This transcript was assembled from Artificial Intelligence (AI) generated captions and may contain errors.

Session 4: Drug-Device Combination Products (Part 1 and Part 2)

Moderators:
William Chong, MD — Director, OSCE, OGD, CDER, FDA

- **Public Comment Presentations on Drug Device Combination Products**

- **Comparative Use Human Factors Studies: Challenges and Recommendations**
  Brandon Wood, BS — Director of Regulatory Affairs, Generic Steriles (Teva Pharmaceuticals USA, Inc.)

- **We Muddled Our Way Through the CUHF Process, Now What Does It Mean?**
  Melissa Lemke, MS — Regulatory Human Factors Engineering Advisor, Human Ability Designs, LLC

- **Comparative Threshold Analysis – So Near, Yet So Far …**
  Vivek Viswanathan, PhD — Manager, Research & Development, Rubicon Research Canada Ltd.
  Daliya Bharati, MS — Director, Regulatory Affairs and Intellectual Property, Advagen Pharma, Ltd.

- **Industry Perspective: Development of Generic Emergency Use Products**
  Amy Lukau, BA, BS — Senior Human Factors Lead, Kindeva Drug Delivery

- **It's Hip to be Square: Demonstrating Equivalency without Inferiority in CUHF Studies**
  Heidi Mehrzad, MS — CEO and Human Factors Expert, HFUX Research, LLC

Panel Discussion

In addition to moderators and presenters listed above:

Public Panelists:

Tim Briggs, MSc	Senior Principal Human Factors Engineer Global Device Development, Viatris
Megan Conrad, PhD	Associate Professor of Mechanical Engineering, Univ. of Detroit Mercy
Manoj Pananchukunnath, MP	Chief Scientific Officer, Scientific Affairs, Biocon Ltd.
Vivek Viswanathan, PhD	Manager, Research & Development, Rubicon Research Canada Ltd.

FDA Panelists:
Robert Berendt, PhD — Supervisory Chemist, DPQA V, OPQA I, OPQ, CDER, FDA

Ariane O. Conrad, PharmD	Associate Director for Human Factors, DMEPA I, OMEPRM, OSE, CDER, FDA
Jason Flint, MBA, PMP	Deputy Director, DMEPA I, OMEPRM, OSE, CDER, FDA
Kyran Gibson, BS	Biomedical Engineer and Lead Reviewer, DHT IIIC, OHT III, OPEQ, CDRH, FDA
Stella Grosser, PhD	Director, DB VIII, Office of Biostatistics, OTS, CDER, FDA
Edna Termilus, MD, MPH	Associate Director, DCR, OSCE, OGD, CDER, FDA

Closing Remarks

Robert Lionberger, PhD	Director, ORS, OGD, CDER, FDA

Session 4: Drug-Device Combination Products (Part 1)

William Chong (FDA): Welcome back everybody. Hopefully you were able to enjoy your lunch, or at least a break from all the discussion and the wealth of information that was presented earlier today. My name is William Chong. I'm the director for the Office of Safety and Clinical Evaluation in the Office of Generic Drugs. It's my pleasure today to be moderating our Session 4 on combination products, specifically drug-device combination products, and focusing on what are the research needs to help support generic drug development in this space.
Our session is going to start with one public comment that we received. We'll pivot into our presentations, take a short break, and then get into our panel discussion at the end, where we'll start with some questions for the speakers that other panelists might have, and then we'll turn to our panelists who haven't had a chance to share their perspectives and see what they heard, what they thought, if there's anything else that they'd like to add, and hopefully stay on time.
First, let's take a look at our public comment. William Feldman, who works with Dr. Aaron Kesselheim at Harvard Medical School, will be presenting one public comment for us, and then we'll pivot straight into our presentations. Hopefully, Dr. Feldman will be able to join us during the panel for any questions that you may have for him. I will turn it over now to Dr. Feldman. Thank you.

Public Comment Presentations on Drug Device Combination Products

William Feldman (Brigham and Women's Hospital and Harvard Medical School): Hi. My name is William Feldman. I am a pulmonologist, intensivist, and health services researcher. I'm presenting today on behalf of the Program on Regulation, Therapeutics and Law, an independent research organization at Brigham and Women's Hospital and Harvard Medical School focused on prescription drug prices and policy. I'd like to thank Dr. Aaron Kesselheim. I'll be talking today about real-world evidence for generic drug-device combinations.
The United States pays among the highest prices in the world for brand-name prescription drugs. Drug-device combinations, which contain active pharmaceutical compounds built together with their delivery devices, are among the costliest products now on the market. Forty percent of the top 50 drugs by gross Medicare Part D spending in 2022 were drug-device combinations. These included blockbuster drugs like Ozempic and Lantus SoloStar for diabetes, and Trelegy, Advair, and Symbicort for asthma and COPD.
High prices in the US are often accompanied by high out-of-pocket costs, which can lead to reduced adherence and worse health outcomes for patients. A key mechanism for reducing drug prices is generic competition. Generic drugs represent 80% of prescription dispensing in the US

