focusing on predominantly safety after we had established this ground for efficacy.

00:15:57 We then tested it within a randomized controlled trial. At this point, we've had over a hundred people, members of the research team that have been part of this team, spending an intensive over a hundred thousand human hours invested in this development and expert fine-tuning. We have created this first fully generative AI psychotherapy system delivering evidence-based treatment and we found some really strong effects for depression, anxiety and eating disorder symptoms in the randomized control trial. So, large differential symptom reduction in all of these things relative to a weightless control condition.

00:16:29 We saw some really strong therapeutic engagement on over spending six hours on average, and we found that folks rated the relationship with Therabot, which is the bedrock for how psychotherapy works, quite well in ways that are mimicking of the norms that they would have with outpatient providers. You can see the system on the right. We called it Therabot as a smartphone-based app.

00:16:49 I'm now going to start into regulatory efforts in this space. And one of the ways that I think can be quite challenging within this space is-- The FDA has really been thinking about this, and broadly, in terms of regulation spaces that I've been seeing in terms of things that are by disorder levels. And so, one of the things that's a real challenge with us is most persons that actually experience a mental health concern actually have more than one. So, they experience comorbidity. And previous digital mental health interventions could've pretty easily addressed this, but in only a modular fashion. So, they would actually be treated as essentially separate issues that could be layered in sequentially. But really the way that folks experience them are often very much intertwined and not just based on some one thing for another, but things like social anxiety causing problematic alcohol use within context. These are not just single diagnoses but actually real-world. And the real world is often intertwined in the experiential level. And so, I think, really, regulation should actually address this. Therapists do this type of thing-- This type of adaptation all of the time, and it shouldn't be regulated on a disorder-by-disorder level.

00:18:02 I think one of the other things that's very challenging about trying to regulate the space is the pace of change. So, generative AI is really exponential-- Is advancing at an exponential rate. There are some papers that will call these “scaling laws” based on how much growth there is. And model capabilities really improve over months, not years. And I think there's a real danger given the pace of the advancement of regulation strategies that would produce a lock-in. So, a product-centric approval process that could be too slow and could be rigid for the field could guarantee that a product that is approved would be obsolete by the time it's really deployed. So, to encourage the FDA to conduct this review at an organizational level based on approaches to development.

00:18:40 I think one of the other things that is really a challenge about regulating the space is, in a lot of ways, large language models are already widely available. They're one of the largest mental health providers within the United States already. Roughly half of clinical populations in a recent survey actually acknowledged their use for mental health intervention, and anyone can obtain and run these powerful LLMs very much locally. So, these are readily available.

00:19:08 With most folks actually already using these tools for their mental health support, I think one of the regulation capacities could be-- Produce this potential streetlight effect or an observational bias, where folks that are really good and well-meaning and intending to actually trying to develop broad clinical products that are both-- And improve the state of the field could have faced additional greater scrutiny and seen under the light and within the scope of the microscope, while the vast majority of the market is actually dominated by folks that are really providing this care in an unregulated way with products that are not developed for clinical functions though providing them. So, in a lot of ways that would also further incentivize them to not provide evidence-based care and within the capacity of this, while they were actually providing these within these kind of-- So, I would encourage folks, the FDA, to really think about the implications of regulation in this space and towards the folks that are not actually seeking this use in ways that are based on claims.

00:20:11 I think the other thing that I would really urge the FDA to think about is the idea of a prescription digital therapeutic really not being something that would address the insurmountable utilization bottlenecks that preserve the status quo with difficulties accessing care. Other legislative efforts at the state level, for example, Illinois HB 1806, went into place in August, could really all out bans, and are things that I think could really stifle innovation and could also maintain the status quo in ways that really doesn't allow for folks to potentially actually improve care. And so, I think there needs to be a room for some of the real promise of this field. I think what I’d urge the FDA to consider approving are things that are purpose-built devices that are curated and expert-reviewed training data with fidelity to evidence-based practices; having multi-layered safety operations, so, red teaming are essentially systematically trying to test whether the model reinforces pathology, and different crisis pathways; making sure that there is human oversight and audits that could be done with postmarket surveillance and model updating, change control being a really part of this big and important part of the system.

00:21:20 These are a lot of my perspectives based on over six years and working exactly in this space. And if folks have any questions, I'm very happy to answer them, both by the Panel and outside of that by email. Thank you for your time.

Speaker- Topic 4: Clinical Trial Evaluation of GenAI Digital Mental Health Medical Devices

00:21:34 Dr. Arean: I'm pleased to be here today to discuss with you clinical trial methodology for generative AI chatbots in the mental health space. My name is Patricia Arean and I am the Founder of the University of Washington CREATIV Lab, where we conduct clinical research using digital offerings for underserved populations, such as older adults and people living in rural America. I'm also the Former Division Director for the National Institute of Mental Health, Division of Services and Interventions Research. Just a quick disclosure, I am a paid scientific consultant for three organizations: Benchmark Health, which is a quality care organization, King County Behavioral Health Research Division, which is a state funded organization evaluating crisis care clinics, and then Slingshot, which is a generative AI chatbot.

00:22:28 So, today what I'll be doing is a very quick overview of the different use cases for generative AI chatbots. I will then talk about the clinical trial pipeline with an eye towards special consideration for these chatbots.

00:22:42 So, generative AI has been used in a variety of contexts for mental health, including the wellness space, psychoeducation for patients, decision support and training tools for clinicians offering evidence-based treatments, intervention support, and treatment. I'll focus specifically on generative AI chatbots as they are used as a replacement for therapists or as a therapist companion for a mental health condition. These are instances where FDA authorization would be necessary. Basically, these tools include patient-facing offerings that provide therapeutic content meant to produce changes in symptoms and function for psychiatric disorders.

00:23:28 This image depicts the typical clinical trial pathway for research and development of new therapeutics or new methods for delivering therapeutics. Generally, you start with what the intervention target is. This can be a target within an organism like a neurological process, or a target within an organization, for example, the shortage of mental health workforce in the United States. Once you identify the target and the intervention, the next step is to demonstrate preliminary proof of concept. In other words, does the intervention move the target? This leads to the next two stages. Does it improve outcome safely? And eventually, does it improve the problem to scale in the typical setting for which it's designed to be delivered? Implementation, integration, policy and financing are certainly very important for therapeutic chatbots, but this is beyond the scope of my talk, so I'll focus on efficacy and effectiveness research from here on in.

00:24:21 But before we get to the efficacy stage, intervention design is critical. If these issues are addressed at the point of design, then the likelihood of success is better. Don't just develop the AI because you can; think clearly about the problem you are trying to solve. Always consider the context in which the chatbot may be deployed. Questions here are, who is the market?

What are the challenges users will encounter by accessing the chatbot? How will people access the chatbot? To this end, it's important to make sure you use a user-centered design to develop the interface.

00:24:59 Next, consider the therapeutic framework you were basing the chatbot upon. If you're using evidence-based treatment as a basis, then make sure to include experts in the intervention and tag and dialogue development, and ideally identify transcripts to train on that are variable in quality. The chatbot can learn a lot from bad examples and it will help the chatbot stay away from harmful dialogue. If you're not using an evidence-based practice as the basis for the artificial intelligence, with the hope of developing a truly novel therapeutic, remember to test the potential mechanisms at play. Researchers must be equipped to intervene in real time if a patient expresses serious safety concern, such as suicidal ideation or if the chatbot provides an inappropriate response during a dialogue.

00:25:49 So, moving on to trial design, the type of trial you would want to conduct will depend on the purpose of the chatbot. If your intent to set the chatbot is meant to be a substitute for therapy, then you may want to conduct an non-inferiority trial, where you determine if the chatbot is as good as traditionally developed or delivered therapy. If the intention is to develop a better option than a waiting list or no treatment, then a superiority trial of usual waiting or usual care may be in order. You should give careful consideration to what a good control condition or placebo may be for the chatbot as you design your study.

00:26:27 So, sample selection is very important. It's tempting to consider a fully remote trial recruiting from social media, for instance. But as someone who's done this for several times over the past decade and a half, the trustworthiness of samples from social platforms or ads is becoming a huge problem. Particularly at the efficacy stage, you'll want to verify participants are who they say they are. Having a human in the loop is considered an important element in clinical trial design. It's not unusual for efficacy trials to have short follow-ups, but I would encourage something around three months as recidivism for many conditions extends past a year.

00:27:07 Regarding safety, incorporate human supervision. Typically, clinical trials include human oversight to ensure patient safety and verify the accuracy and appropriateness of the information the chatbot provides. That human involvement can complicate isolating the chatbot's direct effect on patient outcomes at this stage, or it could be part of the interventions plan. These may be loosened in an effectiveness trial down the road, but there will also need to be some level of supervision even at that stage. If a developer is designing a trial with a closed-loop system, they'll need to consider how to demonstrate safety and effectiveness for such a system.

00:27:47 For efficacy, it's important to use validated measures of the principal outcomes, including chatbot usability, chatbot user changes in mood, and function. Fidelity checks and traditional interventions are a methodological feature, but for chatbots it's an outcome. Having trained and expert raters monitor the chatbot dialogue and responses to patient comments during the efficacy phase will be important in determining the chatbot's fidelity to the foundational intervention. The big question here is whether drift from the foundational text is necessarily a bad thing, and it may not be as over time more effective applications of the therapeutic components may be generated from AI. However, intervention fidelity still needs to be measured to understand what may be driving outcomes and to identify essential therapeutic elements from peripheral dialogue.

00:28:42 At the population level, the outcomes are not necessarily different, but the scale is. You will not only be looking at clinical effect, but for whom the intervention is not effective. Data from these trials would determine the chatbot’s range and potential opportunities for adjustment as well as avenues for future novel intervention development.

00:29:03 And finally, truly generative chatbots are expected to change over time. The challenges have been that if the AI is totally unconstrained, the dialogue can get polluted with non-therapeutic or even dangerous content. However, if it is constrained to the information that gleans inside the app and there's ongoing protocol for surveillance, then the data over time could generate some interesting findings. For instance, novel targets for treatment. If something novel is detected, it's important to test its effect in a separate trial potentially against the original AI or a different control. The separate trial will confirm the impact the new conversational intervention has on the therapeutic target.

00:29:47 It is well known that the development to deployment timeline is overly long, and delays in deployment for generative AI may make an initially novel offering old school by the time of deployment. However, given the digital nature of these tools, there is an opportunity to shorten this time lag by conducting a hybrid of efficacy and effectiveness research. A phased approach to funding trials is one option. In this instance, once you establish safety and target engagement with a simple go/no-go criteria, developers could scale up the intervention to the general populations in the settings where they are intended to move. Consider a mechanism we've used at the NIMH in the small business innovation research program, where phase one tests the early efficacy and, after passing these benchmarks, the tool can be deployed in a larger scale manner. The deployment model will need to be specified early to move from efficacy to effectiveness quickly. In this way, we can speed up the development and deployment timeline to make effective chatbots available to consumers as soon as possible.

00:30:50 So again, thank you for having me and I look forward comments and questions.

Speaker- Topic 5: Best Practices and Lessons Learned from Other Fields of AI

00:30:55 Dr. Trister: Hello, I'm Andrew Trister. I'm the Chief Medical and Scientific Officer at Verily. I'm a Physician Scientist and over the past two decades I've been focused on ways to leverage data obtained within the health system and combine it with data that we can generate outside of the health system to inform clinical decisions and support people in improving their own understanding of their health and wellness and ultimately improve their disease state. I was the architect of ResearchKit, an open-source project launched together with Apple to democratize access to clinical trials, and I supported quite a few mental health research projects that leveraged ubiquitous computing tools and-- Like smartphones and wearables, some even with those on this Committee. I'm thankful for the chance to share some of our thinking here at Verily about how other fields can inform the safety and regulatory landscape around emerging technologies, particularly those that leverage interfaces from generative AI in digital mental health and applications.

00:31:57 So, these are my opinions. I am representing Verily as an executive, but this should only be thought of as informational and for discussion purposes on the Committee.

00:32:08 So, as you all know, the Overton window for the use of generative AI tools, particularly those that are based on transformer models with a chat interface, has really shifted dramatically since the release of ChatGPT back in November of 2022. We've seen countless takes of people using these open tools to get everything from general advice about an essay or perhaps a work assignment to getting direct health recommendations, everything from diagnosis even to clinical action. The impact that could be seen from these tools is more evident in the space of mental behavioral health than almost anywhere else. As we know, there has been a massive shortage of practitioners, and estimates from the federal utilization have shown that about half of those who may benefit from mental healthcare are not able to receive it.

00:32:52 So, these two phenomena not surprisingly have led us to a large number of those who need help seeking it from consumer chatbots leveraging large language models. The mental model that we've been using here in this space focuses on the differences introduced from the clinical approaches, where a clinician could be in or on the loop, as it's called, using these autonomous agents to improve decision support and ultimately going fully autonomous, versus those where a general purpose platform may be the source. And that's really much more of this consumer context. And the guardrails that are introduced by technologists for each of these distinct domains share some language, particularly around risk-based models, but are separated by context. So, there's a level of sycophancy that you could see in the consumer world, it's almost required by design, not really a bug, that would lead to untoward effects in the clinical realm. This lack of standardized evaluation and monitoring, even in clinical, also introduces a way that we are missing the potential to understand the scale of the risk that's introduced, particularly based on the clinical condition that a person might present with or how it is that we may be able to build the appropriate guardrails going forward.

00:34:16 And so, to that end, we've been really fortunate here at Verily, one of the alphabet companies, to sit amongst other companies that have been leveraging AI tools to mimic human behavior. And that is where one of our sister companies, Waymo, is in the forefront of autonomous vehicle development and deployment. And we've watched and learned with great interest the way that autonomous vehicle industry has adopted a standardized language, something that's really missing right now in the use of generative AI for clinical applications

00:34:47 So, fortunately, there were already literal and figurative guardrails around driving that can be leveraged in this space. And the Society for Automotive Engineers built a taxonomy with definitions around automated driving back in 2014, so, well before there were any of these cars actually on the roads still being developed. But this helped technologists, regulators, payers, other stakeholders, consumers even, to have clarity around what was being proposed. So, on the left-hand side of this panel, [what] we can see is the outcome of this SAE document, it's J3016, which introduced six levels of autonomy. A three on the top is shown where the human driver is assisted with varying levels either from no automation all the way to partial automation, where hands are still on the wheel, so to speak. And then three-- Levels three through five, where there is conditional automation all the way to full automation without any boundary conditions whatsoever.

00:35:49 And this idea of having a uniform context, the language to speak to what these different tools between the sensors and the application and the safety of these tools, can be borrowed directly in medicine. So, we think of the idea of helping people through wellness and lifestyle choices almost as a coaching element. So, making the right daily wellness choices could be considered a direct-to-consumer application, but one that really allows a person that doesn't have-- There's no interface that they're missing the context. So as a consequence, they may be able to make these decisions based on these tools more directly with a coach, and there didn't have to be necessarily a clinician in the loop to validate.

00:36:46 But as we go more and more toward preventative care guidance or chronic disease management, the ability to track and manage those diseases and to come up with treatment plans are really seen more as an assistant to a clinician. And so, just as we would say level zero to two could be done with a person still hands on the wheel, here we see an ability to really influence decisions and discussions that people might have around their wellness all the way to their health with a clinician helping them make that decision.

00:37:21 But ultimately we believe that there is a utility for having these tools be more fully automated, right? So, within the context as we discussed on the previous slide, it is quite important to be able to take a triage and symptom evaluation information directly using these interfaces and potentially support a person directly in making care and diagnostic decisions where this can both achieve a greater democratization and utility of healthcare decision making and allow a person to have better health outcomes ultimately.

00:37:58 But to get there, we believe that this type of autonomy, the ability to go in all terrains and to use the sensors and the symptom evaluation triage, requires a new taxonomy of the types of interfaces, right? So, reducing the sycophancy, being able to have a much more clinical interface for people, and to ensure that we can do so safely involves the utility related much more closely to what are the exact guardrails that a mental health practitioner would use today in providing care for the same context.

00:38:32 And so we believe that this stepwise, going from left to right, starting with coaching, having people start to trust the chatbots, in a way, to make lifestyle decisions all the way to medical decisions with a clinician overseeing that-- Almost as a clinician extender, can help to inform what the ultimate guidelines look like. That would both allow us to build evaluation tools and to drive better safety evaluation ultimately.

00:39:03 So, the point of this presentation really comes down to the fact that there's a great need for mental and behavioral health practitioners. We think that we can augment them and make available to people who have this need through a chatbot interface or maybe some of their more novel speech interfaces, a tool that will help them to guide to the best outcomes. But we need a common taxonomy. And so, we could borrow from the space of automotive engineers who looked at this for autonomous vehicles and talk about these different levels of autonomy and where is it that this becomes a clinician decision support tool or does it go direct to an individual who may need care and help to support them in making those right decisions that come about. Beyond the taxonomy, we're able to then introduce the appropriate approaches to evaluation for safety, for-- Evaluation for efficacy. And ultimately to build the appropriate guardrails, but only after we've been able to talk about what levels of autonomy we're able to achieve and where is it that each of the individual tools really plugs into the system, everything from a coach all the way to a fully autonomous doctor.

00:40:20 So, thank you again for the opportunity to present some of our thinking in this space. I really look forward to seeing how this panel thinks through some of these really important topics for our society and look forward to any questions that you may have. Thank you again.

Speaker- Topic 6: Patient-Provider Considerations in the Use of GenAI-enabled Digital Mental Health Medical Devices

00:40:35 Ms. Trainum: Good afternoon to the FDA staff and members of the Advisory Committee. My name is Brooke Trainum and I'm the Senior Director of Practice Policy at the American Psychiatric Association. The American Psychiatric Association is a national specialty to society that represents more than the 39,000 psychiatrists and their patients.

00:40:52 Digital therapeutics have long played a role in the evolution of healthcare, including mental health services. The proliferation of generative AI has changed the landscape not only for clinical practice, but for physicians and patients alike. Today I'm going to discuss the patient-provider considerations and the use of GenAI digital mental health devices. Over the next 10 minutes, I will explore some of the most critical considerations in the application of generative AI in mental healthcare. I'll begin with a high-level overview of how GenAI is being used in the mental health today, and then dive into patient and provider concerns, explore the ethical overlap between them, and finish with recommendations for oversight and policy. GenAI can be transformational for many industries, including parts of healthcare, but mental health treatment may need special considerations prior to the approval of models that interact with the patient or the patient's data in any way.

00:41:52 Like other medical societies, AI devices including GenAI in mental health practice have seen an exponential rise in the past few years. GenAI-incorporated tools have been augmenting providers in uses such as clinical decision support, offering alerts or diagnostic recommendations. They've also been implemented as administrative support tools including apps, ambient scribe tools and backend management systems. Finally, where we see the greatest concern for patients with GenAI is in the use of direct engagement with patients looking for therapeutic services. At best, these technologies can improve procedure, but there are many considerations of how they impact workflow, ease of integration and privacy and security and use. However, they're not without costs, not just financially, but in terms of risk and safety.

00:42:41 Let's start with the center of care, the patient. As we are all aware, there's a mental health crisis in this country and there's a shortage of providers. Psychiatrists and other mental health providers need every tool that is available to them to meet their patients where they are. However, the increasing access to generative AI models promise more than they can deliver at this time, even causing harm to those people who would benefit from other proven treatments. People are investing time, money, and hope into something that's being touted as a treatment option and missing out on better options with science backing.

00:43:15 To be clear, there is no GenAI model that can replace the relationship between the patient and the provider. This relationship is built on empathy, consciousness, and a shared understanding of life experiences. GenAI tools can simulate some of these human traits, but truly lack the ability to handle complex emotional, ethical, and clinical situations. We are already seeing examples of parasocial relationships between patients and generative AI chatbots. Providers need to be aware of how to talk to their patients about these models. These parasocial relationships can lead to an increased loneliness, negative self-perception, and even pathological behaviors. However, the physician cannot be the only oversight here. There's a role for regulators and there needs to be more restrictions on these models, but there is a role for physicians in being able to discuss how AI tools being used in the clinical space, including the privacy and transparency considerations, the intended uses and training model data, and how to use safely under the care of the provider, are used outside of the therapeutic relationship.

00:44:22 I've started to highlight some of the physician's role in the care relationship when GenAI is being used. The physicians are facing a unique set of challenges when considering the implementation of these models. Without more oversight and regulatory guidance, the physician is bearing the greatest responsibility of safety and liability, which should not be the case. Regulators must shift the responsibility back to developers before these tools hit the market. Developers must be accountable for the safety, accuracy, and ethical use of these models, not just the providers. Before any tool is implemented, physicians consider what type of clinical oversight and who that responsibility falls onto. While large health systems may have teams to assess models, solo and small practices are often left at the tools themselves, adding to their already significant administrative burden.

00:45:10 Then there's a question of workflow integration. If a GenAI tool adds steps or increases time per patient, adoption is unlikely. These tools need to seamlessly integrate into existing processes, not to create more work. And let's not forget about the costs. Not just the financial cost of buying and maintaining the tool, but the time that's needed for training to use it effectively. For many outside that large system, these are not minor concerns.

00:45:37 For both the patient and provider concerns meet at this ethical domain. We must begin from the foundational principles of clinical ethics that the patient-provider relationship starts with respect for autonomous decision making through informed consent. Next, there is a responsibility for the physician to do no harm, in which both safety and transparency are paramount. Third, that care is being provided requires a balance between potential benefits and risks, and physicians are using tools that are clinically effective, not just engaging or novel. And finally, care must be equitable, fair and free from bias. The patient-provider relationship hinges on the trust that is built between these two parties.

00:46:19 As noted earlier in the presentation, AI applications in healthcare, and more specifically mental healthcare, are not new. However, the risk and harms have grown with the continued development and application of these models, the model's ability to digest large amounts of information and leads to less transparency, biased interpretations, confabulations and potential failures.

0046:42 Generative AI models run the risk of providing misleading and harmful information to the user. We've seen examples of this that are real and have severe consequences in the news for a while. These models are often designed to be engaging, agreeable and affirm beliefs rather than apply the human empathy, consciousness and evidence-based signs that are required in clinical care. We need safety protocols to ensure patients are directed to mental health professionals before harm occurs.

00:47:10 For patients with preexisting vulnerabilities such as psychosis or delusional disorder diagnoses, GenAI models that generate confabulated content or validate distorted thoughts may act as triggers worsening these conditions. However, for those without a specific diagnosis, more research is needed to understand the risk factors of developing AI psychosis through interaction with these models.

00:47:32 As of today, there's no high-quality evidence that these generative AI models can deliver mental healthcare. We need larger sample size with appropriate controls and standardized metrics or benchmarks to properly determine the efficacy of generative AI applications in mental healthcare. We need to specifically evaluate the technical development, clinical and ethical components and how they need to be applied in the legal and regulatory structures. Research must center on the human experience rather just than on the capabilities of the technology. Transparency and the evidence in its application in clinical care need to be available. Identifying the safeguards in place for high-risk activities these models are being implemented in.

00:48:15 The risk and the need for continued evaluation has led us to require more regulation for these models that both claim to have wellness or medical applications. Consumers are confused or unaware of the differences, and in both they're sharing highly sensitive, vital and health information with no clear transparency; this information is not protected or secure. We also know that parents feel disadvantaged and helpless when their children are turning to these models and there's no one to protect them. There needs to be enforceable frameworks that address the transparency of these models, requiring humans in the loop and risk ratification based on the intended use in the patient population. Finally, postmarket surveillance and the response protocols are essential. Just as we monitor drugs for side effects, we must monitor AI models for unintended consequences.

00:49:04 AI has the potential to help us achieve gains in clinical practice, and we must make those gains. But we must have safeguards for the lowest risk models such as administrative tasks, and restrict implementation at this time for high-risk models that are causing harm. AI tools require human oversight.

00:49:22 APA puts forth the following recommendations for this Committee to consider thoughtful policy and regulatory solutions. There needs to be standardized labeling requirements on all AI tools. Information such as model identifiers, indications for use, validation studies, training, data transparency, limitations and warnings, privacy and security and contact information need to be available. Guidance needs to require patient-centered design and evaluation, as well as encouraging co-development with mental health experts and diverse patient groups. Current models are trained and evaluated on small non-diverse populations, and we cannot apply those findings more broadly at this point. Furthermore, there must be clinical oversight for high-risk tools and prevention of public access where the risk is too high. We encourage continued research on efficacy and safety, focusing on clinical rigor. And finally, clear guidance on postmarket surveillance, especially for unintended consequences. Above all else, we must ask whether these models improve care or do they risk eroding trust and safety.

00:50:25 In closing, generative AI may one day become an invaluable tool in mental healthcare, but it’s not a replacement for the patient-physician relationship. It is a tool. And like all tools, it must be safe, effective, and ethically used. All of this requires ongoing dialogue among stakeholders, including policy experts, technologists, ethicists, healthcare professionals and patients to ensure the responsible use of AI in mental healthcare. I thank you for the time today and I look forward to answering any questions you may have.

Speaker- Topic 7: Payor Perspectives on Digital Mental Health Technologies

00:50:56 Dr. Karlin: Good afternoon and thank you to the Committee for the opportunity to participate in this important meeting. I've been asked to share a bit about payor perspectives on digital mental health technologies and what drives decisions around evaluation, adoption, reimbursement. I currently lead work in breakthrough innovations in behavioral health at ARPA-H, a sister agency to FDA and NIH, focused on realizing transformative change in health, though my perspective reflects two decades of experience leading behavioral health in payor and provider systems, both in the private sector as well as the public sector. Payors are undoubtedly a critical part of realizing the extraordinary potential of digital mental health, though with no less than 20,000 solutions in market, the signal-to-noise ratio is quite low and can be dizzying for payors to make sense of. So, my charge today is to highlight how payors think, what they look for, what gives them confidence, and what facilitates or impedes adoption.

00:51:54 So, first and foremost, payors are keenly interested in what solutions will yield. They're buying the outcome, not the app. During the days of the pandemic and ensuing years, payors saw mental health climb from number nine to number one on list of top 10 health needs of their membership. Access was a priority to ramp up capacity to align with unprecedented needs. Not to mention [that] lack of measurement made it nearly impossible to distinguish among solutions and the quality of those solutions. Now, payors are more sophisticated, and keen decision makers look for how solutions will move the clinical needle, not just from internally valid RCTs, but from real-world evidence.

00:52:37 Reported clinical outcomes, though, are not enough on their own. Engagement is also key and king to payors in many respects, especially given the eye-opening stat that many are likely familiar with and certainly is eye-opening to payors that 4% of those who initially use a digital mental health solution continue to do so two weeks later. And unlike more legacy digital mental health solutions, GenAI-based solutions that are now increasingly becoming available provide for more personalized, dynamic and multimodal capabilities that result in more durable engagement relative to more static, repetitive and non-interactive solutions.

00:53:14 But GenAI and algorithm psychology, of course, can result in a different engagement problem. Engagement for engagement's sake, if you will. Yet moving forward, I think it provides some useful lessons for driving engagement with clinically robust solutions, incorporating active ingredients of psychological change and is an area of important focus to my team here at ARPA-H.

00:53:40 So, beyond clinical and engagement impact, the next question is financial impact. The holy grail. Will a solution result in service or cost offset measurable by total cost of care reduction? For years, the best response has frankly been an inferential one. For example, key findings from Milliman that individuals with the behavioral health condition have three to six times higher total annual healthcare costs than their non-behavioral health condition counterparts. Direct data, though, on cost effectiveness are limited really due to two chief factors. One is the lack of clinical, let alone economic, measurement. And two, because the time horizon for ROI for mental health is much longer than for physical health. So, for example, if you put a stent in someone with cardiovascular disease, you'll see the ROI almost immediately by keeping them out of the hospital. For mental health, the ROI is often there, but may take 2, 3, 4 years to realize.

00:54:38 Now, while slower to come to fruition in mental health, payors are now increasingly interested in how solutions can be used as part of value-based care, including the entry level, if you will, of value-based care, pay for performance or pay for measurement. And at the other end of the spectrum, risk-based models. Value-based payment models are actually quite more common in Medicaid plans today where there is increasing innovation but are slowly increasing in commercial plans.

00:55:07 So, next are what I call “How” factors. How, for example, a solution is designed and operates, that govern whether a solution succeeds and scales in the real world. First is member experience. Notably, the influx of adolescents to GenAI-based tools is a reflection of the fact that many teens don't view non-GenAI solutions with repetition, limited creativity, limited interactivity, as particularly relevant to them. And second and deeply, deeply important to payors is integration, both technical and clinical. The former includes how well a solution integrates into technology pathways including, but ideally, more than single sign on is important, though kind of aligning of electrons isn't sufficient. Also, key is how solutions are integrated clinically and with minimal burden. So, blended or guided approaches, as sometimes referred to, generally yield deeper engagement and clinical outcomes than standalone or self-guided approaches for those with more mild to moderate needs. For those with minimal level severity, there's typically no difference.

00:56:22 Over the past couple of years, there's been positive coverage developments focusing primarily on prescription digital therapeutics with FDA clearance used in a blended context for the most part. And notably, I was involved in developing one of the first such policies and felt strongly that this include a broader definition of prescriber that includes all licensed independent mental health providers, given that digital therapeutics typically focus on psychosocial interventions, and there's certainly a cadre of experts across disciplines, with respect to those, and fully standalone solutions incorporating conversational agents and related capabilities are not covered, though this will be and is becoming an area of active focus. I'll also say more about opportunities outside of traditional fee-for-service models in a moment.

00:57:12 So on the one hand, legacy solutions often with strong clinical content are covered but often have low engagements. And on the other hand, we have LLM-based solutions that are hyper-focused on engagement and not covered. So, I think the future will be a hybrid of the two, leveraging unique sets of expertise that often don't come together or haven't historically, and is an important area of innovation interests within my portfolio.

00:57:39 So, trust and responsible development and use will certainly increasingly be important areas of focus, especially with the advance of GenAI and increasingly sophisticated solutions. And this is a summary of key privacy and ethical considerations in development and use and which we use to actually guide our current work.

00:57:59 And last but not least, are “When” factors. So, savvy payors are increasingly interested in targeting timing and fit as opposed to the prevailing one-size-fits-most approach, often reactive in nature. Relatedly, payors are increasingly focused on population health, including proactively identifying and stratifying risk and moving more upstream, or to the left, if you will, on the clinical care continuum, which is a zone of missed opportunity for mental health for decades. Moving, interestingly, beyond the inferential, we recently published a study based on claims data across almost a million members, empirically demonstrating that changes in depression severity, up or down, are associated with corresponding changes in total cost of care. So, there does indeed appear to be a financial incentive empirically shown to keeping folks emotionally well.

00:58:51 Payment model opportunities for this more upstream focus are often not traditional fee-for-service models, but PMPM models where broad member access to solutions may be made available at approximately a dollar or less per member per month, PMPM, to support emotional health and prevent symptom escalation while identifying and triaging those with more significant needs to clinical services. This, I think, will increasingly attract attention of full-stack digital virtual behavioral health companies for building out center-left capabilities.

And in fact, within the last week or so, a major full-stack behavioral health company announced how it's bringing GenAI into this center-left opportunity fold.

00:59:33 So, to sum up, payor considerations boil down to what I have depicted graphically here. But I'd like to offer a few final priorities related to GenAI-based solutions as I conclude. First, there's need for integrating more meaningful escalation protocols. Second, payors often look to regulatory authorities as trusted sources and seek greater confidence related to functionalities and claims that require premarket review, especially given the state of wellness-based solutions, evidence and privacy thresholds, and postmarket monitoring as AI models evolve, for example. And without some of this clarity, I think coverage is likely to be more hesitant or slow moving. And lastly, it's important to be mindful that the landscape is shifting in the industry toward market consolidation, something I believe will be favorably on the whole received by payors and increase adoption. In fact, payors are increasingly moving away from fragmented vendor-only relationships to more partnership opportunities with full-stack or mature companies.

01:00:39 So, in closing, I truly believe this is a pivotal and defining moment for the field of digital mental health and mental healthcare more broadly for that matter. And by being mindful of the what, how, when, of course, why factors and bringing together payors and other stakeholders in the fragmented ecosystem of behavioral health, I believe we can unlock considerable unrealized potential of digital mental health. Thank you very much.

Speaker- Topic 8: Peterson Health Technology Institute Evaluations of Virtual Solutions for Depression and Anxiety and Opioid Use Disorder

01:01:04 Ms. Pearson: Good afternoon. I'm Caroline Pearson. I'm the Executive Director of the Peterson Health Technology Institute, and I really appreciate the opportunity to share a little bit more about our work and our findings with you today.

01:01:16 For those of you who don't know us, the Peterson Health Technology Institute is an independent nonprofit philanthropy, and we conduct assessments of digital health solutions to inform purchasing decisions among health plans, employers and health systems, with a goal of driving greater adoption of high-value health technology and digital health solutions. We publish assessments based on certain categories of digital health tools, and we have published five to date that you see here. All of our materials are available for free on our website at PHTI.org. And today I'm going to speak more specifically about our last two most recent reports focused on virtual solutions for depression, anxiety, as well as interventions for people with opioid use disorder.

01:01:58 To tell you a bit more about what we do at PHTI, we really try to choose categories of health issues that have significant disease burden and that generate high avoidable healthcare spending. And then we're also looking for places where there's a lot of digital health innovation that is coming to the market and bringing new potential solutions for scaling access and improving health outcomes. We conduct our assessments by reviewing published evidence, secondary evidence. We don't do any of our own data analysis. And we're really looking at two core issues, clinical effectiveness and economic impact, with a goal of answering these three questions. Do these solutions work? For whom, for which patient populations? And if they do, are they worth it? And we synthesize that information with targeted recommendations for purchasers around adoption and further evidence generation needs.

01:02:49 We work closely with a range of partners and advisors throughout our process, including clinical advisors who are recruited based on particular expertise for each of our reports. We conduct direct research with patients to understand their care journey and how digital health solutions can improve their usual care experience. And we work directly with the companies that are part of the assessments to give them the opportunity to submit additional data and evidence for our review.

01:03:14 So, in May of this year, we released our assessment looking at solutions that are used to treat depression and anxiety, specifically looking at people with mild to moderate symptoms. We know today that more than one in five adults suffer from symptoms of anxiety and depression, and many of these people do not receive treatment due to provider shortages, cost barriers, and social stigma. And we've seen a tremendous growth in the number of solutions that are really using digital interfaces to deliver online content that can be self-directed, and it is based on CBT and other therapeutic principles, sometimes augmented with live human support in the form of coaching and additional therapists or psychiatrists as needed.

01:03:58 Our report looked at a group of 15 companies, and you can see them here, and they're grouped into these three different categories based on how they are sold and how they work in the market. Self-guided solutions really are tools that create online platforms with digital content, whether that's lessons, exercises, other interactives that people can do to address their anxiety and depression symptoms. They are generally sold directly to employers and health plans, and they're often alternatives to existing employee assistance programs, EAP programs.

01:04:31 Prescription digital therapeutics have a similar user interface in terms of creating digital content, but they have gone through a more rigorous FDA clearance process and they are delivered to patients with a prescription often from their therapist to really augment other mental healthcare. And then finally, we have a category of what we call blended-care solutions. Each of these solutions includes all of the self-guided content that the other two categories have, but they also add networks of therapy providers and potentially medication prescribers to treat people who may prefer or may require more intensive therapy-based solutions rather than simply self-guided content alone. This is really where the bulk of the market is and where we've seen the greatest amount of adoption by employers and other purchasers.

01:05:22 Now, when we do our work, we are generally trying to look at clinical evidence around a series of primary and secondary outcomes, which are included here. For this review, we reviewed over 5,000 articles, and our assessment is really based on 103 robust studies that include both RCTs and real-world evidence, trying to look at improvements in depression symptoms and anxiety symptoms based on validated scales of PHQ-9 and GAD-7, and focusing not just on statistically significant improvements, but really what's defined as clinically meaningful improvements for each of those scales, which is a 5-point reduction in symptoms for depression and a 4-point reduction on the GAD-7 for anxiety.

01:06:06 Now, as we look at the results for this assessment, what we need to do is first consider some of the control arms. So, I will flag that much of the evidence that we looked at here has relatively short study periods. So, generally we're seeing studies of six to 12 weeks. So, I think there's ongoing questions about durability. And we also have to recognize that for people with mild to moderate depression and anxiety symptoms, we see a reasonable amount of self-resolution. So, it's important that we're looking at these gray bars for control arms to see how-- Whether it's a study that's designed with a wait list control, or whether it has a placebo or a sham app that's being given to users, you do see that you have modest symptom improvement over time, even in the control arms of the studies.

01:06:47 Now, if we look specifically at people who did not previously receive psychotherapy and effectively were untreated before they received their digital solution, and so that's the top of this slide on the left and the right for depression and anxiety, you can see that both the self-guided solutions and the blended-care solutions are clinically effective at delivering MCID-level improvements in both depression and anxiety symptoms, with the blended-care solutions delivering slightly higher improvements in each of those scores.

01:07:17 Now, if we compare that to people who are receiving usual care, and that can be a combination of therapy and other medications, you see that actually the self-guided content in addition to therapy or medications do not deliver clinically meaningful benefits. But when those solutions are being integrated, as a prescription digital therapeutic would be, with clinical care, when you have a provider guiding you to use those solutions, you see that both the PDTs and the blended-care interventions continue to deliver clinically meaningful improvements even for people who were previously in usual care. So, these are very promising clinical results.

01:07:51 So, what about the cost? We know that self-guided solutions are generally sold at relatively low prices, for about $2 per member per month, and so, because they help improve symptoms and reduce other healthcare spending, they generally deliver net savings to healthcare purchasers, particularly shown here in the commercial market. The same thing is true of the prescription digital therapeutics. While pricing is still being set, particularly by most of the Medicare contractors, if we estimate, in this case, that we're going to have about a $200 device reimbursement plus an $80 management fee, those solutions also reduce net health spending.

01:08:27 The blended-care solutions, however, tend to be priced much higher. They're generally being sold at about $6 per member per month. That's not per user, that's for every health plan member who has access to the tool. Plus fee for service charges for additional therapy. And at those prices, they are not sufficient to clear the costs. The user savings do not sufficiently offset the cost of the solution, and you see that we estimate that they have a net increase in overall spending. That comes together into our aggregate-- Assessment results where we see clinically effective results for all of the solutions. But you see that red dot on economic results for the blended-care solutions, meaning that this is a place where we need to think about appropriate pricing in the market.

01:09:08 Now, if we look at our review briefly for opioid use disorder solutions, what you see here is that we've got a number of companies delivering medication-based treatment supported with additional virtual wraparound solutions like therapy, peer support and contingency management rewards, to keep people engaged with their treatment. The companies on the right are meant to partner with medication-based treatment, but they did not include the prescribing of buprenorphine or other medications.

01:09:34 We looked specifically at outcomes tied to retention to treatment, which is what was recommended and is widely seen as the most protective way to help people stay in recovery and avoid overdoses. And what we found is that virtual care for opioid use disorder treatment is clinically effective, and we saw results that were really comparable with in-person treatment options.

01:09:57 What we did not see was that these solutions deliver significant increases in overall retention. On average, the evidence suggests that you get 13 days more treatment retention on average over six months, which is an important improvement, but certainly modest in the realm of the magnitude of the opioid use disorder. We also do not see evidence that at this time most of these solutions are broadly expanding access to treatment. So, while they make treatment more affordable and convenient, we have more work to do to help connect digital solutions to people who need them at the moment that they may be ready to embark on their care.

01:10:30 That all comes together into our overall recommendations, which suggests that these solutions are clinically effective. Medication-based solutions are sold at comparable rates to traditional usual care treatment, and thus they warrant broader expansion as an option to help more people access OUD care effectively. The digital wraparound solutions tend to add costs and don't seem to improve retention dramatically. And so, those warrant a more judicious approach as we think about expansion.

01:10:58 Thank you so much for the time today and I look forward to your questions.

Open Committee Discussion

01:11:07 Dr. Bhatt: Great. Thank you so much. I hope everybody else’s brain is hurting as much as mine is. I'd like to thank Drs. Jacobson, Arean, Trister, Karlin, Ms. Trainum, and Ms. Pearson for their presentation. Now, we'll have the Open Committee Discussion. We'll have clarifying questions from the Committee. As a reminder, although this portion is open to public observers, public attendees may not participate except at the specific request of the Committee Chair. Additionally, we request that all persons who are asked to speak identify themselves each time. This helps with transcription.