and just around 20% of pharmaceutical spending. Yet generic competition for drug-device combinations has often been slow.
A key reason is that brand-name firms obtain large thickets of patents on their drug-device combinations, including large numbers of patents on delivery devices. In addition, the FDA applies a necessarily rigorous set of testing requirements when approving drug-device combinations, given the complex interactions between the device and the active ingredient and the need for proper technique by patients.
Despite the many regulatory challenges associated with evaluating generic drug-device combinations, the FDA does not actively surveil adverse effects after approval, relying instead on data submitted through FAERS and manufacturer pharmacovigilance databases. While these efforts are important for cataloging adverse effects, passive surveillance is subject to lots of potential for bias, including reporting bias, under-reporting, difficulty capturing event rates due to missing denominator information, and insufficient granularity to distinguish between specific brand-name and generic products.
Longitudinal healthcare databases can help address these limitations by allowing researchers to study the safety of drugs in populations of patients treated in routine clinical practice. Well-conducted studies generating real-world evidence rigorously control for confounding based on comprehensive data relating to patient comorbidities, procedures, prescriptions, hospitalizations, and so forth. These studies have the advantage also of broadly including patients across racial, ethnic, and socioeconomic groups that are often underrepresented in clinical trials.

We think that routine use of real-world evidence in the post-marketing setting could provide important reassurance for patients and physicians relating to complex generic products. It could also potentially allow the FDA more flexibility when approving complex generic products.

So we would recommend rigorous, replicable surveillance of complex generic drugs like drug-device combinations. We would also recommend that the FDA develop policies and procedures for triggering further studies from generic manufacturers if safety signals are detected using real-world evidence. And finally, we would encourage the FDA to obtain feedback from the public, from academics, and industry on ways to streamline the approval process for generic drug-device combinations, knowing that post-marketing studies will be completed. Thank you very much for your time.

William Chong: All right. Thank you to the public commenter for their presentation. I do apologize. I forgot to introduce the panel. I do want to welcome everybody. Thank you for your participation. For our public panelists, I'm just going to go alphabetically.

We have Daliya Bharati, Director of Regulatory Affairs at Advagen Pharma; Tim Briggs, Senior Principal Human Factors Engineer with Viatris; Megan Conrad, Associate Professor of Mechanical Engineering at University of Detroit Mercy; Amy Lukau, Senior Human Factors Lead for Kindeva Drug Delivery; Heidi Mehrzad, CEO and Human Factors Expert at HFUX Research; Melissa Lemke, Regulatory Human Factors Engineering Advisor with Human Ability Designs; Manoj Pananchukunnath, Chief Scientific Officer for Biocon; Vivek Viswanathan, Research and Development at Rubicon; and Brandon Wood, Director of Regulatory Affairs Generic Steriles at Teva.

For our FDA panelists, I'd like to welcome Robert Berendt, Supervisory Chemist, Office of Pharmaceutical Quality Assessment I, OPQ, for FDA; Ariane Conrad, Associate Director for Human Factors in the Division of Medication Error Prevention and Analysis I in the Office of Surveillance and Epidemiology; Katharine Feibus, Director for the Division of Therapeutic

Performance I in the Office of Research and Standards, OGD; Jason Flint, Deputy Director for the Division of Medication Error Prevention and Analysis I in OSE; Kyran Gibson, Biomedical Engineer and Lead Reviewer in the Center for Devices and Radiological Health; Stella Grosser, Director for Division of Biostatistics VIII in the Office of Biostatistics; and Edna Termilus, Associate Director in the Division of Clinical Review in my office, Office of Safety and Clinical Evaluation in OGD. Thank you again for your participation.

With that, I think we'll turn towards our invited public speakers. Our first talk will be from Brandon Wood, Director of Regulatory Affairs at Teva, who will be speaking to us about comparative use human factors studies: challenges and recommendations.

Brandon Wood (TEVA): Wonderful. Thank you for the introduction, Dr. Chong. First, I'd like to thank the agency for the opportunity to speak today. Second, I would like to acknowledge everyone who's still here on day two, session four. I know post-lunch presentations, at least the first couple, can be a little sleepy, so I'll try to bring the energy and keep your attention.

The title of my presentation today is "Comparative Use Human Factors Studies: Challenges and Recommendations."

We have our standard disclaimer slide. We'll jump right into the agenda. I'm going to go through regulatory requirements for generic substitutability, current expectations for comparative use human factors studies, challenges with current expectations, alternative study designs, recommended areas of FDA research, and conclusion. A lot of topics, but I promise you it streamlines.

Given that I am a regulatory professional, I thought it was appropriate to start with the regulatory requirements for generic substitutability, and specifically the history and statutory framework. Sometimes to move forward, you really need to look back. For us, our story begins 40 years ago—pretty amazing, 40 years ago at this point—with the Hatch-Waxman amendments, where sections 505(b)(2) and 505(j) were added to the FD&C Act to describe abbreviated approval pathways for products regulated by the agency.

Ultimately, this reflected Congress's efforts to balance the need to make available more low-cost generic drugs by establishing a generic drug approval procedure. So the 505(j) was born. The abbreviated new drug application was born, and an ANDA is a drug product that is a duplicate of a previously approved drug product where we have relied on the FDA's findings that the previously approved drug product—the RLD—is safe and effective. Ultimately, the ANDA is going to contain information to show that the proposed generic product is the same as the RLD with respect to the active ingredient, conditions of use, route of administration, dosage form, etc., and is also bioequivalent to the reference listed drug.

Importantly, an ANDA may not be submitted if clinical investigations are necessary to establish the safety and effectiveness of the proposed product. We also have therapeutic equivalents, so the scientific premise underlying the Hatch-Waxman amendment requires that a drug product approved in an ANDA is therapeutically equivalent to its reference listed drug.

Therapeutic equivalence really consists of three pillars: one, that the product is pharmaceutically equivalent—same active ingredient, dosage form, route of administration, strength, etc.; bioequivalent, self-evident; and also is expected to produce the same clinical effect and safety profile as the RLD under the conditions specified in the labeling. I'm highlighting this specific pillar of therapeutic equivalence because this is very topical when it comes to comparative use

human factors studies. There's a correlation back to comparative use human factors studies in which we're trying to demonstrate the same clinical effect and safety profile. So we're really going to hone in on this specific pillar of therapeutic equivalence.

In terms of substitutability requirements, the generic and RLD products do not need to be identical so long as the differences do not preclude approval under an ANDA. Generally, as a rule of thumb for any aspect relating to the drug-device combination product—and this is aside from our CMC and bioequivalence requirements for the drug—this means that the labeling should be the same, and the product must be able to be used by an RLD user without the intervention of a healthcare provider or additional training prior to use.

So what we'll do is conduct comparative analyses in accordance with the draft guidance, which systematically identifies any differences in the user interface. Part of the comparative analyses consists of a labeling comparison, a comparative task analysis, and a physical comparison. Then we'll use this comparative analysis collectively with the use-related risk assessment and hazard analyses in order to identify and characterize any difference as either no design difference, minor difference, or other difference.

Once we have characterized our differences or lack thereof, what we'll have is either no design difference, which really is the ideal scenario because then additional data will not be required to support approval of the ANDA; or a minor difference. Minor difference will be a difference that does not affect an external critical design attribute and is likely to be viewed as acceptable, provided that we can provide information that demonstrates that the differences are, in fact, minor. Again, here there's going to be no impact to the clinical effect or safety profile.

Or there's going to be other differences. Other differences are differences that may impact an external critical design attribute that involves administration of the product and may impact the clinical effect and/or safety profile. First and foremost, we must strongly consider modifying the design of the user interface to minimize the differences present between our proposed product and the RLD. However, if the differences are present in the final design user interface, then we need to provide additional data such as data from a comparative use human factors study. The purpose is to determine whether or not the differences that have been identified introduce a new risk that might impact the clinical effect or safety profile. So you can see it's come full circle here, going back 40 years ago to Hatch-Waxman, where we're still trying to demonstrate the same clinical effect and safety profile. It's just for this specific instance, and with the presence of other differences, we're utilizing a comparative use human factors study to do so again when there are other differences.

So what are the current expectations for comparative use human factors studies? I have a very quick snapshot here that provides the objective, execution, evaluation, and statistics as well as success. The objective is to demonstrate that the use error rate associated with a change in external critical design attribute for the proposed user interface does not preclude approval of the proposed drug product.

In terms of execution, we'll allow RLD-experienced users to use both the RLD and the proposed generic and take objective measures of their performance while undertaking the tasks that are impacted by the other differences. The study will predefine success for any use task investigated, and other outcomes are considered use errors. Performance of the identified tasks is then compared between the RLD and the proposed generic in a statistical non-inferiority test. Performance by the user on the RLD is taken as the baseline performance, and the proposed generic must be non-inferior to the RLD performance based on the assessed tasks.

Important to note that only use errors are statistically analyzed. Close calls and use difficulties are not part of the statistical analysis as these do not result in medical consequence and are thus excluded. There's also artifacts, which are use events that only occur due to this being a study, and really this reduces the sample size on the tasks that are being assessed.

So what is success? What does success look like here? If the proposed generic is found to be non-inferior to the RLD, then this is taken as evidence that while there are differences in the user interface of the proposed product, those differences do not introduce new risks that would impact the clinical effect or safety profile of the product.