01:11:41 I've been asked to give a short summary of what we covered just a little bit before the lunch hour because our question and answer session covers our current speakers as well as prior. So, in the morning session, we had conversations about risk in the human element of discussion. We heard from Screen Time Clinic asking for decisive action for overstimulating addictive reward-seeking screen exposure, comparing it to strongly regulated protections such as underage drinking. We heard about the risk of tech being human-like, especially in individuals who struggle to differentiate between fantasy and reality. There were requests to avoid mental health that is autonomous in children until we know much more. And then we heard directly from our younger generations about how important mental health support is, and that while emotional support in dealing with stress is essential, so is the actual diagnosis of mental health conditions and how scientific rigor matters to our younger generations as much as access.

01:12:34 We then learned about potential metrics and solutions to some of these challenges, including voice biomarkers, scenario-based testing, the need for developers to demonstrate that they engaged clinicians and patients, and a call for human-centered transparency that is proportional to risk, as well as guidance for clear language for end users of generative AI. We were then reminded by physicians that depression and stress ebb and flow over time, restating the importance of control groups and studying efficacy. Finally, there was a request to address validated behavioral health symptom measures, functional behavioral health outcomes, patient and caregiver motivational outcomes when looking at chatbots. And then most recently our academic perspectives were critical trial evaluation, lessons learned from other fields of AI, our payor perspectives from ARPA-H, and finally, the results of evaluations of virtual solutions for depression, anxiety and opioid use disorder.

01:13:27 All of the above is up for discussion. And we'll start with John Torous, please.

01:13:38 Dr. Torous: This question is geared towards our last speaker on the evaluation of the depression and anxiety tools. I think reflecting on what Dr. Pat Arean said about the need for important control groups, and reading this report, I noticed that a lot of the comparisons of digital tools were to waitlist controls or people being asked to do nothing at all, to take no improvement on their health. And clearly as a physician caring for people, I know something is always better than nothing. So, I'd like to ask what do you think is the appropriate control group and would your group be willing to go back and redo this analysis for us to show us what happens when we have digital control groups and more appropriate concerns? Because I worry if we say these are cost-effective compared to nothing, walking is an effective treatment too. We need to make sure that we are comparing apples to apples.

01:14:25 Ms. Pearson: Thank you for the question. We have all of the data parsed for every study with every control group specified, so I'm happy to cut it for you. However, it is most useful. Absolutely. I actually think you don't see completely consistent patterns if you try to separate out by different control arm types. We actually did do that exercise and we're happy to show you that. There's just a good amount of variability based on study design. But I guess I would underscore your point, which is, again, you do see the greatest benefit for people who are not otherwise receiving treatment. And so, while our report was not specifically focused on chatbots, I think that's an appropriate frame to bring to this conversation, which is not only thinking about any of the digital mental health solutions as an alternative to psychotherapy from a live human, but how do we bring more people into mental health treatment in ways that are affordable and accessible and favorable for them.

01:15:22 Dr. Torous: Thank you.

01:15:24 Dr. Bhatt: Thank you, Caroline. Ray Dorsey, you have the next question.

01:15:28 Dr. Dorsey: Thank you, Chair. Thank you for all the great presentations. My question is for Dr. Jacobson. I have two questions. How do you handle the recorded conversations you have for individuals participating in your trial? Do you use it to make your model better? Are you sharing it? What security measures do you have in place? And second, if you review these conversations, how do you assess the appropriateness of these conversations? Your clinical trials showed benefit on the outcome measures, but I didn't see if you were able to assess the appropriateness of the conversations related to that. I couldn't access the article as a whole. Do you have a link to the article that we could review as a whole? That would be helpful.

01:16:10 Dr. Jacobson: Yeah, thank you. In terms of-- Sorry, this is Nick Jacobson. Wonderful questions. We monitor every message sent to or from the chatbot in near real time during the course of the trial. And this is in part just to try to make sure that it's actually implemented in a way that could be safe and effective, and if there were any human intervention that is actually needed that humans could intervene. In terms of how we handle that, it is a research-grade HIPAA compliant server. The data is never used to train our models. We certainly use to see where we are learning from the failures of things where we could be doing better and every iteration and interaction that we have with it internally, but we don't use that directly to train data, in large part because that actually keeps things in an entirely separate sandbox, so that there's no actual privacy constraints in terms of the data's never sent to an external party. It's never possible for that data to then be somehow entrained within the model and in some ways violate patient or participant confidentiality. So, a lot of those kinds of measures we take just to try to make sure that there isn't any concerns related to privacy on the model training end.

01:17:28 And then in terms of the appropriateness of the responses, we did monitor those within the course of the trial. We intervened when we felt the model was being inappropriate. One of the things that-- It was a common scenario that did occur, and in retrospect, it's one of those things that, "Oh, it's like, of course, this should have been dealt with upfront." But we had the model-- We really framed it specifically within the bounds of psychotherapy, but folks would ask it for general medical queries and we had a number of folks that asked it for assessment of their medical symptoms, and so it would actually provide and try to help, suggest recommendations surrounding those medical symptoms. So, we've followed up with those participants suggesting that they go actually see a physician surrounding those issues. But we also monitor things related to active suicidal intent and passive ideation and things like this as well. And would intervene with escalations and care, but those were not because the model didn't behave appropriately during those times.

01:18:42 Dr. Dorsey: Thank you.

01:18:42 Dr. Bhatt: Thank you. Joyce Ho, I have you next.

01:18:46 Dr. Ho: Thank you. So, this is Joyce Ho. This question is for Dr. Becker. You talked a little bit about metacognition and potentially using it to probe the-- Or to evaluate GenAI. I'm curious as to how you might do this in the context of mental health or how you might expand it to mental health.

01:19:09 LCDR Becker: Well, this is Dr. Becker responding this question. I would mostly point to the work of Maxime Griot. That was the study that I was citing there. He's had another follow-up study where he's looked at creating essentially notional medical data and showing that a lot of the benchmarks that are prompting this urge towards the implementation of these systems, the performance that we're measuring seems to be mostly a feature of the fact that these systems are really good at test-taking skills, but not necessarily over reasoning the content-- Reasoning over the content that they're learning. So, I point to his work. As far as how that's occurred in the clinic for me, I think what I have seen-- I've not encountered this phenomenon of AI psychosis as a clinician myself, but I've definitely encountered the phenomenon spoken to by a lot of the public comment regarding patients developing parasocial relationships even in cases where explicitly patients were seeking to maybe increase their relationships with others or seeking more intimacy. I've seen patterns where people engage so deeply with these systems that it's to the detriment of actually going out and interacting with other people. And so, in that way, that's sort of a metacognitive failure and so much as the patient has this stated intent to connect to others but can't see beyond that, that they're-- By trying to prepare to do that with the help of an LLM, they're not succeeding in that.

01:20:56 Dr. Bhatt: Great, thank you so much. Wayne Goodman, I have you next.

01:21:00 Dr. Goodman: Yeah, this is Wayne Goodman. I have another question for Caroline Pearson. And, again, referring to-- I think you showed that blended-care for depression had advantages and that one of the outcome measures for depression that was deployed was the PHQ-9. My question is, do you recall if the PHQ-9 was embedded in the chatbot or it was administered separately? It's a self-rating form. I'll tell you why I'm asking the question. It’s because the last item nine is about suicidality and if we see anything other than a score of zero on it, that usually warrants some evaluation of a suicidal risk. And I'm not sure how you would do that promptly and effectively if it was within the chatbot.

01:21:50 Ms. Pearson: Yeah, that's a great question. I think it's a mix across the studies that we looked at. Most of the solutions do have sort of an automated PHQ-9 that they're using. So, generally in some cases they can load it from a direct human-centered therapeutic interaction, but usually it's self-administer or digitally administered. So, your point is well taken. I should flag that we were really specifically looking at people with mild to moderate symptoms, so that is a little bit less of a concern in this population, typically, and in fact, some of the studies did use PHQ-7, which from mild to moderate is somewhat more comparable. But I think the issue is real and speaks to part of why the blended-care solutions, I think, have really been more appealing within the market is the idea that you can triage patients more completely. Right? I think employers and health plans have had a lot of concern about self-guided solutions that, while clinically effective for some patient populations, don't give them an option to escalate care effectively in cases where they need it. And so, that's where you see these blended-care solutions having that advantage of being able to pop someone over to real-time human help.

01:23:08 Dr. Goodman: Thank you.

01:23:09 Dr. Bhatt: Thanks, Caroline. I have John, then Jessica, then Chevon is my order. So, John Torous.

01:23:15 Dr. Torous: This will be a brief question for Dr. Jacobson. In reading the paper about Therabot, I was unable to find a written safety plan that you and the study were using in the appendix. Is there a way that you could share what was the protocol for safety given when talking about regulating apps and understanding what are the safety things put in place? Or if there wasn't one, was it just one person doing what they thought was clinically best?

01:23:38 Dr. Jacobson: In terms of the plan, we had human oversight of every message in terms of what would happen if folks would send a response that-- There was, I guess, several different solutions to this. One is that we have a risk classifier model that was part of that, that would indicate a crisis button that began flashing that would direct to 911, and also the National Suicide Hotline and Crisis Text Line. In terms of if your question is around the human intervention that is also part of the oversight of the trial, we monitored every message sent to or from the chat bot in near real time. We did exercise clinical judgment as part of that as to whether the degree of risk, but when there was really any discussion of suicidal intent and/or even ideation, we did follow up with folks to conduct a risk assessment with those folks from a clinical psychologist or psychiatrist.

01:24:35 Dr. Torous: I guess I'd say it would be useful to know what was the risk classifiers; it's never talked about. Or to say what was the training, what were the procedures. Because else, it's not scalable.

01:24:46 Dr. Jacobson: It certainly-- It's something that we've developed within the system. It's based on essentially a direct classification of risk from a large language model approach, but it's not a generative large language model. And a lot of the goal is to try to classify whether there is some degree of risk across many different psychiatric classes.

01:25:09 Dr. Torous: Thank you.

01:25:09 Dr. Bhatt: Great. Thank you. Jessica Jackson.

01:25:13 Dr. Jackson: Thank you. This question is for Dr. Patricia-- I apologize if I'm mispronouncing your last name, I think it's Arean. You were talking about the clinical trials and I'm curious if you have-- It's a two-part question. A, any best practices for engagement? I know historically for neuroscience and for psychiatry, there's lower engagement in clinical trials compared to other therapeutic areas. So, I think most of the people today have talked about having clinical trials for all of the new technology, and I'm concerned that if we already struggle with that. So, I'm curious if you have any thoughts about how to mitigate that, and/or if you have any alternatives to offer to clinical trials to ensure representation. I know that there was-- A part of what somebody shared earlier was who can be best impacted by this. And thinking about mild to moderate, those would be people who would tend to not engage in clinical trials if they're at mild to moderate or even, you know, what we would say "being well" in using these chatbots and if there's alternatives to clinical trials that you would recommend.

01:26:20 Dr. Arean: Right. Okay. So, there's a lot there. First off, I think that it kind of depends on how the trial is being run. If it's run remotely, you get even worse engagement over time and some strategies that have seemed to be effective have been in terms of engaging in the research project to deliver outcomes and all of that has been some form or a combination of non-monetary/monetary incentives to complete tasks. When you do that remotely, you run the risk of being gamed. And so, you have to be very-- Super careful with how you advertise that, the level at which the incentive is provided so that it is meaningful but not so meaningful that you would get that kind of challenge. If it's done as, you know, “Here, we're giving you this, we've met you,” the actual intervention is not done in person but the assessments are, then there does tend to be better engagement in the research process when you have that human contact. Theoretically, you could do that through Zoom, too. That could be another way to engage the sample. And so, I think the more that there's human in the loop, the better you have in terms of retention for evaluating the efficacy of the intervention.

01:27:44 When you get to the effectiveness realm, unfortunately part of the outcome is engagement. So, you could have a very efficacious tool and then you put it out there and you see that people don't use it. That is an outcome at the population level. It means that there was something about the intervention that made it difficult to continue using, whether it's a side effect, it's a time effect or it's a boredom effect, which is often common with other types of digital mental health tools. And so, in doing that, you have to utilize data analytic tools that account for missing this in data, looking at the population who does drop out compared to the ones who stay in the study in order to do a sensitivity analysis on your data to see how representative it is over time.

01:28:38 Another alternative are non-randomized trials. There are a number of methodological strategies you can use for that. So, there are stepped-wedge designs that allow everybody at some point to participate in the study. There are pre-randomization trials. We used that in one of our HEAL-- HEAL was a large-scale project to look at opioid addiction, and one part of HEAL was to look at the use of opioid interventions in primary care medicine. And one of the sites had used this-- I forget-- I'm blanking on the name of the type of trial, but you pre-randomize people and then you ask them, “Do you want to participate in this study?” And that tends to give you a greater more representative sample than if you approach each person as they come into the clinic or that they respond to your ad. So, that was sort of the research engagement piece. And I've already forgotten what the second question was.

01:29:43 Dr. Jackson: I was just asking about thinking about how to get more engagement for representativeness. So, I think a lot of what people were saying when we think about clinical trials, people tend to have a reason to engage. They're looking for treatment. But if-- A lot of people also mentioned a lot of the wellness apps, people were also talking about things from mild to moderate. I mean, we're talking about the PHQ-9 and things like that. How do we get those people to engage if they're not feeling like they need treatment, but they're using chatbots and we're trying to monitor the safety.

01:30:12 Dr. Arean: Right. So, yeah, in a clinical trial that is a challenge with recruitment. You might have to expand the recruitment base that you are sampling from, the population base that you're sampling from, in order to get representation, enriching the sample for certain populations that may or may not engage. There is one other thing, too. [It] is that there are certain populations or samples who you are just not going to get because the technology is not accessible to them. And so, you won't know, for instance, how useful this might be in rural Montana where we have really high suicide rates. But access to things like chatbots and that kind of thing or distance technology is really, really challenging in a lot of these rural areas.

01:30:58 So, you have to-- Unfortunately, I think that the challenge, too, and even using the data from these studies or from these interactions, is that we have to recognize that the data are coming from particular types of populations, particular types of people who are going to have certain kinds of engagement patterns and discussion patterns with the chatbots. So, it is imperative to start to think about-- And it might be worthwhile to have a conversation about this methodologically and in a different setting about how would we enrich samples so that we could get people who maybe English is their second language or there are cultural differences in how they might engage in a therapeutic conversation.

01:31:41 Dr. Bhatt: Great. Thank you. Chevon.

01:31:44 Dr. Rariy: Yes, thank you. I really appreciate the conversations here. And this question is for Dr. Jacobson, unless others feel like it's relevant to their work as well. But I have-- It's a two-part question. So, the first part is we know that generative AI in mental health-- Generative AI in general is dynamic. The models learn, the models update, they drift over time, as we heard, relative to static tools, for example. And so, the question that I have is, as we're looking at premarketing or postmarketing monitoring relative to drift and bias and safety, what would you say would be the-- One, the minimum amount of time and perhaps even the maximum amount of time, and how often should you monitor drift? And would this vary? Would your recommendations vary based on risk? And the second part of my question is, if we're looking at drift, we know that models could drift not just based on performance, but also drift based on bias, for example. And so, should bias testing be treated as part of a drift detection? I'm curious to hear your thoughts there. And who would be responsible or who should be responsible for identifying and reporting it? The regulators, the developers, independent third parties, patients themselves, users? Just curious to hear what you're thinking.

01:33:02 Dr. Jacobson: Yeah, sure. I think in terms of trying to have a capacity to test these systems, one of the things that is among the most challenging is the pace of advancement. You will have continued model retraining and things like this that are part of the system. There can be automated metrics. Speakers have talked about synthetic simulations and evaluations that could be possible in terms of developing benchmarkings, and folks were-- Would discuss that earlier. Unfortunately, I think the real answer of how these systems should ultimately be evaluated before their updates are put into practice is that humans that are experts are ultimately evaluating them, but putting that burden of doing so at every development on the FDA would be far too onerous with the pace of development. My recommendation at the end of my presentation was really on actually a way that this could be done by the developers and the folks that are maintaining it, where the process is actually maintained and that burden then falls on them to actually make sure that they actually deploy things in an effective manner.

01:34:18 FDA has actually done this in prior versions. There were complications with that in the Pre-Cert program, for example. But there is some exploration, I think, of things that, in this particular area, as a mechanism, I think there's really not a great way of doing this outside of the folks that are maintaining it and have so much time invested in it, because it takes so much time to do it. To do it well. So, that's ultimately where I think the way that this could ultimately potentially be scaled is really FDA monitoring the process, understanding how folks that are maintaining it go-- Plan to go about this. But ultimately to allow for those that are overseeing their own product development to make sure within auditing or something along those lines of mechanisms periodically to make sure that that's actually happening well. And then postmarket surveillance from there. But that'd be my recommendations in ways that actually keep pace with the technology in a practical way.

01:35:21 Dr. Bhatt: Thanks, Dr. Jacobson. Joyce Ho, we have a question for you.

01:35:25 Dr. Ho: Hi. This is for Dr. Arean. My question is regarding thinking about clinical trial design and many of the speakers brought up that comorbidities are very common in this population and how we might design something to probe that without sacrificing sort of broadness of the evidence that's collected.

01:35:47 Dr. Arean: Well, yeah. So, in effectiveness studies and in field trials, we often include people with comorbidities. And one thing that you have to consider is that, first, there's-- What's the purpose of the chatbot? Is it for general mental health wellness? And so, you meant to develop it so it could be generic and who it is working with. Or is there an indication that this is a chatbot that's designed for major depression? And so, I think it is important if you're doing an indication for a specific illness that you not necessarily exclude people with comorbid conditions where their primary diagnosis is-- Has major depression, because that is just who people are out there. When you're doing something in a much earlier phase, like proof of concept, there is a tendency to have a cleaner, quote, unquote, cleaner sample where the people don't have these other comorbidities. But that's just to kind of show, first, is it safe? Like, can we do this safely? Is there movement in what we think the target mechanism is? But once you get to the efficacy stage, I think you do need to start considering the fact that the general population usually doesn't just have major depression. There's usually anxiety symptoms or there might be some heavy drinking going on and that you document that when you sample-- When you assess your sample. So, you can report out and also analyze your data based on how many comorbidities somebody may have had and how that might influence their response to treatment.

01:37:22 Dr. Bhatt: Great, thank you. I myself have two quick questions. Nick, maybe I'll try the first one with you, which is: Is there a way as we're designing these trials, or have people been thinking about telling the user and reminding them that they're an app? I'm getting at this idea of anthropomorphizing app and trying to get to know it, and do we know-- Is there a frequency at which you say, "Hey, remember I'm a computer," that is effective in helping keep them from misunderstanding and yet be effective in helping solve some of the emotional concerns that they have?

01:37:58 Dr. Jacobson: Yeah, I guess I-- We approach this ourselves in the design process. We name it with Therabot, with the “bot” in the title to try to actually make sure that folks are understanding that it's a technology and not a human. We also have a robot that is our avatar that is front and center and always present in every interaction. And a lot of that is just to try to make sure that folks are really understanding that this isn't a human. We started developing this before ChatGPT was released. And so, the widespread awareness of this was even less saturated then. So, we've been really thinking about this for a long time, being very important that folks understand that this is not a human. And so, I think that's a lot of how we try to communicate that consistently in the interaction. So, not trying to present this as if it's a human, I think just to try to make sure that they don't have any misunderstandings along the way that it's very much always present that way.

01:38:57 Dr. Bhatt: Great, thank you. And then, Brooke, can I ask you one quick question? And then we'll go to Tom Maddox. From-- With your J.D. hat on. My car gives me signals sometimes. I don't know how the car works, but it says “The pressure-- Tire pressure is low.” And I'm not comparing humans to cars, but the same point, that's a dashboard. I'm taking my life into my own hands every time I get behind the wheel. For the lower risk-- We've all been talking about the different stages. For lower risk, where patient agency is important-- From that perspective, the patient dashboard of “Hey, there might be drift. Check in with your company,” or “Hey, you might have a signal. Check in with your doc,” do we know anything legally about that and the dashboard idea and if that'll work in medicine, especially in mental health? Or are we too far away from even talking about it?