So there are challenges with the current expectations with the comparative use human factors studies and the current paradigm in general. Really, the focus of the current comparative use human factors study is based on a binary endpoint—the statistical non-inferiority between the RLD and the proposed generic. Our position is that this focus does not account for the root cause of use errors. A use error can occur for multiple reasons. While another difference between the RLD and the proposed generic may be responsible for a specific use error, it doesn't necessarily mean that it's the end-all be-all. There may be other reasons or other root causes for the use errors, but the current approach presumes that all use errors are caused by other differences only.

Root cause analysis here would identify negative transfer between the RLD and the proposed generic device. I refer to negative transfer a lot in this presentation. Negative transfer is defined as the interference of previous knowledge with new learning, where one set of events could hinder or adversely affect performance on related tasks.

The exclusion of use events such as close calls and use difficulties also excludes considerations of the root cause analyses of these use events. Similar to use errors, these may or may not be linked to the other differences between the RLD and the proposed generic.

Additionally, there's no consideration of RLDs that are used for chronic conditions. Repeated use of a device with other differences will result in the user adapting to the device differences, whereas the comparative use human factors study design currently is one and done. So what we're not accounting for is the human ability to learn and adapt. It's not considered.

Additionally, there's no consideration of new users that will adopt the treatment as the generic comes onto the market. Lowering the price and increased availability will lead to new adopters of the treatment that are not being considered under the current paradigm.

One of the points that we would like to make—one thing that we'd like to highlight—is that a use error is not a data point. There are significant statistical difficulties in comparing use errors. The reason is that use errors are not binary. They're multilayered, and ultimately root cause analysis is paramount in determining whether or not that use error is related to negative transfer. If it's not related to negative transfer—so it's not related to a difference in external design attribute—then it shouldn't be counted against us for our substitutability evaluations.

Here, if we're conducting a study and we see a failure to complete the dose within the root cause analysis, if on the left here you see we interview the participant and they say, "The button wasn't where I expected it to be," well, that's negative transfer. That's a use error that is the result of a difference in external critical design attributes.

However, if we move forward with a root cause analysis and we interview a participant and they say, "I don't look at my dose display. I just count the clicks the device makes," well, that's not necessarily related to a difference in the external user interface. So that's not necessarily something that should count against us in a substitutability evaluation.

Similarly, for the other example: "I missed my breakfast and my blood sugar is low, so I got confused and I made a mistake." That's not related to a difference in external critical design attributes, so it's not a use error that should be counted against the substitutability evaluation.

Further, you have to ask the question on the two on the right: Would there be success on subsequent use? Which is to say that passive learning is not the same as intervention and retraining, and that's something that should be considered within the scope of these evaluations.

So we also have challenges from a statistical perspective. Calculation of the non-inferiority margin requires information from the RLD that is not available. So what this forces the generic industry to do is to investigate RLD performance prior to conducting comparative human factors studies. It's also been particularly tricky to get agreement or obtain agreement on non-inferiority margins and endpoints. What we've seen in our experience has been multiple attempts, and at the end of the day, this results in time and cost and resource allocation on both sides.

Additionally, criticality of the drug and context of use are not considered in the sense that if we have a chronic condition versus emergency use, when it comes to the statistical evaluation, use errors are not weighted due to their clinical consequences, and this is something that should be explored.

For practical implementation, the comparative use human factors study guidance leads to very large recruitment targets, and there are likely to be diminishing returns based on these very large recruitment targets. New use errors are unlikely to be detected by increasing the sample size.

The agency has also required recruitment of indicated but non-adopted users. What I mean here is, if a medication is indicated for adolescent self-administration, for example, but it's not actually adopted by adolescents, then that's going to present significant challenges during recruitment. So understanding indication versus adoption is a really important topic.

Also, the agency has excluded the use of surrogate participants that are using non-indicated drugs in the correct RLD device platform. So what we have here is real-life gained experience from the RLD device platform not being included in comparative use human factors studies.

So what we're suggesting is an alternative study design consideration. Really, what this is is IEC 62366 applied to generics. So really, what we're saying is this is an evolution of IEC 62366 and not the revolution that comparative use human factors studies are.

In order to facilitate such a study, high level, you create a use specification that's focused on RLD users. You would also allow for overlap of similar RLD populations, and then you would prepare a use-related risk assessment that is focused on user interface differences and specifically sources of negative transfer.

Then, within the human factors study, we would recruit RLD device users driven by the use specification. Notice that "devices" is underlined here because what we would recommend is if

there is an RLD device platform being used for another product, but there's an overlap of the usability comorbidities with our proposed product, then they should be considered for inclusion in the recruitment for our study.

Then we'll also scale the sample size according to the severity of outcome, and in order to facilitate the detection of negative transfer, ultimately the pass/fail determination will be shown by successful use and accompanied with a benefit-risk analysis. With all of the studies that we do at the end of the day, what matters is successful use—successful administration. So what we're suggesting is that the equivalence to substitutability would be shown by successful use and a benefit-risk analysis.