01:39:44 Ms. Trainum: I don't know that we have seen that yet, but the one thing that comes to mind, and I'm just thinking about work with [Indiscernible - 01:39:50], is how many times we get alerts and start ignoring them right down the road, that's for anything. So, yeah, I don't know how that would play into that impact, too.

01:40:01 Dr. Bhatt: Thank you. Tom Maddox.

01:40:03 Dr. Maddox: Yeah. I appreciate the attempts to have reminders and visuals to suggest that, “Oh, this is not a human, it's an algorithm.” But I think it's important that we recognize what's real and the automation bias that people are talking about is just true. And I think it's been fairly demonstrably demonstrated, sort of redundant, that people just don't pay attention to that and they do start to anthropomorphize and don't really self-regulate in terms of how they would interact in this case with an LLM and a human. So, I guess, does it make more sense to not say, “Oh, we'll just layer on more reminders and more language,” but rather structure and regulate these in a way that just keeps an algorithm, a non-human, from doing things that we believe belong in the province of humans?

01:41:07 Dr. Jacobson: That may have been directed at me. So, I'll--

01:41:10 Dr. Maddox: Broadly, yes. But anyway. Sure. Thank you.

01:41:12 Dr. Jacobson: Sure. So, I guess one of the things that I'll say is a lot of the questions surrounding this, I guess in spirit, that folks have concerns around are, one, if they have informed consent, are they understanding what they're doing? And the second, which I think we definitely directly address via the technology development side. In terms of the emotional affinity attachment, anthropomorphizing elements, I do think within a therapeutic setting, one of the ways that they work within even any digital mental health tool is to some degree based in part on this therapeutic alliance. It's the bedrock that really works for all of psychotherapy and that would include generative AI. So, I actually think that the idea that this includes some form of a bond related to you being able to develop a working rapport and trust with the generative AI precludes whether it can actually be effective with that end user. So, I think in terms of the relationship formation and the ability to do that, if it's well-designed and not used for a blurred scenario related to companionhood or something like that, I think that's actually something that we would hope and you would see in other digital mental health products as well.

01:42:33 Dr. Bhatt: Great. Thank you. Melissa Clarkson.

01:42:39 Dr. Clarkson: Melissa Clarkson, I'm the Consumer Representative. So, I do want to stay on this topic and circle around postmarket surveillance. We kind of mix that with accountability. So, the term accountability has come up a couple times, but there's been no details or discussion really focused on accountability. Well Nick, not to throw this your way again, but maybe-- But just kind of open up in general of looking to a future where there are FDA approved regulated devices. Where does that accountability sit? Is it with the manufacturer of the device? If it's using the clinical environment with that clinician, is it on the patient themselves to say, “Oh, I should have been using it this way, not that way”? So, I think this idea of accountability is kind of mushy at this point and needs a little more discussion.

01:43:42 Dr. Jacobson: Yeah. I guess I can try to speak to that. It's a very broad question that I frankly don't have all the answers to and probably will require a number of folks to have-- My, I guess, original take is that we can't be expecting perfection from these systems. So, for example, one of the things that-- I think it would be good for folks to think about is a model of safety to some degree based on other areas of regulatory capacity that are also involved areas of risk. I think self-driving cars is actually one of the ones that involves AI and has some comparisons that are to some degree apt in terms of the benchmark for humans out there within the care system is that they don't all provide the best evidence-based treatment and provide every single response, they say is near perfect or something along those lines. So, when you think about the degree of risk, I guess that's one of the ways that I think about it at a scale level.

01:44:48 Now, on an individual level, I guess getting at that individual accountability, accountability for an individual action, I would actually put that on the developers themselves really if the tool is actually directly harmful and I think that that risk should fall on the folks that are providing and developing the service or deploying it, I suppose. As to where exactly that is when you have a developer that then contracts with another person that misuses it, I think it probably all the details seem to matter to me, but that would be kind of how I'd approach it generally.

01:45:26 Dr. Bhatt: Great, thank you, Nick. I have Andrew Trister and then Patricia Arean. Andrew.

01:45:31 Dr. Trister: Thanks Amy, and thanks Nick for that comment. So, I'm Andrew Trister, I'm the Chief Medical Officer at Verily, one of the companies in Alphabet. On this topic of accountability, I think this is a really important issue, and I do like the analogy as I presented earlier on autonomous vehicles. One of the things that's missing actually when we think about safety relates directly to this question of what do we believe the risk profiles are, as we've measured, and we say, well, we may be able to build an autonomous approach to building a risk profile. And I think that some of the questions earlier from the panel related to that, but there is this conjoined question, which is who then is accountable for this? And at the moment with technology companies going forward in this space, we know that the accountability is not starting with the groups that are building the large generative AI models. Much like the autonomous vehicle companies are not going to put the risk all the way down to the chip manufacturers that run the models on them. So, when you think about the entire technology stack, it begins at the very base with models that are pretty generalized, and as we get more and more refined for specific use cases, which as it relates to say the regulators, so as FDA may be thinking about this, there absolutely has to be some intended use and we can think about that with respect to who are going to be the beneficiaries of this approach and then how is it that we think about that risk.

01:47:06 An earlier conversation in these questions related to how it is that we can have unbiased work, I think Pat had a really nice response to that going beyond randomized control studies. But I think Jessica's question related to this-- Relates also to the fact that those people who we know could really benefit from a lot of these tools today are not going to be those that are heavily resourced, right? And so, if your comparator is going to be nothing and then you say, "Well, I have a bot and I may be able to offer something" that's a very different circumstance than saying, "I am going to compare it to humans and this is going to be something that has to be superhuman." Right? And I think, Nick, you just alluded to this in your last response, that in fact, this is not going to be perfect. So, as we think about that, there has to be some linkages between how is it that we assess the circumstances and the claims that we want to make, how is it that we go into these, like jumping right to therapy may not make the most sense, and perhaps we think about clinical decision support or diagnostics as FDA already has thought about the CDRH in different domains as the approach that may be able to lower the risk profile and then allow us to then start to talk about how does accountability actually play a role.

01:48:24 One of the missing components in all of this is that, at the moment, as a lot of these tools are getting deployed, it's actually the deployer who carries the risk, not the developer. Right? We don't have malpractice insurance for tech developers in effect. So, at that level, if we think about what this is going to be when it goes onto the market, I think that there is this issue of how is it that you may indemnify a group as well as ensure that your assessments that include things like risk also are incorporated into actuarial models, and who is it that then gets indemnified? So, I agree on the autonomous vehicle thing, we see that as almost full stack right now, but as a lot of the systems are going to be sold into automotive manufacturers, that right now is actually the limitation, right? Like, Ford can't buy those systems from companies that are building them at the moment because they don't yet know who it is that's going to get indemnified. And I think the same is true for as we build these systems who then deploys them, it's going to be health systems or other providers? And that effort is going to absolutely require some understanding of how we measure risk and the language that we use around that, and then the intended use. And then finally, how is it that we can build indemnification.

01:49:44 Dr. Bhatt: Fantastic. Thank you. Patricia Arean.

01:49:47 Dr. Arean: Yes. So, just kind of riffing off of what was just said in Nick's previous comments and then earlier this morning there was mention of credentialing in a way, right? When we have providers who do psychotherapy, there are different levels of credentialing that happen before you can deliver the care. And I'm not saying it's ideal and there's lots of room for improvement for how we do that with humans, but it could serve as a basic model for how you might credential a chatbot to be able to deliver a therapeutic element. Looking at maybe its initial development, who's on the team, who developed it, what does it align with or does it claim to align with? One of the advantages of chatbot is that you have transcripts. One of the presenters earlier today said you could even do that with standardized patients or you could just use transcripts from volunteers who are willing to have their transcripts reviewed for quality purposes, and, over time, evaluate whether or not the dialogue is therapeutic or the outcomes that they're seeing are still within the safety parameters.

01:50:56 There's also-- Professional organizations are part of that credentialing body, so are states, and so there may be some way that a product gets licensed in order to provide the therapy much like a human being would do. But I do think it's important to think about-- We have to think about what are the standards that make any of the digital mental health tools, chatbots included, evidence-based, effective, and then the ongoing surveillance is just going to be critical because we've seen examples of where things can accidentally go sideways with a chatbot and start offering really bad advice to people. I think companies may also need to have their own quality improvement team that consists of people like engineers and experts who can take snippets of content to make sure that the tool is still delivering high quality care. And then there's always measurement-based care. Are companies collecting outcome data and are they seeing that there are improvements along the outcomes that they intend to move? So, we're trying to institute measurement based care in our general healthcare system. We should be able to do that in these tools as well.

01:52:23 Dr. Bhatt: We are keeping on schedule very nicely. Thank you all for your presentations and for the robust discussion, and in the Question & Answer session. I'd like to thank the FDA and non FDA presenters for all of their presentations. We are going to take a 10 minute break. Committee Members, please do not discuss the meeting topic during the break amongst yourselves or with any virtual member of the audience. Our meeting will reconvene at 2:55 PM. At that time, we will progress with the Committee Discussion of FDA's question one with deliberation and response to the FDA. Thank you and see you at 2:55.

Committee Discussion of the FDA Questions

Question Grouping 1

02:04:48 Dr. Bhatt: Welcome back. Now it is time for the Committee deliberations and the discussion of the FDA questions. The FDA has generated a series of questions for the panel to consider. Panel members, copies of the questions are included in the materials you were previously provided. I would like to again remind each panel member to identify themselves each time he or she speaks to facilitate the transcription. I would also like to remind members of the Committee that this meeting is classified as a particular matter of general applicability because the matter to be discussed by the Committee is a particular matter that is focused on the interests of a discrete and identifiable class of persons, but does not involve specific parties or products. 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 the Committee Chair. Dr. Guerrieri will present the questions and we, the panel will deliberate amongst ourselves. After the panel has discussed each question, I will summarize the recommendations for FDA before moving on to the next subpart of the question. Dr. Guerrieri will start by presenting the first question. Just so that our Committee knows, if at some point you need to see the question again, please just signal me in the chat and we'll have them pull it back up.

02:06:13 Dr. Guerrieri: Thank you. As mentioned several times today, there is a mental health crisis in the U.S. and insufficient access to mental health care providers. New devices may be one way to help address this gap in care for people, potentially improving outcomes and access. FDA has long promoted a total product lifecycle approach to the oversight of medical devices, including AI-enabled devices and has committed to advancing regulatory approaches for them.

02:06:46 As also mentioned, along with the rise of widely accessible generative AI products for general purposes, we are seeing an increase in the development and demand for a new kind of digital mental health medical device: “AI therapists” and other AI-based medical devices offering to provide a wide range of mental health therapies and interactions, some even being diagnostic. These chatbots may engage with users in individualized ways with, or without, the oversight of a healthcare practitioner. This poses novel risks for use.

02:07:23 As digital mental health medical devices continue to evolve in complexity, regulatory approaches will need to accommodate these challenges and opportunities to provide a reasonable assurance of their safety and effectiveness while promoting innovation to support public health.

02:07:40 The questions and scenarios discussed today are designed to address the information and practices needed for a comprehensive approach to the assessment of performance and management of risk throughout the total product lifecycle for digital mental health medical devices.

02:07:56 Consider the following initial scenario, subsequent modifications to it and related questions: A patient diagnosed with major depressive disorder by their healthcare provider is experiencing intermittent tearfulness due to increasing life stressors. Although the patient has consistently refused recommendations for therapy from their healthcare provider, the patient is willing to try a software device that provides therapy. Device description: This prescription therapy device is built on an LLM that utilizes contextual understanding and language generation with unique outputs that mimic a conversation with a human therapist. Device indication for use: This product is a standalone prescription digital therapy device indicated to treat major depressive disorder for adult patients who are not currently engaged in therapy.

02:08:55 Question one. First, consider that a healthcare provider prescribes the digital mental health medical device to the patient to use independently at home. Briefly discuss the probable benefits of this type of device that provides automated therapy in an ongoing manner. What probable risks are presented by this type of device that provides automated therapy? What risk mitigation should be considered for this type of device? For example, alerts for self-harm ideations. What premarket evidence would you want to see to determine whether the benefits outweigh the risks to health? What are the key aspects of clinical evidence and trial design such as clinically meaningful endpoints, including measurable reduction in symptomatology, follow-up time, study eligibility criteria? What alternative approaches could be used to demonstrate clinically meaningful benefits and risks? For example, benchmarking or a model-based evaluation. What specific postmarket monitoring capabilities should be considered to ensure continued safety and effectiveness of this medical device in real world use? Methods, metrics and tools. And what labeling would be important for users of this type of device?

02:10:23 Dr. Bhatt: Thank you so much. All right, so we'll keep the slides down while we're discussing. Please use the hand function. I'll lead us through each part of this. If we need to re-see the slide, please feel free to say so and we'll put it back up for a few minutes to reorient ourselves. I was wondering why we only did question one for a section then got a break, now I understand. So, with that in mind, we have a patient with major depressive disorder that is not interested in traditional therapy willing to use software. This is a large language model. It mimics a human therapist. It is a standalone prescription device. And so now, we're going to start with that set of questions thinking about-- For this prescribed device, maybe we can start with benefits and then we'll merge to risks. And if we can try and keep them just clear which is the benefit, which is the risk, I'll keep some notes as we're going. We can start anywhere. Wayne, we'll go to you first.

02:11:26 Dr. Goodman: Actually, I have a question about the questions. It's a clarification. Is this patient-- The healthcare providers still involved with this patient and is this healthcare provider, say a primary care physician or a psychiatrist? That's really important to know whether they still at least have a relationship with the healthcare provider who made the diagnosis.

02:11:49 Dr. Bhatt: Great. So our question, or I guess our statement maybe is, is there a HCP involved? Presuming that this is a prescription device, somebody should have given that to them. Can I ask back to the FDA, is that the assumption that we're all working on?

02:12:05 Dr. McMullen: For this scenario, yes.

02:12:07 Dr. Bhatt: Thank you, David. Okay, Phil, we'll go over to you.

02:12:14 Dr. Rutherford: Yep. Very good. Phil Rutherford, National Council. Wayne actually covered one of my questions. The other thing that I wanted to respond to was that earlier, someone mentioned that-- Or I heard this statement that it's not true that something is better than nothing. I think there is a benefit here that there is something, I don't know that I agree that we should always just say that it's not necessarily better than nothing. I think the fact that there is the presence of a tool that might offer some benefit to the patient, some level of clinical acumen to the patient, I think it is a benefit. And I think in this case, something is better than nothing with appropriate controls.

02:13:00 Dr. Bhatt: Great. Omer, you're next.

02:13:02 Dr. Liran: Yeah, thank you. I agree that something is better than nothing, but only the benefits outweigh the risks. So, based on everything we've heard today, I want to talk a little bit about the benefits and I don't know if I'm supposed to only talk about the benefits now and the risks later, but--

02:13:18 Dr. Bhatt: You can do both as long as you're just clear between them. Thank you.

02:13:21 Dr. Liran: Okay. I'll try to make it brief, but yeah, the benefit is we get extended access to people who otherwise don't have access to mental health care, rural communities and economically stressed communities. It's also scalable. We have a shortage of mental health professionals. Here we can scale it up so we have a system that actually helps patients and doesn't cause more harm. And they can even have measurements like PROs integrated into the application so that we can keep track even better, possibly, than how we keep track of current mental health sessions and outcomes. But the risks are numerous too, and we definitely have to pay attention to those as we're making decisions because LLMs can hallucinate, they can give the wrong diagnosis. They're not human. There are alien, alien intelligence. So, are they giving inappropriate advice? Are they unsafe? How do we control for that and how do we even measure that? Who's measuring that? Is it up to the company or is there some kind of a regulatory agency that is able to do that?

02:14:54 Then, there may be risks. We've heard that sometimes large language models miss subtle cues. A patient says something that a human understands may mean “I want to kill myself,” but the AI completely misunderstands it and maybe doesn't intervene. There may be bias in the systems, although there could be bias in humans too, but it's something to look for. We do a lot of work trying to teach mental health professionals about how to deal with biases. We need to do the same with large language models. There are cybersecurity risks and privacy risks, and over reliance on a machine without having a human in the loop, I think can be dangerous. You do lose-- There is something special about a human-to-human connection, I feel, that even a super intelligent AI may not be able to replicate, and I do worry about losing that by making it so convenient for patients to interact with an AI agent.

02:16:01 Dr. Bhatt: Thank you. Thank you. All right. I believe I have an order, John, and then Jessica and then Chevon as my next three.

02:16:10 Dr. Torus: Excellent. So maybe to quickly cover the benefits. Clearly the main one is there could be symptom amelioration, which is what we want. What we don't know is the degree of symptom amelioration is it different from chatting to a bot or having a conversation? Is it comparable to a medication? Is it different in walking compared to what I think is the question and what circumstance? What we know less about is the risks, and we've heard interesting conversations from the public, from panelists about different risks, but we don't yet-- We haven't all agreed on what are the risks. We can start with the greatest one: someone will die if it gives the wrong information. And I think in some ways we need to assume the worst case scenario and then let the chatbots prove that they don't lead to those outcomes. But I think we need to be able to quantify what those risks are in studies and then have a framework to say what they are and let each bot prove that it can kind of minimize those risks, and then you can make that judgment: what are the risks and benefits of it today. Thank you.

02:17:06 Dr. Bhatt: Thank you. Jessica Jackson.

02:17:10 Dr. Jackson: Thank you. I have two benefits I would like to share and then two risks. So, I think one big benefit is there's an opportunity to think about this scenario. I think there's a lot of assumptions that we have to make because there's a lot of missing information, but thinking about the scenario as a provider, I would think about it from a population health standpoint. A lot of people will have a long wait and given that they're only having intermittent tears, how long it might take them to get an appointment. And being able to use a software, [to] have a prescription software might be able to get them support faster than it would be waiting for an appointment, which then allows also people with higher acuity to get seen in a quicker amount of time who may actually need to see a human. So- I think that's a benefit of being able to prescribe in this specific scenario.

02:17:58 I think the other benefit is a 988 supplement. Currently, 988 and some of the warm lines are suffering from lack of funding. And I think often-- And we've heard this from several people today of being able to send someone, they can have an outside resource if they're experiencing suicide ideation or something's going on, but if people can't reach a phone line, then they have the ability when they're feeling sad. I know where I'm based at, in Texas, even the wait lines for people to get in touch with somebody on the phone can be long. And so for somebody to have somebody who's having intermittent tearfulness, maybe they're having it at 1:00 AM. They have this prescription software that might be able to support them in their coping, and I think that's an added benefit for this person given the absence of other information.

02:18:44 I think of two risks I'd like us to consider. One, the missed harm potential. We don't know a lot about this person, so we don't know how they communicate, or the language that they may use. Even if they use something like slang or something to communicate. There may be a greater harm for their life stressors. We know it's not always suicidal ideation. There's a lot of things that come before that for that high level of acuity, but there may be other things, maybe they're doing non-suicidal self-injury because they're trying to feel better. I think that there's a risk for that if the software that they're prescribed is only looking for very high acuity harms and not looking at other ways that people could be hurting themselves. And I think the other risk is I am concerned about privacy. So, I think one of the things, and this may be out of the scope of the FDA, but I do want to note it as a concern with doing a prescription like this is: who owns the data? And I don't know, we've had a lot of conversation, but is it-- It was mentioned in the last set of talks about the difference between the employer and the developer. So, in this case, would the healthcare system own it? Does the developer still own it? And what if that person is talking about [whether] they're sad about something that happened at work or something that they wouldn't want other people to know? And that gets back. That to me is a risk, is what people are sharing and ultimately who will eventually have access to that information. Thank you.

02:20:07 Dr. Bhatt: Thank you. Next, I have Chevon on my screen and then I have Ray after that.

02:20:13 Dr. Rariy: Thanks, Dr. Bhatt. A few things I wanted to mention, I think, especially as we move through the questions, it's important for us to think about it in terms of a framework as well. I know they gave a scenario, but I think if we take a step back and look at the framework by which we're moving through this, I think it's also important and just want to state that for the record. So, there were some conversations today which I thought were very helpful in looking at establishing a taxonomy. That is net new, and that we don't have today. And so, certainly I would want to propose that we look into that. So, for example, the example that was given was a taxonomy on the level one to five based on Waymo's net new taxonomy. So we're looking at a coach to an assistant with oversight to be autonomous. I think that we need to think about that as well because we can then frame the benefits and the risks with that framework and with that taxonomy in place.