Some additional considerations: We would target recruitment of RLD users to simulate the high-risk scenario where the user has had their prescription changed at the pharmacy. We would focus on use of proposed generics by the RLD device users—again, RLD device users. We will not artificially reinforce negative transfer through the use of an RLD in a randomized fashion. We would focus analysis of results on negative transfer between RLD and the proposed generic. Really, that's what we're concerned with here—differences, use errors that have occurred due to differences in the external critical design attribute. That's what we want to focus on, and that's what the focus of the study should be.

Sample sizes should be scaled according to the safety profile, so you could have smaller, moderate, or larger, depending on what the clinical consequence is for failure to complete a dose. If it's not life-threatening, smaller. If it is life-threatening, moderate. And clearly, if it's critical and may result in death, it should be larger.

There would also be an emphasis on root cause analysis of use errors, use difficulties, and close calls rather than a statistical analysis of use errors alone.

When it comes to pass/fail criteria, 10 to 15% failure to complete the dose may be okay, depending on the root cause and clinical impact and safety profile. This is a product-specific determination. This is going to be evaluated on a case-by-case basis and also something that is very directly correlated to the risk profile of the product. Successfully completing the dose on subsequent use may also be okay, depending on the clinical impact and safety profile of this specific product. This is another case-by-case specific consideration. But again, bioequivalence substitutability is measured through successful use and accompanied by a benefit-risk analysis.

We have done a successful case study where users with reference products—various reference products—were replaced with a two-step auto-injector in a mock IFU. What you see here is we have a three-step auto-injector, two different four-step auto-injectors, and a pre-filled syringe, all being transitioned onto a two-step auto-injector in a mock IFU. Within this, we also simulated the product being switched at the pharmacy without intervention, without additional training, and the results are at the bottom. They're pretty remarkable: 20 out of 20 from the three-step auto-injector were able to successfully complete a dose; for the first four-step auto-injector, 18 out of 20 were able to successfully complete a dose; for the second four-step auto-injector, 20 out of 20 were able to successfully complete a dose; and from the pre-filled syringe, 20 out of 22 were able to successfully complete a dose. So what we're trying to show here is that while there are clearly other differences present in the final design user interface here, dependent on the risk profile of the specific product, they were able to successfully move over to a two-step auto-injector at a success rate that, if commensurate with the risk profile, should be approvable.

So in terms of recommended areas of FDA research, we'd suggest: one, understanding how users adopt techniques for new medical devices over time; two, understanding of what devices have been approved for multiple indications and how different use cases have been afforded by the same design; understanding of indication versus adoption gap—so I had mentioned this before, indication versus adoption, a lot of times what we're seeing is a gap; research and understanding of which areas of medicine are likely to have high incidence of non-indicated drugs being prescribed.

Additionally, an understanding of use errors which are caused by negative transfer versus use errors that are caused by device naivety, and which group are at most risk from the introduction of a new combination product.

Sample size is another very important consideration. What we could suggest is research to conduct large sample size testing to understand normal use error rates for different device types; research to focus available user error data on currently marketed products, making this available to industry—that would be incredibly useful; and lastly, research to investigate how a more traditional, qualitative human factors study design with appropriate root cause analyses of use events may be appropriate to provide evidence of device comparability and substitutability.

So in conclusion, here in the spirit again of Hatch-Waxman 40 years ago, the goal is still to make available more low-cost generic drugs. Availability of high-quality generic drug-device combination products is essential for making important treatment options more accessible to Americans, but there are still challenging regulatory and scientific obstacles that are hindering timely approval and development, one of which is comparative use human factors studies. As they're currently constructed, these are prohibitive to generic drug availability, and alternative considerations are needed.

What we feel is the human factors expertise needed to facilitate timely approvals already resides within FDA and within industry. So let's get back to the basics—the evolution of IEC 62366, fit for purpose for substitution, and not the revolution that comparative use human factors studies are.

So alternative study designs and data analysis considerations could be the key to unlocking critical generic drug-device combination product availability.

With that, I would like to acknowledge every individual on this slide: Henri Akouka, Colin Roscoe, Sophia Edmonds-Allen, Mark Destefano, Carrie O'Donnel, Louise Dunphy, and Leslie Sanchez Torres. This presentation would not have been possible without these brilliant engineers and human factors professionals, so sincere thank you to them and to you for listening. Thank you very much.

William Chong: Thank you, Brandon. I'd like to invite next Melissa Lemke to the podium to do our presentation on "We Muddled Our Way Through the CUHF Process. Now What Does It Mean?"

Melissa Lemke (Human Ability Designs): All right, thank you, everyone. Thank you, FDA, for having us all here to talk about this important topic of comparative use human factors. My title—we really are muddling our way through human factors, and hopefully I can muddle my way through the green button.

So again, I'm Melissa Lemke. I'm a regulatory human factors engineering advisor. I've been in the industry as a consultant for 21 years now, running studies, submitting to FDA and outside the US. I primarily do, at this point, pre-FDA reviews for my clients and training on human factors engineering, and then subject matter expertise. So kind of on-call advice across the industry. I am a caregiver-turned-professional human factors engineer. I'm a biomedical engineer with human factors training. I say that because we really are in kind of an innovative science here of human factors, whereas in medical device we are kind of sitting in 20 years of focus on the human factors validation study, which is a very kind of custom approach for FDA submissions. This obviously is new thinking and a new method.

Importantly, I have a 100% success rate with submissions through the agency, and across the 21 years that's come with creative solutions. So we've been doing a lot of that, which I'll present today—kind of some of the challenges.

Disclaimer: This does come from my client case studies and some survey data, but it's anonymized, just given confidentiality agreements.

So we're going to do kind of a systems analysis of the comparative use human factors world that we're sitting in right now. This diagram comes from "Applied Human Factors in Medical Device Design," which also this diagram is basically covered in the human factors CDRH final guidance.

So we're going to do a play on the comparative use human factors draft guidance as the user interface that we're going to focus on here. We'll think about who are the users of this draft guidance—the user interface obviously still in draft form—and then the use environment. What is going on within industry that we'd like the FDA to be aware of as we're trying to all get to approved generic products on the market for our patients? We need a lot of patients. We heard about a lot of issues happening in the real world right now that patients don't have access. And partly, unfortunately, right now, there's a lot of use error happening with implementing this guidance that is—you know, human factors is holding up drug approvals in the generic space.

So if we think about our draft guidance users for comparative use human factors, obviously we have our important regulators, and there's the pre-ANDA pathway, and then the ANDA reviewer team. So those are two separate teams within the agency. They have processes, and obviously they collaborate and work together, but they do have different responsibilities within the FDA process—the regulatory process.

We have consultants. A lot of consulting firms are picking up more business in comparative use human factors, not only in generics but in biosimilars. We kind of borrow from this guidance as well in that realm. We have kind of new people sitting at the table within consulting and industry, where we're now bringing in statisticians to consult. Typically for human factors validation studies, we're not working with statistics that we know of in our data analysis. Our method is founded on statistics, but a lot of our consulting firms and consultants admit in a survey that we conducted and in interviews, they're not statistics experts, and they don't know how to apply statistics. They've never done that in their work. So that's a big one here that it's really a hurdle to get statisticians at the table that know how to apply human factors statistics.

I would say that's unique in medical human factors. We borrow from aviation, nuclear, and they are doing statistical-based science. So I feel that's a void that really needs to be bridged and better taught and understood for medical human factors. It's a really big hurdle that we're sitting in right now.

Pharma companies—so obviously within our pharmaceutical development, we have a lot of cross-functional stakeholders that have to come together and get behind this method and human factors. You know, we're still—validation studies, we're still challenged with. So drug developers, the regulatory team, quality, risk management—all of those cross-functional stakeholders are still now coming to the table for this method.

Device developers, engineers, designers, platform device developers—so we might want an off-the-shelf product to be used. Well, we're relying on the device constituent-only manufacturers to come to the table with how is your device working? How is that going to work with this new generic drug?

Collectively, we know that there's a lot of confusion right now with this draft guidance, and I would say that FDA has done a very good job of trying to promote this guidance and educate on this guidance. There's a lot of smart people in industry trying to implement this guidance. Ultimately, we're all trying to work together to understand, implement, and get these studies done and the products on the market. But we're not being successful right now. And that's really a public health issue, right? We need these drugs on the market.

We have a lot of experimentation going on with the human factors method in the ANDA process. That's not really a good place to be in science at the regulatory submission stage, right? We want to know that we're going to have a successful plan that we put forward to our reviewers, and they buy into it, and we go and run our study, and it's efficient. That's not happening right now.

And so ultimately, I think probably FDA is frustrated with what they're seeing. And we definitely know that industry is very frustrated with where we're sitting right now on not getting things approved and spending a lot of time, budget, and years of waiting to get a product on the market. And we heard about a lot of hard work happening this morning that's aside from just the human factors challenges that we're now facing as well.

So then again, we think about our user interface. This guidance from 2017 is new, and it brought forth new statistics. And these are the terms that showed up that really had never been seen before. Comparative threshold analysis—that was the new kind of process. I would say that for the most part it's straightforward. Industry is not having a lot of problems implementing the comparative threshold analysis part, documenting what they need to document and going through that. It's pretty straightforward.

But then we get the statistics. And we have that non-inferiority model. This model or method is traditionally based in clinical evaluations and effectiveness. And the interesting thing—so Brandon talked about these other design differences that relate back to critical tasks and safe and effective or successful administration of a drug. Well, that's very different than thinking about the clinical effectiveness, right? We have either you get the drug or you don't get the drug. And even on a chronic use product, if we don't have learning happening, and there's not a learning effect—and how do we measure that? We have all these questions. But basically, if we come down to, did you get the drug? I didn't get the drug. Well, how many people shouldn't have to get the drug for it to be considered equivalent or substitutable? So that's a really big kind of challenge with just thinking about applying this statistical method.