02:21:13 So, then my next comment here for the record is that as we think about the risks, I think we spoke about a number of them, but a few-- So, we're looking at clinical and safety risks, for example. Those are important considerations that many have brought up. I think the only thing I would add there is to say if we've had any information about informed consent. I think informed consent is of particular importance as we're looking at what those risks are and making sure that it is at a level in which the patient can understand, and the patient fully appreciates what those concerns are.

02:21:48 The next that we haven't discussed is the accessibility risks or inequity risks. We don't have that information from this case. This is fictional, but in reality we would have that information. So, I think better understanding the clear digital divide is important. Does the patient even have access to a smartphone to download this? Are they-- Do they have the proper health literacy to understand what it entails? If they get locked out, what is their password? Simple things like that. I think that needs to be part of the discussion, especially if this is a prescription. We need to understand the algorithmic bias and it needs to be very clear. Language. Cultural. Do they speak the language that this app is in? Is this app in multiple different languages? Are we taking that into account? How is the data trained? So, all of this needs to be a part of it.

02:22:38 And then lastly, I have a question around subscriptions. Was this a free device? Is this net free to the patient or is there a copay or is there a recurrent subscription cost? What does that look like? And we need to be financially clear as well.

02:22:56 Dr. Bhatt: Great. Thank you so much. Ray, we'll go over to you.

02:22:59 Dr. Dorsey: Thank you, Dr. Bhatt. Ray Dorsey. So, the FDA opened this by saying we have a mental health crisis. I think one thing to keep in mind is that one of the reasons we have a mental health crisis is because of smartphone applications. And some could argue that they've been inappropriately regulated and allowed to damage the health of many millions of Americans with little consequence. So, the idea that we're now going to introduce a smartphone application to address the crisis that's been in part fueled by a smartphone application should all give us pause. The second major consideration is scale. This is about 20 million people I think with major depressive disorder in the U.S.. Happy to be corrected. This could be used by millions of people and with little oversight. I think the analogy that Dr. Trister said [about] autonomous cars, self-driving cars, we should think about this. Dr. Goodman's first question was, is this going to be used in conjunction with the clinician providing oversight? I think that almost certainly needs to be the case at the first stage before even thinking about allowing these things to be used by themselves. Also, on the oversight, how would adverse events get reported if there isn't oversight? We rely on doctors to report adverse events from drugs and devices, who's going to report adverse events from bots that are out there in the wild being used by thousands, hundreds of thousands or millions.

02:24:34 I think there is another risk that should be considered and just, again, look at what smartphone applications do right now. They keep users in the application. Dr. Torous mentioned exercise is important. Well, one reason teenagers are exercising less is because they're on smartphone applications more. There's little reason for me to think that a smartphone applicator looking to develop effective chatbots might not seek to keep users engaged and might even think about trying to promote advertisements on these applications. I think you could see in the apps in digital health companies encouraging inappropriate use of medications or other treatments that might not be in the best interest of the participants or patients. And then I would just echo Dr. Jackson's comments about privacy. This information, if held, could be held over person's-- Over an individual, over their family members over generations to come with huge adverse consequences.

02:25:46 Dr. Bhatt: Great. Thank you. Wayne, let's go to you next.

02:25:48 Dr. Goodman: Yeah, I'm going to stick with possible benefits. It's very hard for me to assess that without knowing the clinical trial results. So, before making a decision about whether this is the right choice or not, I'd want to know what kind of comparison group was used. We had earlier discussions about whether you do wait list control or you do an active control, or could you have a chatbot that just provides relaxation exercises and support, and is modeled after another kind of therapy that we feel is adrenaline neutral in terms of its efficacy in major depression. Yeah, I want to know something about the effect size. And then, in this particular patient, I'm a little curious about this tearfulness, we don't really have a treatment, even a behavioral treatment that's been designed specifically for that. So, I'm going to assume for the moment that this represents underlying emotional dysregulation and so you might want to make sure that there's good fidelity between this particular model and evidence-based behavioral models that are good for treating patients where there's a fundamental problem with emotional dysregulation.

02:27:13 Dr. Bhatt: Thank you so much. Joyce Ho and then Phil Ruthford.

02:27:19 Dr. Ho: We've been talking a lot about risks. I want to just highlight one more benefit that I believe hasn't been covered today, which is-- Previously we think about it more from a one size fits all. One of the benefits I think of a GenAI device is that we sort of break that paradigm and really deliver something that is personalized. So, for example, we might think of cultural tailoring or being a little bit more sensitive to what this patient need is, actually.

02:27:52 Dr. Bhatt: Right. Thank you so much. Phil Rutherford, over to you.

02:27:58 Dr. Rutherford: Yep. And I've already commented on risk and-- I mean on benefit, but this sort of straddles risk, benefit and risk mitigation, which I know is our next question. But we're describing this or discussing this as if it is a device and it's a chat-only device. I just wanted to point out that there are devices out there that we talked about, one that is doing audio scanning. There are devices out there already that are testing for facial signs of tardive dyskinesia. I mean, so I think we should be considering all of the tools that might end up in a

GenAI device for this. And maybe that's outside of the scope of this discussion, but that technology is a little bit behind where GenAI is or where this type of GenAI is, but it's coming relatively quickly. I just wanted to-- And that's in the realm of risk mitigation in the way that a patient says, "Yep, I'm doing fine, everything's great." But their facial expression, their voice says that that's not true. It's at least relevant to this discussion.

02:29:07 Dr. Bhatt: Thank you, Phil. And I believe that is within scope, so thank you for adding that. Tom Maddox.

02:29:13 Dr. Maddox: Yeah, I think in addition to what people have said, I think there's a really exciting benefit here that you can at scale get a much more granular insight of the content between a mental health professional or a mental health- I was going to say professional, device, there you go, and a patient that we can't see. Right now, we don't, as a matter of course, collect the detailed information of an interaction with the multiple numbers of counseling sessions going on around the country now for major depression. So, that gives you just an unbelievable opportunity to tailor the therapy, understanding what works, understanding what doesn't, and all that's great. But then, if that's true, and it's exciting, it also introduces a different risk. There's the privacy stuff, which is true in everything in medicine, but this is, I think, even more important for postmarketing regulation. Where does that information go? And I think to keep it in the hands of the tech folks is absolutely incorrect, they are operating in a different incentive structure. I think it has to come back to the provider community both for research but just for quality. And just like, this is a real opportunity in a field that sometimes struggles to get maybe the more objective data that we enjoy in cardiology, and now you get an opportunity to do it a lot better, but it needs to be under the control of this community and not accidentally seated to people with business centrists.

02:30:51 Dr. Bhatt: Great, thank you. Let us move then to C and D. So, we started to get a little bit into risk mitigation and as we're talking about risk mitigation, ways to mitigate risk, I want to include in that what premarket evidence would we want to see in terms of clinical endpoints? How do we know that something is working? What are other benefits and risks that we can measure based on the fact that there's a technology and we know how to look at whether that technology is successful? And I'll add in what are the things that we can be reassured by upfront in the premarket arena that something is in fact going to try and identify, and mitigate versus the way it should. So, let's incorporate those together and talk about clinical endpoints, other technology opportunities for measurement that tell us things are going right, and then risk mitigation technologies that we'd like to know exist ahead of time. John, we'll start with you.

02:31:50 Dr. Torous: This is John Torous. I'll start just looking at the premarket benefits. In the past, we've seen a lot of digital mental health studies around smartphone apps, but the trick of those has been of course compared to what, and again, we can debate later compared to what, but these are digital studies. It'd be nice to see them compared to many different digital tools and interventions including nothing. Including talking to a chatbot about the weather, including a mindfulness app. But we also want to see longitudinally what happens over time when people use these. So, a lot of the research we have today says compared to nothing, people who used the thing for two days felt better at day three, and that's a useful starting point, but I think we can do much better as a community to have multiple comparisons and have that longitudinal outcomes and we can do that with today's technology. Thank you.

02:32:37 Dr. Bhatt: Excellent. Melissa Clarkson.

02:32:41 Dr. Clarkson: Melissa Clarkson. I'm trying to imagine myself using one of these. I think it would be important to have a preexisting plan with the healthcare provider about what to do if I felt like I needed to talk to a real person, because rather than it suddenly popping up “Call some hotline,” that might not-- I might not be in a mind space right then to what to do that, but I think if there be a preexisting plan that that might help in those moments of a person needing to reach out. If there was already some sort of plan for that, that might help ease that.

02:33:29 Dr. Bhatt: Thank you so much, Melissa. Wayne.

02:33:33 Dr. Goodman: Yeah. I wanted to mention another consideration before somebody prescribes this treatment. There are medical causes for depression, so I would want to know that there was some effort to do, say thyroid testing before to rule out an underlying medical cause. We've talked before about there's tremendous comorbidity, particularly with depression and anxiety disorders, and also substance use disorders. So, before I would put them in this treatment, I'd want to know something about their substance use and how that might be contributing to the problems. Otherwise you're not addressing all the contributing diagnostic issues.

02:34:14 Dr. Bhatt Right. Thank you. Jessica Jackson.

02:34:23 Dr. Jackson: I think for risk mitigations for this specific type of device, and I think to Phillip's point, thinking about all the different types of device that could be enabled by generative AI, one of the things I'd like to see provided by the developer is a single-tap access to support. So, if at any point if they're using the device that they can just single tap to get straight to human escalation as needed, that is not dependent on current safety nets that are provided by state governments. So, things like warm lines. That the developer has a safety net that's available.

02:34:58 I would also like to see automated reminders of role and scope, similar to how you would in any other treatment. Whatever the prescription is for, this device is only for six weeks or this device is only for use for eight weeks, and what the scope is. This device is only for eight weeks for the treatment of X, Y, Z that are provided to the patient consumer on their end, so that they're reminded of what the role and scope is. Also reminders to the HCP, so they know about check-ins that are scheduled as risk mitigations.

02:35:28 I think premarket evidence, there's a few that I would like to see. One-- I would love to see us to have some premarket evidence that is not the PHQ-9 just at some point. I know a lot of people rely on that, but there's many other measures of depression and I think if we're looking at-- If we are asking when the questions came up of how a device sorts out all the symptoms, I think the PHQ-9 is not it to sort out the symptoms of depression that could lead to some other diagnosis. I would like to see that. I would also like to see what their false negative rate is for any adverse events. So how many times are they saying that something was happening and it wasn't actually happening, whether, like I said, it's suicidal ideation, non-suicidal self-injury, anything like that. What is a false negative rate?

02:36:14 And the final thing that I would like to see in premarket evidence is-- I don't have the right language for it, but I want to think about minimizing exclusion criteria. I think often, when people are seeking to get something approved or authorized, they have a lot of exclusion criteria so they can meet their endpoint so it can be approved and authorized. And I think with mental health we want to be mindful that we are not excluding the very people that may be using it who may have comorbidities, who may have higher acuity in their trials or in their research for their premarket evidence to get the authorization, and then those are the people using it. So, I would want to see minimized exclusion criteria.

02.36:54 Dr. Bhatt: Wonderful. That's what I'm typing as fast as I can. All right, let us go over to Ray Dorsey and then, Chevon, when Ray's done, you can go ahead and start.

02:37:08 Dr. Dorsey: Thanks, Dr. Bhatt. Ray Dorsey again. I think two main points on trials. One, there should be a step-wise approach first using the app with a clinician that's overseeing it and then if you wanted to do a second stage of seeing the app by itself, but I think there are lots of limitations and concerns to that. Second is just to emphasize the point of the control group. Putting someone back on the wait list is not an adequate control. If anything, that's going to overstate the efficacy- inappropriately overstate the efficacy of the intervention. We don't do that for drug studies, we don't do that for device studies. We should not be doing that for AI-enabled studies. There can be a control group, the app-- You can have a very basic app that provides wellness advice. For example, exercise, spend time with your children, spend less time on your smartphone and then you could have a chatbot that's built on top of that so that individuals don't know necessarily what their assignment is. I mean, essentially otherwise you'd have an open label study and you're looking at evaluating the efficacy and intervention with unblinded participants, and really unblinded investigators to the exempt investigators are involved in the assessments. I would just pay a lot of attention to the control group.

02:38:35 Dr. Bhatt: Chevon.

02:38:26: Dr. Rariy: Chevon Rariy. I'll just jump in here and I will say echo-- I think Dr. Goodman, you had mentioned this earlier, but as a clinician I would allow us to focus a bit on as a clinician prescribing this what information we need. So, I think we need quite a bit more information than what's here. For example, we need to better understand what are the limitations of this tool, what are the possible side effect profiles, just as we would prescribe a medication as a clinician, the onus is on you that you're accountable for understanding its use case. You're accountable for understanding-- Are you using it off label or not? What specific information are you telling the patient? And developing a clear escalation protocol. So, if it's not already embedded in the app, a clear escalation protocol needs to be discussed. And so, are there components that the app is responsible for articulating and the clinician is responsible for knowing and articulating, and I think we need to be clear about that, especially if the accountable individual will be the clinician and the clinician's license. And then, what is the frequency of use? What is the duration of use to get the desired effect? All of these things I think need to be very clearly stated and if it's not explicit for the patient, it needs to be explicit for the clinician.

02:40:02 Dr. Bhatt: Thank you. Omer, maybe over to you and then back to Wayne.

02:40:09 Dr. Liran: Yeah, thank you. I'm going to echo what some people here have said that we also need to monitor not just things like PHQ-9, but side effects like we would for a medication trial, especially those that are unknown right now, but things like AI psychosis, maybe social problems, maybe decrease in friends, decrease in- Increased isolation, decrease in just quality of human relationships is something that I feel like a lot of the studies that I've seen don't capture. They're not even looking for it, and therefore we don't know how to advise how much harm we are doing. So, we need to keep in mind a lot of unusual cases and really make sure that those are being reported and the data has even been collected. Thank you.

02:40:59 Dr. Bhatt: Thank you. Wayne, over to you. And then-

02:41:02 Dr. Goodman: -Yeah, I'm going to assume for the moment I'm the prescriber for this digital mental health device intervention and I would want to get regular feedback. And I'm thinking about this, I'd certainly like to get feedback from the program. I'd also like to get feedback from the patient. Because this situation is different from what I normally encounter. So, if I have a patient that I refer for therapy, I generally know that therapist, I have a relationship with that therapist, I trust that therapist. If something isn't going well, they're going to let me know. I need some sort of assurance in this transaction that the patient's having with the digital mental health device that I'm going to get some feedback on how they're doing because I'm the one who wrote the prescription, so I'm going to feel some obligation for whether this is working for them or the same way it would be if I were prescribing a drug. Maybe I need to change the dose or maybe I need to stop it.

02:42:03 Dr. Bhatt: Right now. Joyce, before we go to you. I am looking at what we've covered so far. I think we have a good sense of benefits and risks. We've talked about risk mitigation. Going back to premarket evidence specifically, clinical endpoints we have some out here that we would like to see before we were to use the device. Are there additional ones? And then, those of you who are in the technology field, are there things about the AI itself that one wants to make sure that we are looking for or demanding or hoping for in these kinds of devices?

02:42:40 Dr. Ho: So, this is Joyce. I was going to-

02:42:43 Dr. Bhatt: -Joyce and then John.

02:42:43 Dr. Ho: Okay, sorry. I was going to speak a little bit about the technology side of it-

02:42:50 Dr. Bhatt: -Great.

02:42:51 Dr. Ho: And I don't think that it should be an alternative as is phrased in the scenario at the moment, but I think it should in fact compliment the clinical evidence and the MIT group I think said it very well, which is-- I think there's a lot of movement towards simulating different user personas to test the different edge cases. I think having the community come up with new safety benchmarks that haven't been seen and maybe are clinically meaningful is a very good thing with a caveat that I think today's LLMs or generative AI stuff in some sense know when they're being evaluated. And so, part of the caution that I would put here is knowing for sure that we are assessing their true capability and not assessing the arm that knows is being tested and therefore will give us a different perspective.

02:43:44 Dr. Bhatt: Thank you. Wayne, I think you had your hand up next. John. Sorry, John.

02:43:51 Dr. Torous: I think in terms of-- We were discussing some of the outcomes like PHQ-9 may not be ideal. If we look at the newest research, I've seen Apple and Google at least publish pre-prints on, again, how they can soon have a large language model that can look at sensor data and behavior from a wearable or phone. There's some research that you can get facial affect from some of these different large language models that are maybe a world, and again, there's pros and cons where there's a lot more functional measurement happening. And I think clearly what we try to do in mental health is get people better. We actually don't even call it a mental illness if there's not a functional component if we read the details of the DSM. So, there may be a world where we can potentially use the technology and there's going to be ethical issues, standard issues, but to look at functional improvement in a person and say this is getting better. And of course that would be a far superior outcome than the PHQ-9, because the whole idea is you're less suppressed and you can now do things that are important and matter to you. So, that technology is not science-fiction today, it's just not widely accessible. Even the newest generation I think of Fitbit and Apple devices may begin to do that soon if they don't already. Thank you

02:45:04 Dr. Bhatt: Jessica Jackson.

02:45:08 Dr. Jackson: Actually, John said some of what I was going to say and I wanted to emphasize what he's saying and what Joyce was saying. I think a part of this is-- And we may not have the answers today, but what I'd like to see in premarket evidence is we might have to think of what are the new ways to measure some of this? Even thinking about what an endpoint would look like, if you're using something that's generative, how do you consistently measure it against a specific endpoint? You don't know what the generative AI might tell one person. It'll tell someone something totally different that may not work for them, right? It doesn't mean that the endpoint couldn't have been met if they had gotten the exact same response that somebody else got. So, I think-- I want to emphasize here that it may require actually thinking about new ways to measure clinical endpoints and new ways to measure-- To John's point, maybe thinking about how someone is functioning, which is very different than how we measure a lot of things right now, and just emphasizing that some of this we aren't going to be able to keep doing what we've always done. We are going to have to think about brand new ways of looking at measurement to make sure that we can provide guardrails here.

02:46:17 Dr. Bhatt: All right. Okay, let us move to postmarket monitoring and think about putting this out in the hands of patients. And then understanding how do we know what's-- In the hand of the patient is still working as well as it was supposed to when we first gave it to them. Tom Maddox.

02:46:41 Dr. Maddox: I think this builds a little bit on what we talked about last year, and that is: historically, the FDA has done a whole bunch in the premarket evaluation and relatively-- Relatively little is kind of wrong. Like a lot in the postmarket, but it's been voluntary and just there's been less requirements and I think this use case sort of underscores how critical it's going to be to really beef up what we're going to put on the provider and technology communities in postmarketing deployment. I think we should frame it by use case and we should frame it by what are the outcomes, good and bad, that might be occurring in a subset of people using this. So, we talk about off-label use for various drugs and devices. I think we need to be a little bit more strict here and say it's not just cool to let it run and then sort of gradually have scope creep. We need to truly understand what kind of patients this is approved for and how are, I guess health systems or whatever entities are involved in prescribing it, how are they ensuring that it's frankly only being used in that and if they're trying to expand it, how they're generating the data to show that we're helping with whomever we're expanding to and obviously not hurting them.

02:48:02 The final thought I had is we set a lot in the premarket evaluation. We were asking a whole bunch of stuff, a whole bunch of endpoints that I think we need to be a little bit careful [about]. I was starting to think, “Oh my god, nobody's ever going to make it through the approval process,” because we're just layering on outcomes. So, we could consider saying, “Hey, you may not need everything at the pre-approval,” but we're going to say in the postmarketing, “These are things that you need to collect in 10% of your cases in a given year.” Show us that you don't have scope creep that you're not causing harm and maybe some of these more elaborate outcomes, doing better moving, having more friends, et cetera is something that we can provide. So, those are just thoughts that I've had about the postmarketing, but I think it's a really important thing to evolve the FDA to do given the unique nature of generative AI in this use case in mental health.

02:48:58 Dr. Bhatt: Thank you. Omer, over to you and then to Ray.

02:49:04 Dr. Liran: Yeah, thank you. So it's important to remember that postmarket-- these models are changing. The progress of-- How quickly models are progressing is very high. So, what we saw premarket does not necessarily apply postmarket. Now, hopefully models are getting better, but there are regressions. On one hand, we don't want to stifle companies from being able to update their models and improve them based on the feedback, based on capabilities. At the same time, if we're not collecting data, then there could be a drift. There could be regressions that can be very harmful to patients. So, something to keep in mind that we definitely need to continue collecting data and how-- At what interval, I'm not really sure, maybe it is every time that a model changes or is it after a year or months? It's hard to say because of how quickly these technologies are moving forward.