The external critical design attributes—this is an interesting term in that sometimes the internal mechanisms—how I interpret this is we're looking at the user interface, so the external parts of

the device that the user's interacting with and those design attributes. So the buttons, the safety shields, caps, those types of things. But sometimes a feature might be the injection time, or how long you have to hold an auto-injector against the skin. Well, that could be influenced by spring force or some change in viscosity of the drug. Well, if that internal kind of feature changes, is that considered an external critical design attribute? It certainly can impact safe and effective use or complete administration of a drug.

A big stickler—I presented on this two years ago, Brandon just talked about it—but not bringing in the use-related risk analysis. So the foundational kind of document or process that we go through in human factors to bring forward the safety lens is the use-related risk analysis or use risk management. So without this, we are kind of putting all users on a level playing field, and that in implementation, getting back to the root cause analysis, it's really not fair to the user interface. And it doesn't really get us at the level of granularity that we need to know if our user interface is optimized in the human factors validation study, or in this case, if it's equivalent, or substitutable, or equal, or just as safe, or kind of all those terms you might hear.

The minor and other design differences—again, there's kind of some confusion around what does that practically mean in my process, and what am I meaning there? And then use error rates—that's also a big one Brandon talked about. So just counting use errors—is that really getting at does a person get the drug in their body effectively?

And then the use environment—we heard about this this morning. But there's patents that obviously are in this space. There's a long time to wait to be able to develop a generic. And we know that there's work going on to be able to get to the market first. Very competitive landscape. So we know that collaboration is minimal to none, right? We're not friends with our competitors necessarily. We're not publishing data in this space. We don't want to give competitors an advantage.

And then we also know that pre- and post-market use safety data are limited, either flawed, not reported, that type of thing. Real-world evidence is limited. So just kind of the known background information in this space is also limited.
Given that, ANDAs require proof that the generic is bioequivalent in the dosage form, strengths, and such. We know that no preclinical and clinical data are required, so that should give us an advantage, right? That should shorten our process. But we know that's not really happening.

And in my mind, I say, well, then that should open up some budget, right? You're not having to run a clinical study, which we know are very expensive. However, human factors budgets are significantly less than clinical study budgets. So when we get to these comparative use human factors studies, literally getting funding to run them is problematic and might take extra time.

So again, if we think about the use cases or what's going on, ideally we would see this process happening early in the design process. So as we're selecting a device, or as we're thinking about developing a generic. And ultimately we know that the ANDA pathway is not for innovating. We're looking to basically be comparative or equal to the RLD.

So a sponsor for a generic rarely has access to human factors data for the RLD or a right to reference. So we're looking at replicating any data that or creating new that we want to be able to leverage. And ultimately, the regulation says our goal is to minimize or avoid those design differences.

The ANDA does require that we also run a human factors validation study. So it doesn't get us out of that stipulation for the Code of Federal Regulations. And sometimes that could be used as evidence. We do know that that might be something that you could use to justify some of these differences. Again, the URRA is a very important piece of information for the regulatory submission.

And then, if we're looking at kind of what's going on, though we may identify those other design differences—and I know for a fact in these submissions, the industry's goal at the top, we know that we're trying to minimize design differences. Well, industry in our documentation during the comparative threshold analysis, we're trying to minimize or avoid those. So there's problems actually identifying—if the button's a different color, is that considered another design difference? We're trying to kind of minimize the differences, but FDA is not letting us hide them. So a lot of times that's stalling the submission process because there are features that are related to critical tasks that are not even being identified. And then you're not kind of stepping through the process that the agency needs in order to do their review.

And ultimately we know that both experienced and novice human factors professionals, consulting firms that have been doing this a long time, they are having difficulties in categorizing and justifying. I was just on a webinar. I hear a lot of kind of misinformation being put out there right now. The justifications should be data-driven. So I think of the Jerry Maguire movie "show me the money." I say, show me the data, right? We need data behind our opinions so that it's not an opinion-based justification that we're really showing the data. And even though IEC 62366 is a design standard, it's not based on human performance data. So a lot of times we might start with a design that's what's recommended, and then when we put it in the hands of users, it doesn't work. So we really need human factors design or human factors data, and that can come in different forms.

So ultimately, we have challenging methods going on. Sometimes we can't recruit participants. We might have to recruit adolescents, which are very difficult, and when we're getting into large sample sizes, experienced RLD users might be really hard. The RLD may no longer be available on the market for sale. Procurement of the RLD—we hear about supply chain issues. The cost of the RLD can be prohibitive. Maybe it's—there's not a lot on the market. Now we're taking more away from actual patients who need it to run a study.