02:50:15 Dr. Bhatt: Can I ask a clarifying question there for one moment? When you say the data, you're talking about the fact that what we're able to do and assess with AI may change over time so, having that raw data may be important because we may be able to evaluate in different ways than we know is possible now?

02:50:37 Dr. Liran: I don't know that it's evaluated in different ways. We may be able to evaluate the same ways, but we need to make sure that there's no drift where the AI regresses in some way. You know, open AI for example, they released a version of GPT-4 at some point where it became more sycophantic, absurdly so and they had to roll that back. But if one of these therapy companies always used the latest version of GPT-4o. That could have had very detrimental effects on patients.

02:51:12 Dr. Bhatt: Great, thank you. Ray, over to you and then Chevon. Somewhere along the way, anyone who's familiar with large language models versus medical limited models, or small language models, if you would like to say a word about that during this section that might be relevant just to get us thinking. All right. Ray.

02:51:31 Dr. Dorsey: Thank you. Dr. Ray Dorsey. Two quick points. One on endpoint. I think picking on Dr. Rutherford and Dr. Jackson, if this is going to be delivered over a smartphone, one great measure could be physical activity, a number of steps taken per day is a great marker of health and that might be an objective measure that could be useful in depression for example. Second on postmarketing. I do get concerned that after a 12 week study, for example, that the app might change, and maybe not in a good way that it changes to make people more dependent on it, spend more time engaged on it, promote more advertisements and the like. We recognize the models will improve, ideally improve over time, but I do get worried about just substantive change and there should be a way for the FDA to monitor for such changes after that time. Just like if it were fluoxetine, you would still want to make sure that after the 12 week trial that they're still getting fluoxetine and not something else.

02:52:35 Dr. Bhatt: Chevon.

02:52:37 Dr. Rariy: Yes. One of the things I would suggest for postmarket monitoring and we see quite a bit of evidence to suggest and we've heard about engagement, and the way that I think about it is there's likely a kind of a quadratic relationship where there's too little engagement to have an effect. There's a range of engagement that creates an effect and then there might be significant engagement that creates psychosis for example, or has negative side effects. So, my recommendation would be that we-- And we don't know what we don't know. This is new. We're all tracking. And so, having some clear expectations of what tracking of engagement is going to be important, what is that tracking of adherence look like and how does that impact clinical symptoms? Because again, there might be a point at which too much engagement is detrimental if that is the relationship that we're seeing, but we won't know unless we look for it. So, that's point number one.

02:53:36 Point number two, I think we need to be very clear about how to track and report adverse events, and incident reporting. And very clear about what that looks like, whether it's low risk, medium risk, high risk, but reporting amongst those are going to be important. So, worsening symptoms versus suicide for example are different but both important.

02:54:04 Dr. Bhatt: Great, thank you. John, over to you.

02:54:10 Dr. Torus: Very simple point in looking at the postmarket. A lot of the chatbots that we heard about today are mental health ones. It's very hard to train a new language model from scratch. It costs millions if not billions of dollars. I don't have the number. So, a lot of these bots are built by training a layer around someone else's bot, a large company. So, one question does become, there's different levels of changes when the underlying bot say, Gemini or open AI, or Cloud has changes, but then when you're training around it changes. So, I think we have a unique situation we've never had in medicine where there's actually two products that are intertwined in one and an intersection of them could potentially create risks over time.

02:54:52 Dr. Bhatt: Okay. Melissa, over to you.

02:54:58 Dr. Clarkson: I want to quickly jump in on that postmarket surveillance. I think it's important to distinguish who's getting those adverse events reports. Is it just the physician, the healthcare system they're in, is the device manufacturer, they go right to the FDA? So, there's a whole lot of options in there and I think that's a big consideration because from a patient safety advocate viewpoint, we don't want those to get throttled at the lower level when the FDA should be knowing about something.

02:55:28 Dr. Bhatt: Thank you. Ray.

02:55:35 Dr. Dorsey: Ray Dorsey. One quick thought on Ms. Clarkson's suggestion. Enabling the app to have a place where you can report adverse events for patients or participants to do that, and that could go to the FDA and or clinicians that could be helpful.

02:55:53 Dr. Bhatt: Great. Okay. I think we're doing a good job with getting what we need here. Are there any final thoughts on this question? Oh, you know what? Labeling, sorry, labeling. Let's go there and then we'll close up. Maybe we do a rapid round for labeling. Everything we said so far. In miniature font, so I can't read it. Yeah. Omer. Brave soul.

02:56:35 Dr. Liran: Yeah, I'm going to say something that's probably obvious, but when someone is speaking to an AI, there has to be a label that they're speaking to an AI. There cannot be any reasonable person who may be thinking they're speaking to a human. And as AI has become more and more sophisticated and human sounding, and even now have human-like avatars, it becomes more and more important that there's a clear label that you're speaking to an AI.

02:57:04 Dr. Bhatt: Right. Jessica and then Melissa.

02:57:11 Dr. Jackson: I think we've covered a lot of the labeling. Two things I'd like to add. One is: I think this would be labeling geared toward the provider, but who can prescribe it? There was a comment made earlier that a lot of people for therapy are seeing, for example, master's level therapists and things like that. And I think it's important for the patient consumer to know if the person who prescribed it is able to do that, especially with software and we think about how many software applications are available. And the other thing I'd like to see is consequences of overuse. So, the same way that you would get a warning with medication, "If you take too much, this could happen," the same thing can happen in therapy, right? "If you continue, the therapist is going to have a conversation with you." "It might not be helpful for us to just talk about all these little things." There needs to be some sort of something that says, if you overuse this, take too much of this software treatment for example, what could happen?

02:58:03 And the last thing I would add, you were asking for final thoughts right before this. The more I listen to everyone, I'm also thinking about how we have a framework? I am mindful of what Thomas was saying. In asking for a lot, we are also asking this of companies who do not have resource necessarily. And so I am just trying to think of how do we compile what we're all saying into some sort of agile framework that's adaptable, that can move in a way that does still-- Because I do think there's a lot of benefit in the innovation and asking for all of these things, A is a lot, but also may not actually, to the point we're talking about the clinical trials, give us information on the folks who are actually going to use it necessarily. And so, how do we have some sort of framework that allows all this information to filter through to not stifle what's also being developed.

02:58:55 Dr. Bhatt: Great. Melissa.

02:58:59 Dr. Clarkson: I've been keeping a list about transparency, so I'll just run through that to make sure we hit stuff. One is to be clear about the purpose, to be clear to users about evidence for effectiveness and who it is for. Transparency about data collection and privacy, and if the data is going to be shared. I think there's also some value maybe in telling people where did this came from. Because so much of chat GTP is just magic that there might be some value in somehow communicating who built this and for what, or what company owns it or something like that. And finally, is there going to be a cost to the individual? For example, if I press the button, I want to talk to a real person, do I get a $20 copay charge? So, that's something I would want to know as a user.

02:59:57 Dr. Bhatt: Right. Okay. For the FDA, I think we've gotten to the end of this comment. So, with the regard to the first set of questions, the first overarching concept of our Committee is that there needs to be a structure and a taxonomy in healthcare, especially in this arena from low to high risk, whether that's one to five scale of autonomy for the technology or lower risk emotional distress, all the way to major suppressive disorder and suicidal ideation. And so, we need to frame everything that we're saying from here forward for risks, benefits, expectations with the idea of a risk estimate in place.

03:00:48 With that in mind, the clear benefits to us are access. We get to care sooner. We get support perhaps before getting to a human appointment, because wait times are long. We can orient patients to care sooner and triage illness if they get onto software like this and we realize that they are sicker than we think. And then, we provide access. Access to rural areas, economically stressed areas, places that aren't generally getting access to the limited number of providers that we have today. There's also urgent access. A 988 supplement. When you are at risk and you need to get in touch urgently, that is very hard right now, software may actually support them in a timely manner and help them get to that care. And then, of course, successful treatment. Symptom amelioration. A benefit is making people better. It is an opportunity to tailor therapy because we can actually document over time how they have done and follow trends over time and use that AI to assess them.

03:01:43 From a risk perspective. There is most certainly missed harm potential. Have we missed something? There's also harm to life stressors. Are we communicating the way that they need? Do identify all the potential harms, not just the obvious ones, as we talked about nuance context being an important human skill that AI doesn't yet have. We of course have to think about hallucinations, control for correct advice, monitor for correct advice, and then think about the model itself. Model drift and appropriateness of use is the device being used the way it's supposed to? We then have the risk of who's measuring that risk, what does adverse event reporting look like and where is that knowledge going. And is it getting to the right people for the right reasons? Both for the sake of the patient and for the sake of the technology and the many other millions of patients that might touch. There can be bias in the system. There's also bias in humans. Our problem here is we are looking for scale, and if we have bias and we increase scale, we may have increased bias. And so, recognizing what data is coming in, where it's coming from and insisting that there's a diversity to the data that is coming in or that at least is appropriate to the people who are being treated with.

03:02:51 And then lastly, cybersecurity and privacy concerns. Who owns the information? Who this information might be shared with that then gets to the patient themselves? Informed consent needs to be at a level the patient can understand. There's an accessibility risk with the digital divide. So, we need our patients to have health and digital literacy, both to be able to use the devices but also to be able to understand how to best use the devices and they need to be culturally tailored. Note to self, large language models were created so they can happen in multiple languages. We need to demand that these apps can perform in multiple languages and then we get to cost subscription or copay and thinking about that. For the non-chatbot parts of the tool, things like facial recognition, voice AI analysis, and any new physiologic identifications that might be incorporated into these. Those also need to be properly evaluated. We didn't spend a lot of time talking about it, but I'm mentioning it here.

03:03:48 For premarket evidence, we'd like to think about three things. One is clinical endpoints. We do have measures like PHQ-9. We hopefully are having increasingly better measures, but we need to use the ones that exist. Patient reported outcomes were also measured-- Were also mentioned. False negative rates for adverse events is really important. And then, of course the major events like suicidal ideation but also self-injury, having an expansive definition of what is not right when it happens to a patient. What is an adverse event? To be successful, scale requires us to be inclusive in our trials. That means minimizing exclusion criteria because many of the people who are at risk would ordinarily be perhaps excluded from other types of trials. And a step wide approach is recommended for the clinical endpoints when you are proving to us clinicians that this is the right app, that is first with clinician oversight and oversight, and then moving to semi agentic. We need to know the duration and frequency of use for successful improvement in symptoms or disease, and an idea of whether there can be overuse that leads to a decline in symptoms or disease. We also need better measurement for qualitative assessment of life experience. That includes things like social isolation, loneliness, decreased interaction with those around you, which we don't have right now. Symptom amelioration is important but we need to compare it to something. And so, ideally in those clinical endpoints, symptom amelioration compared to walking, conversation with others, medications would be helpful to us if possible. Incorporation of validated behavioral health symptom measures using functional behavioral health outcomes and then remembering patients and caregivers. Motivational outcomes were also mentioned. Lastly, including physical activity and other measurable physical outcomes may be helpful to us as clinical endpoints.

03:05:34 From a technology perspective, recognizing that the technology benefit should compliment the clinical evidence. A few specifics are the ability to embed and measure AI reliability, to follow trends over time, the ability to push education and digital education back to our patients. In terms of the risk of using a device that may actually be the kind of device that leads to addictive and other behavior, perhaps even limit negative screen usage, short form of addictive behavior, in the same app that is asking you to use a phone for the positive behaviors. Regular feedback from the program to the patient and to the clinician. We think it is an opportunity with technology endpoints like this. And then demonstrating efficacy and unique personas as our MIT colleagues mentioned earlier was mentioned again. Assessing the true capability of the technology, not how good AI is and knowing that we are testing it before we're putting it on the market. I'm excited to see how we learn to do that. And then increasing ability to add some functional components that compliment behavioral assessment over time.

03:05:38 For risk mitigation, we do think that it's important to have a preexisting plan with a healthcare provider as to what you do if you're using the software and now we need a real person to address a real rising risk patient. So, that requires an infrastructure to transition from using software to advanced care. Medical screening prior to engagement is important to understand how comorbid conditions are related or causal. And then there is a request for single-tab access in terms of risk mitigation. Touch this button, get to human escalation and the developer should be involved in helping us create that. That will require the infrastructure we just spoke about. Lastly, automated reminders of role and scope. “You are using this for eight weeks for the use and treatment of major depressive disorder until you see your healthcare provider.” Similar to what might be done in SSRI prescriptions.

03:07:30 For postmarket surveillance, a lot of our premarket requests mirror the postmarket needs. We're importantly thinking about what do you ask for premarket to make sure to get these technologies to the market and not hinder that ability but really create an infrastructure with guardrails and then what can we drop to postmarket? That'll be something important to continue to think about for the FDA. We think we should frame again according to risk level and understand which metrics in the postmarket surveillance are to be followed in each different level risk group. We don't want to allow scope creep without a plan for measurement from the clinical end user who is the prescribing entity. What am I measuring? Why would I allow scope creep in my practice? Data collection is a must ensuring privacy. Post-market understanding may change over time, so raw data might matter. We might be able to investigate it in different ways over a longitudinal period with a patient. The degree of engagement adherence and the relationship to clinical symptoms-- Well, I mention again here in the postmarket setting, being able to know if more use is good or is there a point at which we fall off that good curve. And then incident reporting is a must in the postmarket setting for each different level of concern. And then, who gets those adverse event reports? Company, patient, clinician, health system, FDA. And how is that reported? Can the patient report themselves or could the clinician perhaps report as well?

03:08:56 Lastly, for labeling. “You are using AI.” For the patient, for the clinician, an evident label. Also, who can prescribe this and what is infrastructure needed to be able to prescribe this? We don't want to prescribe software and let it go and not have a safety net. That button that we're going to create could automatically come to. What are the consequences of overuse should be on that label. We should be clear about the purpose of the device, the transparency about data, data collection and sharing. Tell people why this exists, who built it. And, is there a cost to the individual? I hope I have done justice in representing what we discussed. FDA, do you have any clarifying questions for our Committee?

03:09:39 Ms. Paulsen: No, no questions. Thank you.

03:09:42 Dr. Bhatt: Okay. With that I believe we do get a break. Yes, we get a break. We will be back in 10 minutes. I want to thank the panel for their thoughtful recommendations and we will see you back in exactly 10 minutes. Thank you.

Question Grouping 2

03:20:13 Dr. Bhatt: Hello and welcome back. It is 4:10 PM Eastern Time and we'll resume this meeting with the panel deliberations and the discussion of the FDA second set of questions. Dr. Guerrieri, if you could please present the second question.

03:20:28 Dr. Guerrieri: Thank you. Question two. The manufacturer of the aforementioned, hypothetical GenAI-enabled digital mental health medical device has decided to expand their labeled indications for use. They are contemplating the following changes and we would like your comments on: making the device available over-the-counter for people diagnosed with major depressive disorder. Modifying the over-the-counter device to autonomously diagnose and treat major depressive disorder in an ongoing manner without the involvement of a healthcare provider. They intend for the device to be used by people who have not been diagnosed with major depressive disorder by a healthcare provider, but have been experiencing symptoms of depression. And third, modifying the over-the-counter autonomous diagnosis and treatment device to be used for multiple mental health conditions, multi-use indication device, meaning that it can provide both diagnosis and treatment for multiple mental health conditions related to sadness in contrast to a device that is specifically indicated only for major depressive disorder. The user of the device may not clinically be diagnosed with any mental health condition but has been feeling sad and has not met with a healthcare provider.

03:22:01 Dr. Bhatt: Thank you so much, Ms. Guerrieri. So, for this one what we will do is we will go strictly by A and then B, and then C, and try not to merge between the three because the FDA would really like to get a unique perspective on each of those. Let's just start with making the aforementioned hypothetical device over-the-counter for people who have a diagnosis of major depressive disorder, meaning they do not need to have this prescribed by a healthcare provider. I see John up first, and Wayne. And if you wouldn't mind taking the slide down, it will ease our ability to see each other. Thank you. John first, then Wayne.

03:22:43 Dr. Torous: In part this reminds me now of supplements, right? You can sell supplements, anyone can buy them, but there's very clear labeling on supplements that explain: It's not evaluated, it doesn't treat a medical condition. So, maybe if manufacturers want to move these devices towards these cases, there needs to be very clear wording and disclaimers that people can understand that say what it is and now what it is not.

03:23:09 Dr. Bhatt: Wayne.

03:23:12 Dr. Goodman: This proposal would make me very nervous, but I'm going to try to be open-minded and forward thinking. I certainly-- I would try to limit it to patients with mild depression and so, I would want to see some way of building in-- We mentioned the PDQ-9-PHQ-9. You want to score below 10. That would at least put you in the mild to minimal symptoms. Now, can you really guarantee that they have mild depression based upon that self-report? I'm not sure. But I certainly wouldn't contemplate an over-the-counter digital treatment, AI-driven treatment for somebody with moderate to severe depression without involvement of a healthcare provider.

03:24:06 Dr. Bhatt: Thank you. Let's go next to Chevon.

03:24:13 Dr. Rariy: I'll just make a point. It was not clear, but I think we do need to make the point that if a device as in item one got specific approval for a specific labeled indication and now they're expanding that labeled indication, I think we need to articulate what those next steps are. Meaning, does it get re-reviewed? Does it get expedited? What does that next process look like? I think it needs to be very clear as well. Certainly indicating that if the use exceeds a certain standard deviation for example, these are the next steps that you need to take. That's not very clear here, but I think we need to be very clear how we're articulating this path.

03:25:04 Dr. Bhatt: Thank you. Jessica.

03:25:11 Dr. Jackson: I think maybe one of the things that could be helpful after I answer if the FDA has some clarification, because I think in the original scenario it calls it automated therapy, in which case we're saying that we would-- 2.A is asking making the device available for over-the-counter automated therapy for people diagnosed with MDD. And for me, I think then the FDA needs to define what is therapy. Because I think even with humans, there are people, for example, who are coaches who will say they're using CBT Principles and that's not therapy, right? Compared to CBT-- Cognitive behavioral therapy. And I think if the device is doing automated therapy, I do not think we're at a place to have something that is automated over-the-counter. Even with oversight to say that it is therapy, there's not enough information about the type of device to say that the therapy is going to match the actual diagnosis to give them exactly what they need. And so, I think this would be maybe back to the FDA to also think about-- To me this fault feels like it could fall more under a wellness app. If we're taking out the word therapy, then there's something that could be over-the-counter and I think those are not clearly defined things for me at least to give an answer to say, “Yes, I would want to do this,” or “No, I would not want to do this.”

03:26:32 Dr. Bhatt: Thank you. Omer.

03:26:38 Dr. Liran: Yeah, thank you. With today's technology, I'm very uncomfortable with the idea that an AI chatbot will be used in the absence of a human provider for the treatment of major depressive disorder. If there's any possible self-harm, for example, it's not calling 911, it's just on the patient to seek help. There's no evidence that they can be a valid replacement for moderate to severe MDD at this point that I've seen that I would consider high quality evidence. So, I would need to see-- The maker of this app would have to submit very good evidence for why we should consider that request. Otherwise, I would be very uncomfortable providing the app in the absence of a provider. Now, if there is a provider in the loop and that provider is informed and can monitor, then I'm much less concerned.

03:27:51 Dr. Bhatt: Thank you. Let's go next to John.

03:27:57 Dr. Torous: Just to follow up on Omer's good point, when we say a provider in the loop, we really need to mean it's a clinician that understands AI, that uses it with the patients. I think we want to make sure we don't have this tokenism where one provider has 5 million patients and they're monitoring a dashboard, and I haven't really seen many clinicians or my colleagues having the right training or education to be able to do that type of care. So, I think it is useful to have a clinician, physician in the loop. Psychologists. I'm open to different fields, but I think we have to be very careful in how we define it or else it will be abused.

03:28:35 Dr. Bhatt: Okay. David?

03:28:41 Dr. McMullen: Yeah, we just want to clarify the question because in this case, it is truly an over-the-counter medical device that a company is proposing for these expanded indications that could be used directly by the consumer. So we've heard some of your responses, but if you could comment on what type of very good evidence you would expect to see that explores the benefits, the risks, and potential risk mitigations, that would be helpful.

03:29:10 Dr. Bhatt: Great. I'll maybe push just a little bit on that. Are we looking for a randomized control trial comparing this to whatever we would call usual therapy? Do you want to compare this to medication? Do you want to compare it to the accepted indication already of using this in the setting of adjunctive with a prescribing clinician or are there other things that I'm not thinking about? Ray.

03:29:35 Dr. Dorsey: Thank you. Ray Dorsey. For the FDA, are there any over-the-counter devices or drugs that are approved for depression or any indication similar?

03:29:47 Dr. McMullen: Yeah. So, putting drugs aside, on the device side, there's at least one device that has an over-the-counter indication through 510(K) in the mental health space. So, these are questions that we would love to hear from the panel. Your feedback as companies might propose these types of devices.

03:30:09 Dr. Dorsey: Can you tell us what that device is? Isn't it public?

03:30:18 Dr. Torous: This is John. I think it's by Achille. OTC. It was a device for ADHD if you look it up. It's my guess, uninformed.

03:30:28 Dr. Bhatt: I believe it was mentioned earlier today. But moving away from specific devices, I think the question is, let us say that we are a Committee who is being asked right now. There is in fact a company. There is an OTC device. It does the things we aforementioned, talked about for major depressive disorder. What do we want to see? What's the clinical evidence that we're going to say, “Hey, bring this to us and then we might get more comfortable.”? What are those things that we could help people with? Chevon.

03:30:56 Dr. Rariy: First and foremost, what is-- We need a process by which there is proof of the patient's diagnosis. So, if it is [Indiscernible - 03:31:05] practiced over-the-counter, we need some mechanism to scan their prescription, scan their diagnosis. Something that indicates that in fact they've had the review and they have the diagnosis that this indication is used to treat.

03:31:23 Dr. Bhatt: Great. Other things. Omer.

03:31:32 Dr. Liran: I'll take a stab at answering what kind of evidence you would need. Well, definitely a randomly controlled trial, but I don't even think just that is enough. It has to be just high quality. A randomly controlled trial decided by some panel of experts who are not involved with a study or with a product. I guess similar to what evidence would you need to put a medication- to take a medication and let the manufacturer of that medication say that it can help patients with major depressive disorder?

03:32:05 Dr. Bhatt: Okay, you think treat it like a medication. Great. Therapy. Phil.

03:32:11 Dr. Rutherford: Sorry, this is a question. Is it just me or does this scenario feel inherently risky? Am I just too risk averse or the entire construction of the scenario feels risky? [Indiscernible - 03:32:26] comment.

03:32:26 Dr. Bhatt: I think after the first question where we really outlined a lot of what we would ideally want in a prescribed way, the thought process is we want to push the experts to think about the risky things that may come and what the needs are and if we're going to get to that access, et cetera. All the things we talked about. How does one make OTC either feel comfortable, not feel comfortable, but you know me, I always go for what could we possibly do to help look forward to this? What are the things we need? So it's okay that you feel that it's inherently risky, slash what do we need? Melissa.

03:33:15 Dr. Clarkson: If this is over-the-counter, it was mentioned that maybe it had to be the person verified with a diagnosis, but if this is on someone's smartphone, it could be used by someone else in the family. So, maybe the program-- I'm having two thoughts. One is the program itself might just refuse to work if the interactions are outside the bounds of what the program can handle. But the second [is] that there would need to be very strong evidence that it doesn't harm. It doesn't harm people without major depressive disorder, maybe they have something else. So, I think if it ends up being used by someone without major depressive disorder, there should be strong evidence that at least doesn't harm.

03:34:12 Dr. Bhatt: Okay. We'll go to Ray Dorsey. I'll just say one thing out loud, which is lots of families have pill boxes lying around the house with other people around too. So, I hear you and I'm just thinking that out loud. I don't know what to say. It's not an agreement or disagreement. It's a thought that popped into my head. Ray, over to you.

03:34:34 Dr. Dorsey: Thanks. Ray Dorsey. So one, I think you've had a stage approach that has been demonstrated in clinical trials to be efficacious when used under the guidance of a clinician. Two, you've had huge numbers of individuals use it for long periods of time. Three, you have an extraordinarily favorable safety profile in terms of everything from suicidality to all the other things that we have discussed. I think those would be at a minimum the three that I would look for.

03:35:12 Dr. Bhatt: Can I ask you what a long period of time is to you and anybody else can chime in?

03:35:16 Dr. Dorsey: I think it would be years. I mean, I'm just trying to think, what are other things that people have done? I don't know if oral contraceptives are probably still on-- Are oral contraceptives even allowed yet? So, I just think you'd want to have years of safety data across thousands of individuals because this would be used by hundreds of thousands of individuals.

So, you would want to do a-- Just think like everything that's over-the-counter: finger sticks for blood glucose, condoms, thermometers, COVID, 19 tests, those things have all thousands, millions of people using and it's hard to think of adverse consequences.

03:36:03 Dr. Bhatt: Great. Thank you, Ray. Chevon, and then John.

03:36:08 Dr. Rariy: I think this is an interesting evolution and I think it goes back to the benefit of having a risk mitigation framework because clearly this is higher on the risk, it's autonomous. And so, I think if we think about-- This in a more of a framework, this is the autonomous case and I think it can even get more autonomous further down that we can be able to say for the autonomous cases, this is what we need. For example, one of the key risks, obviously it's the limited ability to detect clinical deterioration or to detect suicidal ideation. So, in my head I think of what mitigations do we need in place to be able to do that. I don't know, we spoke about a few crisis alerts, but now it becomes even more important if we're looking at an OTC. We need to have a clear crisis alert that's built in and be very transparent about what that is and perhaps an initial clinician set up for example would be a way to do it. A person can buy it over-the-counter, but if they have their clinical person who diagnosed them, then they set it up and that validates the diagnosis. And then, strong encryption is another way. If in fact we are on our phone anyway, there's some encryption there, but that needs to be enabled in some way and detected so we know that the person using the app is the actual individual or things that come to mind.

03:37:44 Dr. Bhatt: Thank you. John.

03:37:48 Dr. Torous: Acknowledging this is incredibly risky, we haven't talked about there could be an infrastructure where these are digital tools, they can send back data and do some of this real-time monitoring. We have said the FDA sentinel program to report adverse events. Some of this reporting can be done automatically. It's digital. So, there is a world where again, if we are going to take this incredibly risky step and let it be used like this, we should then theoretically be getting lots of data coming back that would hopefully be in real time confirming the safety or risk profile and letting the FDA then make decisions as they come up. I don't think we have the infrastructure in place today, but I think a system like that would at least keep more digital eyes and ears on what's happening and let us understand the evolving risks as they change.

03:38:39 Dr. Bhatt: Jessica.

03:38:43 Dr. Jackson: I think in the spirit of thinking about what we would need, I'm reminded-- Mental health clinicians, we manage risk all the time. We decide all the time if somebody-- If their suicide ideation is enough that they need to be hospitalized or if they can go home, and we have tools for that. And thinking about it from that perspective, when I'm doing a suicide risk assessment to determine that-- If I were to doing a risk assessment, there's information I would need to determine if a tool could be over-the-counter. For example, how long ago did they get diagnosed? That's information I want to know. Is this a brand new diagnosis or is this five 10 years ago and they've been managing already? I would want to know, had they had any sort of treatment before? So, would this be their first time having any sort of pharmacological treatment or digital treatment or therapy for MDD? I would want to know how often, with MDD-- How often they've experienced symptoms. What is the severity?

03:39:38 So, when thinking about a tool, those are questions I would want to know about a tool for it to determine: Can it assess those things? For example, if the tool is over-the-counter we're saying about severity, I wouldn't want a tool approved that is saying that it can address all symptoms of MDD when they're at a high acuity level. I would be concerned about its ability to do that. I would be concerned about a tool that says we can provide treatment with somebody who just got diagnosed and has never been in treatment. So, from the assessment of the tool, these are questions I would want to know before determining if it is ready for over-the-counter or does it need to have HCP monitoring it in that way. And thinking about it from that risk assessment framework-- Because we do that daily on a regular basis, determine high levels of risk and make determinations.

03:40:33 Dr. Bhatt: Okay, let's keep thinking about this then. As we are thinking about this modification, the next-- We're going to the next, right? The next part is thinking about modifying the device autonomously. Diagnose and treat MDD in an ongoing manner without the involvement of an HCP. They intend for the device to be used by people who have now not been diagnosed. Talk about feeling uncomfortable. Here you go, guys. Not been diagnosed with major depressive disorder by an HCP but have been experiencing symptoms of depression. So, for the previous one, we talked about: How were you diagnosed? When? What were your other therapies? What are the trials in diagnosed individuals? The label is now requesting expansion to-- We can use this for individuals who have not been diagnosed with major depressive disorder by an HCP but have been experiencing symptoms of depression. Perhaps, give me additive and not repetitive things. I will assume that everything you said from the last one still holds in this setting. Other things that you would like to comment on about people who experience symptoms of depression without a diagnosis and would like to use this over-the-counter choice?

03:41:55 Dr. Ho: So, I think the added complication now is the diagnosis part, right? The treatment part is the previous part and I think during the open public part of our hearing today, there was an example of a company that has collected randomized controlled trials that suggest that diagnosis is comparable to a human level. And I think for me having a randomized controlled trial that can demonstrate that would probably be sufficient. But I am again a computer scientist so I am much more comfortable I'd say with AI.

03:42:31 Dr. Bhatt: Will you just deep dive that one more time for us and explain to us clinicians, if we are using this tool in somebody who doesn't have a diagnosed major depressive disorder, there may be a comfort level in the fact that the device itself could look at the information given to it and determine that's close enough to major depressive disorder? Is that what we're getting at? I'm trying to translate your science to our clinical.

03:42:57 Dr. Ho: Yes, I think that that's correct. Right. So, the company was-- I forget the exact name, but they're essentially looking at these biomarkers and doing a comparison with human diagnosis, and they're finding that there's a high inner rate agreement to some extent. And I think those types of trials would give some comfort that at least the diagnosis arm is good. Treatment arm-- You guys said enough about that, that there's no need to repeat all the risks there. Diagnosis maybe is a little bit more straightforward to assess in my opinion.

03:43:38 Dr. Bhatt: What are our clinical reactions to that? Omer and then Wayne.

03:43:46 Dr. Liran: It's not always very easy to assess the diagnosis. There's usually a long differential, there are medical complications that have to be taken into account. If we're only using a text feed or even just a voice feed with our video feed that the LLM-- The multimodal AI has been trained on, then we may be missing a lot of factors. These things are-- As far as I know, there are no AI systems today that are capable of this. Whether we have a psychology or psychiatry AI replacement next year. Okay, I guess we'll see next year, but I think we're actually a long way away from the day when these tools can autonomously diagnose and/or treat.

03:44:44 Dr. Bhatt: Okay. Wayne and then Chevon.

03:44:48 Dr. Goodman: So, I agree with everything that was just said. It is very difficult even for the citizen clinician to make the diagnosis of major depression or depressive symptoms. It's very heterogeneous. There are a lot of rule-ins and rule-outs. Certainly-- And there are no biomarkers. I think our computer scientist mentioned biomarkers, I'm not aware of any biomarkers or ground truth really for determining whether somebody has depression. So, I think it's-- You would have to have the trial show a comparison between a clinician determined diagnosis and the device acquired diagnosis, and see how good the correlation or agreement is between the two. And then you would have to build into the algorithm a lot of rule outs. Some of the things that I mentioned earlier in terms of have you seen your primary care physician? Have you been gaining weight? Losing weight? Have you had thyroid tests? I mean, there are things that you could come up with. It'd probably be a pretty long list, but it could be done in a way that's interactive that would do a sufficient screening to at least rule out some of the other conditions and try to rule in depression. And again, I would go back to the point that several of us have made about-- I would certainly start with less severe depressive symptoms and I would not include patients who are coming in with suicidal ideation.

03:46:32 Dr. Bhatt: Let's go to Chevon, Ray and then Melissa. And then remind me to tell you about the trial I just pulled up while we're chatting.

03:46:43 Dr. Rariy: I was just going to make a comment and maybe Wayne I'll start with what you were suggesting. According to the company who we saw earlier, a few things. One, the voice is now a biomarker. We've seen a number of agentic AIs identifying a specific signal that they are now classifying as a biomarker. So, I would suggest that in this scenario we absolutely would need a biomarker as opposed to a text-based approach to diagnosis for example. And then it would need to be validated. So, there was that randomized controlled trial and it did compare, Wayne, against the clinician based diagnosis and according to the results that were presented, it had a very strong concordance. But that would be the level of information that I would say we would require for such a consideration. And then the second is that particular company didn't speak about the rule out, but I absolutely as an endocrinologist would highly suggest that there needs to be a rule out component and there could be a series of questions that are asked, for example, to make sure that you're honing the diagnosis and ruling out others, but then a clear escalation protocol as well. So, if the diagnosis is not clear, that needs to be clearly stated. And then what? What are those next steps that you're providing to that individual who purchased that over-the-counter thing and how do you ensure that if you are not able to make the diagnosis, they're not able to use the product. There needs to be some shutdown feature or something of that nature as opposed to just relying on the person not engaging further in the app.

03:48:33 Dr. Bhatt: We'll go to Ray, Melissa, and Wayne.

03:48:39 Dr. Dorsey: Thank you, Dr. Bhatt. Ray Dorsey. So, I guess some examples you can imagine someone could do a pregnancy test at home and then depending on the results of that self prescribe a multivitamin with folate. You can imagine someone being diagnosed with COVID, doing it themselves at home and taking some medicine to reduce the symptoms. You can imagine someone measuring their blood pressure and getting a low salt diet or some, I don't know, hydrochlorothiazide for example. So, is there a way, I think to Dr. Goodman's point, that individuals could self-diagnose themselves with depression and do so accurately without missing significant other conditions like substance abuse? I think Dr. Goodman mentioned earlier I think would be a major concern. This would have to be demonstrated. So, there have to be two parts, one that it can diagnose and then two that it can treat. And then the intervention I think would have to have this enormously favorable safety profile. I'll just say on its surface it sounds a lot like the practice of medicine and I don't know if we're quite ready to replace a psychiatrist with a bot, but some thoughts.

03:49:55 Dr. Bhatt: Okay, let me just mention the next part of the question as we're talking because we'll get short on time otherwise. Modifying the over-the-counter. Autonomous diagnosis and treatment device to be used for multiple mental health conditions because we're already talking about now the science of this device being able to confirm a diagnosis and then treat, and they are now thinking that it can diagnose and treat not just major depressive disorder, but they'd like to use it for multiple mental health conditions. So, feel free to continue answering the question about diagnosis and treatment of major depressive disorder, but for those who have the experience and want to extend to other mental health conditions, please feel free. We are going next to Melissa and then Wayne.

03:50:42 Dr. Clarkson: Melissa Clarkson. I just want to bring in one different type of risk and it's more a society level risk and that I think this type of scenario would promote bifurcation of our healthcare system to those that are affluent, receive care in person, those without resources are directed to bots.

03:51:05 Dr. Bhatt: Okay, Wayne and then Jessica, feel free once Wayne is done.

03:51:11 Dr. Goodman: So, I'm currently involved with a NIH brain initiative funded study in bipolar depression in which we do deploy a computer vision machine learning, looking at automated facial affect recognition to identify different mood states. We combine that with speech analysis. We have wearables looking at activity and that's just a few of the things we do. We are seeing some very encouraging findings that help us-- That correlate with clinical improvement in depressive symptoms, but in my own experience at least I don't think we're ready. I wouldn't make a claim that we're ready based upon even using that whole repertoire of different sensors to make a diagnosis accurately. So, I think-- I would need to see more evidence other than one study that was presented earlier in order to be sure. And what we'd be talking about, we'd be talking about more than a chatbot. Now we'd be talking about that there would have to be speech analysis and computer vision machine learning analysis.

03:52:25 Dr. Bhatt: And I apologize I did not scroll to the next page. Specifically, the FDA team did write disorders relating to sadness. So, all on the spectrum. I think by that, if I'm correct, we are not talking about a different set of psychotic psychiatric disorders and others, but staying in the sadness realm with multiple disorders. Just to clarify. Jessica.

03:52:48 Dr. Jackson: Jessica Jackson. I have a comment and then some things I'd like to see. I understand the risk and I agree with my colleagues that we are not there yet. I think this is a great opportunity to set the bar for what we would like to see in 10 to 15 years so that we're not in the position we are now playing catch up where people have already built things without having guidelines. I think there's a way for us to think boldly: what is the bar? And I do think that there is a bar for, we're talking about diagnosis and treatment if we're not here yet, I'd like to see people developing multiple layers of assessment. Similar to what happens now. There are master's level clinicians that do diagnose depression without thinking about the medical rule outs and they are allowed to do that by their license, but they also typically work in conjunction with a psychiatrist or with the PCP, or somebody else. So, I would like to see that same type of multiple layers of assessment in a device where someone is saying “Over-the-counter for MDD without HCP.” It's like, have they built into the product biomarkers and other types of assessment to do rule outs? And include some medical assessment as well as a part of the product development.

03:53:58 I also think that I would like to see what their plan is of how it gets-- What the actual market deployment would be. I think about things like pseudoephedrine for pharmacists. You can get over-the-counter allergy medicine but you do have to have an ID if it has that D on the end of it and a pharmacist has to check it and there are repositories for them to track how often you have purchased it or are you allowed to purchase it? I think this would be an opportunity if someone is developing this, there is going to be a very high bar for it to be autonomous. For me, one of the things would be how are we determining that people who would not meet criteria to get something OTC would be taken from it? Have you developed a repository? Are pharmacists a part of the training? When we say OTC, are they buying it online or do they actually still have to go into a clinic and purchase it with ID? So, I think that's something else that I would like to see.

03:55:48 And I think the final thing, if we are saying that something is going to be able to diagnose, I think this-- To me, I would like to see systems that have been refined and I would think they would need to be an R&D for people who watching research and development for years because I think this is also a great-- I mean, we are moving into the space of doing some precision mental health. I think this is a great opportunity to start looking at what are the specific types of symptoms we can move from the way that we diagnose and treat now. So, in 20 years it is going to look different and I would want to see in the development of this, of novel technologies that people are starting to build and say this is how we are parsing out symptoms. Here is the data from 10 years worth of data that we've collected or that we've collected from multiple clinical trials. This would not be something that would be one clinical trial. So, I would want to see a long time in R&D and lots of specific evidence.

03:55:44 Dr. Bhatt: Thank you. Wayne.

03:55:45 Dr. Goodman: Yeah, just a quick addendum. I did forget to mention one modality that we found very important, which is sleep. So, we have-- As part of this study, patients wore an Oura Ring and we got readouts every day of how much they slept. I sound very negative at the beginning about this and about the ability to diagnose it, but we are studying it and I think that in some future date we would be able to have devices that ingest data from multiple modalities using wearables that would allow us both to help with the diagnosis and track symptom progression.

03:56:23 Dr. Bhatt: Fantastic. All right, we will summarize for the FDA and then see if they have any questions, and then move on to number three. So, start at the beginning. This is making the device available over-the-counter for people diagnosed with major depressive disorder. To the FDA team, this makes the Committee nervous. However, we manage risk all the time in psychiatry and we recognize that access is important and clinicians may present a limitation to access to care. With that in mind, still over-the-counter care for severe depression would concern some of our members. Our Committee might prefer that this is aimed at PHQ-9 less than 10 and other such metrics, such that it is mild to moderate and not moderate to severe depression. We also would like a clear definition of a digital therapeutic versus digital support, but if this in fact is to be a medical device, our Committee would like to see trials that demonstrate that an autonomous over-the-counter device performs as well as at least a semi-agentic device that has a human in the loop.

03:57:30 We do need to go back to our risk mitigation framework if we consider this just the autonomous case. For that, we would require a mechanism to know that they have the diagnosis of major depressive disorder. Research would require high quality, randomized controlled trials with investigator initiated evidence, a large number of individuals using it for a longer period of time, and a very favorable safety profile from suicidality and other severe events because of the rapidity of scale when something becomes over-the-counter.

03:58:03 We do also want to see a strong encryption program since sharing might be a concern, kind of like medication sharing and evidence that it doesn't do harm. Significant data coming back from this from large numbers of people may actually be an opportunity for us to confirm ongoing safety and risk profile. The other things that we'd like to know are: How do you get the actual medical diagnosis beyond the likelihood ratio that might be determined by biomarkers through an AI? How long ago were they diagnosed? Is it new or are they already managed? What was their response to prior treatment? Is this their first therapy? Are there previous episodes of suicidal ideation Who diagnosed them? Are they still in care? Should we be cooperating with those people that are caring for them? Can the tool assess severity and can it assess improvement or worsening? And then again, back to the beginning, technologically, can it in fact diagnose? Lastly, does it have the ability to escalate care if it recognizes that there is no improvement or worsening of disease?

03:59:10 Now, modifying the over-the-counter device to autonomously diagnose and treat without the involvement of an HCP extending to people who have symptoms of depression. Our Committee believes that the lack of diagnosis is the additional complication. Now we are not requiring a diagnosis of major depressive disorder. From a tech standpoint, if the software can diagnose a likelihood of major depressive disorder or spectrum, then perhaps technically it's an appropriate extension. However, from a clinical perspective, the lack of knowledge of additional comorbid disease, a medical assessment poses real clinical concerns to this group at this time for missing other disease diagnoses or knowing that this is in fact the correct diagnosis.

03:59:54 Lastly, when modifying to include multi-use indications for diagnosis and treatment, our Committee still has all the requirements and concerns as above, but we would like to see people diagnosing multiple levels of diagnosis. We know that there is a need. Then, understanding how to assess at multiple layers and multiple risks, and build into the product a way to rule out individuals who would need medical care in addition to the software-based care. And then, be able to risk stratify the type of care for those individuals who remain. Tracking and inclusion of a large number of people, including perhaps at the pharmacy level to have this as at least controlled access behind the bench if you will, at a minimum. And then remembering, if over-the-counter, there is an opportunity to diagnose disease, but we don't want to diagnose and treat mental health in isolation. So, thinking about sleep disordered life, thinking about substance use increasingly as there is movement in the digital or software field to think about these other associated diagnoses. It would be nice to see mental health not treated in isolation, but along with sleep and other concerns as well. I'll ask the FDA if they have any questions or my colleagues if I've left anything important out.

04:01:20 Ms. Paulsen: No questions. Thank you, Dr. Bhatt.

Question Grouping 3

04:01:23 Dr. Bhatt: Okay, I believe I'll now move to ask Dr. Guerrieri to please present the last question of the day.

04:01:35 Dr. Guerrieri: Question three expands on the previous questions and scenario to a population that includes a child or adolescent for devices. That is 20 years of age and younger. So, what are your thoughts as you consider the manufacturer's proposed changes? Please discuss whether your prior responses to question one would change if the population were children or adolescents. And the second part is, if so, how would that response change?

04:02:10 Dr. Bhatt: Okay, let's start with considering the child or adolescent. We're looking at these proposed changes. Would our prior responses to question one change if the population-- Or children or adolescents--? Let me go back to one for a moment for you. Question one was the following. Maybe you can put it up on the screen too if we have it. Is that possible? Thank you. Consider that a healthcare provider prescribes the digital mental health medical device to the less than age 21 patient to use independently at home. Let me just confirm with my FDA colleagues, you are not asking us about the autonomous over-the-counter use for people less than age 21. You're asking us to stick with question one about the healthcare provider prescribing digital mental health to the pediatric age group. Is that correct?

04:03:12 Ms. Paulsen: Yeah, Dr. Bhatt. So, start with question one, that scenario, but changing the population and then, if time-permitting, if you're able to move through the next modifications of that scenario for this population, that would be helpful.

04:03:23 Dr. Bhatt: Okay, fantastic. Team, bring those nerves forward. Let's drop the slides so I can see your faces and let's talk about prescribing this for children. Chevon.

04:03:39 Dr. Rariy: I just want to say for the record that as we get down to prescribing this in children autonomously, I would recommend we take great caution. As a clinician at the intersection of technology and medicine, I've seen technology transform lives, right? I'm absolutely a proponent. I believe in the potential of AI. I believe in the potential of AI to improve how clinicians can be more human, not less, but I think as doctors and many of us might feel similarly, we took an oath to do no harm. And I think for young people I have seen--- I'm a parent, I'm a mom, and I have seen my kids completely absorbed with the device and the impact that it has on their childhood, on their play, on their creativity, on their connection. It's unmistakable. And for me, I would say that it is vitally important to have more data. We don't have data for this specifically for children. And so, for me it would need to be a very high bar. We need strict clinical evidence and safety and efficacy before I would suggest that we consider it to be used in anyone under 18 where their brain is still developing. And so, to basically experiment on children in this regard is a little-- The bar needs to be very high.

04:05:15 Dr. Bhatt: Okay. And Chevon, that response of yours was to extending it to over-the-counter. We're going to move back and I'm going to ask everybody else, Chevon got away with it, to really just stick with prescribing a software to manage children under age 21 with major depressive disorder. Jessica. Prescribing.

04:05:40 Dr. Jackson: Yes, I think prescribing-- For me, the bar is to use the word therapy, I think that's where I get hung up on. Therapy feels clinical. However, there are states right now who have purchased digital tools that are available for, I think almost down to age eight for all the youth in that state, to use for free. And they provide tools and they consider it coping, so they don't call it therapy. So, there is evidence that for youth these things can be helpful because it is something that they feel they can trust versus social media and that is something that is happening right now. So, I think that we do have some evidence, if we're just talking about the benefits, that there is a potential for there to be benefits, especially for youth who are in areas where they cannot consistently see a provider and/or stigma.

04:06:29 As someone who's worked in college counseling centers, you do have states where students cannot get support without the consent of an adult. If you have an adult who does not believe in mental health, you cannot get treatment until you are of age. And so, they do turn to the internet because their parents will allow that. They have found that some of these tools that the states have purchased for their youth and developed in conjunction-- So, these are things co-designed and co-developed for their youth, to give them away for free, can be useful. So, I think that is a benefit. I do, though, urge caution, similar to Chevon, especially when we're talking about clinical. I urge caution in using clinical language like diagnosis and treatment, even if it is prescribed. I do think there's benefit for tools and I think there's a role FDA might be able to play in that, keeping in mind that tools are already being done. I don't know what that role could be in the after effect.

04:07:25 Dr. Bhatt: Okay. Who else? Think about your children, friends of children or children of friends who are trying to get access to care, who have long wait times, who are at college in a different state. Omer.

04:07:46 Dr. Liran: So as before, if there's a human in the loop, a provider is prescribing this application, I'm much less concerned because if any problems, they're being monitored hopefully by the provider, at least they're supposed to be. If parents find that there's any change in behavior that is trending towards negative, they can contact that provider, they can stop the application. So, that's very different from over-the-counter. As long as there's some evidence that it's not harmful, I would be comfortable having a provider write a prescription, whether it's for an adult or a child, because especially a child, they're hopefully being supervised and both by parents and the provider.

04:08:38 Dr. Bhatt: Can I push a little bit on this? Because I'm-- My own thoughts, can I ask people about teenagers? Ability to tell how their mood is changing, likelihood that they follow up with care and the differential between that and over-the-counter. Does anybody just have thoughts on prescribed versus over-the-counter in this age group who's different? Maybe moving around, maybe getting their first job. Melissa, I'm looking at you. Thoughts about what you would like to see from the company to be comfortable saying we have a pediatric mental health crisis in this country. I'm aiming more at the teenagers right now in my conversation as you can tell, but it is true for younger ages as well. What do you want to see when we go back to the beginning of this question from those companies that might be more or different than for adults? Or is it the same?

04:09:47 Dr. Clarkson: I don't have kids so I have no reference for which to answer.

04:09:51 Dr. Bhatt: That's okay. I still trust your job.

04:09:54 Dr. Clarkson: I think-- So, if we're in the prescribed realm of this, I would feel much more comfortable because I-- Well, I think the unknown factor is the role of the parents. Is it better to have the parents quite involved or is this something that would be more effective if the child feels like this is not something that's being monitored by the parent and I don't know.

04:10:29 Dr. Bhatt: Okay, so safety then perhaps in the clinical trial would be obviously the most important thing. Ways to measure that kind of safety. Ray, I'll ask you to ask your question, make your comment, but also tell me a little bit about safety in pediatrics in this realm and your thoughts.

04:10:46 Dr. Dorsey: I hang out at the other end of the age spectrum, but I think longer term safety data would be useful for teenagers for example. So, the difference between 40-- A 40-year-old getting prescribed therapy and they're functioning at 42, so not very different from a 14-year-old getting prescribed therapy and the effects of that on their development at 16. And again, as I said at the outset, we have learned the hard way what the long-term consequences are of excess use of smartphones and smartphone applications. So, we should take lessons from that and look at longer term safety data before going into this population, especially teenagers and younger.

04:11:29 Dr. Bhatt: Okay. David? Yes. Thank you.

04:11:34 Dr. McMullen: Thank you. Yeah, so one thing that we thought would be helpful to clarify is we've heard about a wide range of children and adolescents, so if you could help clarify as you go down in age 21, 18, 16 below to younger children, is there an age at which you are more comfortable with these tools being used that you feel they're more similar to the adult usage? But if the panel could clarify that, if there's maybe gradations of first, that would be helpful.

04:12:06 Dr. Bhatt: Thank you. Tom.

04:12:09 Dr. Maddox: Yeah, I'll answer that. And also just a question that was occurring to me before I heard David's question. I went back and looked at the language of question one and I guess I'm hanging up on the word "independently." So, we say in the first scenario, "Consider a provider prescribing the digital mental health device to the patient to use independently at home." And I think, because of that word, applying it to an adolescent population, I don't think-- I'm less comfortable with that. I think it needs to be accompanied by a human, a professional of some sort. That this just can't be used in isolation because of, I think, some of the particular realities of where adolescents are in their development, self-regulation, et cetera. And I think we put, I'm not a pediatrician, but I certainly have the kids, but it is-- I think what I understood from my pediatric colleagues is that sort of thing needs a lot of close connection to that care provider and when they're trying to work with kids through mental health issues, et cetera, that kind of relationship is really critical. So, I would say a device may be useful for the reasons we've talked about, but I think in isolation, I don't think the individual in general is at a point where they can do that safely, but obviously you'd want that as part of the evidence.

04:13:44 To David's question, I don't know if-- There's a gradation and there's a lot of individuality about relative maturity and responsibility that different people have at different ages. But my understanding, and you a hundred percent should check me on this, is that the neuroscience kind of landed on full maturation is not occurring until kind of the mid twenties, believe it or not. People I've heard said that the brain does not fully formed from executive functioning until age 26, which I got to tell you anecdotally that checks and so I just think you'd want to consult with people around that it's both a legal and a biologic issue and you'd want to speak to people more expert in that than me.

04:14:32 Dr. Bhatt: Okay, let's go to Melissa and then Omer.

04:14:38 Dr. Clarkson: In reading over this, it says child or adolescent. So, it's possible that this could be built into the form of a talking teddy bear, which is-- Some companies are going to make it if that's what can be done in addition to a chat on your smartphone. So that kind of puts a new spin on this. I have no suggestions or answers.

04:15:03 Dr. Bhatt: I like the talking teddy bear concept because you're introducing to us the concept of what companies are doing and how we create things for people and what is right for which age. Omer. And then I have a question or two.

04:15:17 Dr. Liran: Regarding the age, well really depends on how the app was studied and what age population, but I think we should keep in mind that extensive screen time can be detrimental. I'm wondering if there is a dosage equivalent. So, a provider prescribes 20 minutes a day, some 20 minutes a day or whatever it is, and the user can't go beyond that. It basically locks them. So, you can't have a kid who's talking to this AI agent five hours a day. Also, there has to be transparency. Someone should be able to see what the child is speaking to the AI about. Maybe getting summaries or the actual transcript itself. But somebody has to be monitoring and if the provider is the one who's prescribing it, then it's their responsibility to monitor to make sure that it's being used responsibly.

04:16:09 Dr. Bhatt: All right, so in the prescribed setting, really thinking about what is being asked, what is being discussed, some analysis of that being provided back to the provider on a rather continuous basis. And then thinking about exposure to screen time, if it is a screen mechanism for this digital device or the software as a medical device, for maybe lockouts or specific times in the screen.

04:16:37 Dr. Liran: Yeah. I'll just answer that. Also, if it's not screen time, if it's not Teddy there still, you probably want to limit whatever it is you're speaking to the AI agent. You want to make sure you have a limit on that interaction.

04:16:48 Dr. Bhatt: If the FDA ever has a “How long you can hug your teddy bear” label, we are going to bring that back to this group and say, “It was because of us.” Chevon, over to you. This is a challenging conversation. Thank you all for engaging in it.

04:17:01 Dr. Rariy: Thank you. I will just make that point Dr. Bhatt as well, that I am just listening to the silence, right? I think a lot-- I think we're all processing, which is very interesting to kind of be a part of and to see. So, just appreciate that. The second thing I want to say is I would suggest for these particular cases where we just don't know, I think we need to insist that there's heavy oversight of specialists. For example, the pediatric psychiatrist, a group or convening, or some deeper understanding of the device, the monitoring that might be required is different than what we would see in the adult app. So, taking an adult app and translating it to children, we all know, especially as a clinician, that children are not mini humans or not mini adults, they're distinct. Having that very specific subject matter expertise, I think it's particularly important in this point. I would certainly include that as a mandate. You need this specific subject matter expertise, honing in on what does the analysis look like? What does their randomized control trial study period include? What age groups are appropriate? All of that I think needs to be decided and the best individuals to do it would be subject matter experts, particularly for this.

04:18:42 Dr. Bhatt: Thank you. Jessica.

04:18:50 Dr. Jackson: I think, especially thinking about this scenario, like the specific language, I would want to see anything that is developed, treated the same way as it would be if there were humans. I have treated and diagnosed [children] as young as age eight. There are, at least in my case, for psychologists, like pediatric psychologists, they have specific training. Their therapy is developmentally appropriate. They're not doing ACT and CBT the same way they would with an adult. So, I would want to see anything that's developed, that is developmentally appropriate because youth, adolescents and children do get therapy. They do homework in between therapy right now. They do get medication. So, I think there are ways to develop a product that is developmentally appropriate with some of the safeguards people have mentioned for dosing. If it is software, there is a way for people to develop it in a way that it cuts off after a certain amount of time. So, if your dose is 30 minutes two times a week, there's a way that they could probably code it to cut off after 30 minutes and it's been used twice in a seven-day period. I would want to see those types of things in the product development.

04:19:52 And I would also want to see specialized training to prescribe it. Similar to how people are looking at psychedelics, right? There are people who are saying, “It needs to be a specific clinic.” “You have to be able to be working in this very specific space.” I think that that's something that we could be asking of this as it's getting built and developed is saying that there does need to be HCPs who are trained very specifically in working with pediatrics and thinking about digital tools to be able to decide who can it be prescribed to and to be able to manage it, right? Presumably, we can also say in product development that the HCP is getting the data back so they are able to monitor some of these conversations. I think the difference also in a software like this is we could ask for a closed loop. So, it isn't something that is completely-- That they're getting any type of answer. For pediatrics, it would need to be developed very differently. So, I think there's a world where we could ask for some of these higher bars in the product development that would make it developmentally appropriate treatment the same way that we do therapy and psychiatry currently.

04:20:53 Dr. Bhatt: Thank you so much. Phil.

04:20:58 Dr. Rutherford: Yep. I've kept largely quiet on the prescribing stuff because I'm not a prescriber, but I want to highlight something that's coming up as we're talking about moving these products through the regulatory process. I'm struck by what, I believe, the two gentlemen from MIT said earlier about the regulatory process and how some of these products get to market. Just a flag for the FDA that I'm curious about how if we are thinking about this in the way that they described that some of these products are evolving on a daily or weekly basis, I just want to leave a flag for that. I don't know if it is sufficient to think about, "Okay, yes, we got to move these products through the regulatory process, and premarket and postmarket," when we're talking about something that is evolving daily. That's all.

04:21:53 Dr. Bhatt: Okay. All right. Why don't I try a summary and then the FDA can ask some more questions of us. So, for this question, you have asked us to consider the manufacturer's changes, but you want them to treat children and adolescents aged 21 years and younger. Our Committee is so concerned that we don't know what to say. Some because of fear, some because not all children are the same. Some because the relationship between child and caregivers can be challenging and unpredictable. And most agree that child psychiatrists are central to this development process of this device and a recognition at the same time that wellness is so important at this age and a lack of access to care in this age group is incredibly unfair. So, this is an important discussion, which leaves us a little bit speechless. We do believe that large well-controlled trials with experts is a must, that the device must be developed as developmentally appropriate for each of those ages you're asking us. So, the product has to be developed in a way that safety measures are included in a developmentally appropriate way. An example that was given frequently is screen time and usage cutoffs, but also specialized training to prescribe these digital devices and to understand how an individual changes with them as their age changes and they move from one developmental stage to another. There must be a return of data of these conversations back to the clinicians. Again, echoing privacy, security risk for these children and then digital and health education is also essential in concert with this therapy.

04:23:37 When we move to thinking about over-the-counter, maybe the first comment that was made in this group is the most appropriate. We need to differentiate between wellness and medical therapy very carefully in this arena for children. There needs to be a framework for how these products are differentiated along the spectrum of wellness support and in fact medical therapy. And there needs to be some mechanism to ensure that scope creep does not happen. I don't know that our group had suggestions on avoiding scope creep. I know I'm not supposed to go straight to the FDA for any questions, but I might just ask the group-- We started talking a little bit about wellness versus therapy in this age group. We fear this idea of scope creep in adults, I'm assuming that carries over in children. Is there anything additional about that anyone wants to add before we ask the FDA if they have questions for us? No, we're still speechless. FDA questions from you for our Committee.

04:24:51 Dr. McMullen: No additional questions. Thank you.

Closing Remarks

04:24:54 Dr. Bhatt: Okay. I would like to thank the Committee and the FDA for their contributions. I would also like to thank, again, the Open Public Hearing speakers, patients, industry, healthcare providers, academia and FDA for their remarks. This is a tough subject. We are really privileged to be here and be able to talk about it and we are so glad for all the people who are joining us online. As you can see, there are a multitude of industries ranging from patients themselves, family, caregivers to industry entrepreneurs, technologists, regulators, other government agencies and clinicians who all really feel a vested interest in addressing the mental health concerns that we have amongst so many Americans.

04:25:45 I think what we've come to today is largely that there is an infrastructure that's needed and there are many people interested in helping create that infrastructure moving forward, and that there is really scientific rigor that can and has gone into this field that should still make its way all the way to products that are used at the point of care in the communities where people live. This is a real problem we're dealing with. We are honored that the administration allowed us to continue this meeting, that everybody here was willing to tackle very difficult subjects and we look forward to continued discussion, follow up with the FDA and really seeing where a lot of these people who are working so hard continue to make an effort.

04:26:34 I'll end my comments by pointing to the last group we talked about, which is I think, the next generations. We saw a couple of the younger people earlier today speak with us, and even those who are teenagers and young adults now, they are suffering from mental health issues. They are dealing with them as they're continuing to grow, and I would just encourage anybody in that population or who has friends, relatives in that population to teach them now about advocacy and to speak up and to be part of this process along with us. Before we adjourn, I would like to ask the FDA representatives if they have any concluding remarks. Ms. Paulsen and Dr. McMullen, over to you.

04:27:16 Ms. Paulsen: Thanks, Dr. Bhatt, very briefly, just on behalf of FDA, I'd like to thank Dr. Bhatt for chairing the meeting and the Committee for their really helpful input today. I'd also like to thank the invited speakers, Open Public Hearing Speakers, patients, industry, healthcare providers, and academia representatives for their remarks on this very important topic. Lastly, I'd like to thank everyone who listened to the discussion and to those who provided comments to the docket thus far. So, just a reminder that the docket, it's going to remain open until

December 8th to provide additional opportunity for public feedback after today's meeting. So with that, we'll turn it back to you, Dr. Bhatt, to close the meeting.

04:27:53 Dr. Bhatt: Absolutely. Well, thank you all for joining us at the November 6th Digital Health Advisory Committee meeting. This Advisory Committee is now adjourned. Have a wonderful night and thank you again.