The statistics—obviously finding a meaningful delta for the non-inferiority statistics. I've worked with statisticians, and I'm like, "Are you going to say 10% of the users don't need to get the product?" I'm not saying that. How are we going to come up with that margin? What's meaningful? What are we really asking here? And then ultimately simulated use testing might be problematic. The RLD has actual drug. How are we going to get around that to do simulated use testing?

Budget constraints in our large sample sizes, the number of days we have to do testing, recruitment and incentives, obtaining, purchasing the RLD. Typically—well, on average, an ANDA validation human factors validation can be $250,000 or less or more. But if you're looking at comparative use studies, we're looking at a million dollars, and a lot of that is the drug procurement costs.

Delays to market—this is a big one. Again, why are we here? We want to have more drug availability for patients in need. FDA-controlled correspondence—I've had clients take years, and it's not for lack of effort, right, to get to that agreement. Recruitment challenges—again, we

can take months to years to get the number of people that we need. And then the data collection and analysis is challenging.

So some of our solutions—this is my research team with Dr. Megan Conrad, and we're on a collaborative grant with FDA. We're in year three. We've had a lot of discussions with them, and that's why I know that they're wanting to get this right. Ultimately, we've published and presented this. We've done dissemination activities to try to get the word out on new ways of thinking around this. So look us up, look for our publications. We have more pending.

And I want to kind of end here where we have gathered and presented and published stakeholder perspectives. So thinking about industry and FDA—we're kind of the system here. We have to come to agreement. We are really proud of a visual taxonomy that we've developed, and we're trying to advance it further to help identify design attributes and user interface features that we should be looking at to kind of standardize that approach and do better at identifying other design differences.

And then, ultimately, we're looking to develop and test a different statistical analysis with comparative use. And importantly, we're bringing in—this is an idea of kind of where we are right now with the process. But we're bringing in the use-related risk analysis. And we're looking at another metric that isn't just counting use errors. But we do bring in kind of that risk analysis and potential harm.

So with that, thank you all for listening, and I look forward to a great discussion.

William Chong: Thank you, Melissa. Next we'd like to invite up Vivek Viswanathan and Daliya Bharati, who will be doing a joint presentation titled "Comparative Threshold Analysis: So Near, Yet So Far."

Daliya Bharati (Advagen Pharma) and Vivek Viswanathan (Rubicon Research Canada: Good afternoon, everyone. Firstly, I would like to thank the agency for giving us this opportunity to express our views at this workshop. Vivek and myself are going to talk about comparative threshold analysis. We have had wonderful presentations from Brandon and Melissa, but we would like to take a step back and talk about comparative threshold analysis, and how especially we feel that the enormous data that is generated during the product development—development of the drug-device combination products involves generation of a lot of data—and how this data can probably be helpful along with the comparative threshold analysis to, and probably can help alleviate the need for doing the human factors studies.

Just a disclaimer before we proceed. The opinions and views expressed here are ours, and not of Rubicon Research or its affiliates.

While I know Brandon discussed about the expectations of the agency and the guidance that discusses about how threshold analysis is to be done, threshold analysis involves these critical aspects like doing the labeling comparisons, doing the physical comparison of the drug product, the drug-device combination products, and doing the comparative task analysis.

When these are done, especially when the labeling comparisons are done, we do a detailed side-by-side, line-by-line comparison of the prescribing information, the instructions for use, the device labels, the carton labels, device constituent part descriptions. So each and every aspect of the label is compared with the RLD, and the differences are put forth and justified.

Similarly, the physical comparisons involve detailed examination of the physical features of the RLD compared the same with the drug delivery constituent part. And details such as the precise photographs and everything is presented in this kind of comparison. Next is also the comparative task analysis, where comparison of the task of the proposed generic drug-device combination product against the RLD is presented, where the detailed sequential use or the tasks that are involved are comparatively assessed against the reference product.

And eventually the outcome from this threshold analysis is either there is no design difference, there are minor design differences which may not require human factors studies or do not require human factors studies, however, there can be certain other design differences, which eventually the expectations of the agency is that we tend to go into the area of human factors studies.

So when these user interface design differences are being compared, eventually it's always that the applicant tries to ensure that these differences are minimal, and that is assessed throughout the development process. And care is taken that these design differences are very minor or remain minimal. However, there are instances where the applicant has no choice but has to get into an area where the design differences no longer remain minor. That's primarily because we have certain patents around the devices which obviously the applicants need to take care of. There could be certain manufacturing issues. There could be certain budget constraints. But eventually, though the attempt is to keep the design differences to the minimum, it gets into the area where we go into the other user interface differences. And this can overall result probably in the need for doing human factors studies, and eventually that can delay the project at times, maybe even killing the project, because that's not the kind of time that maybe an applicant has in order to get the product filed and approved.

Eventually getting a generic drug product on the market is actually all a time-sensitive activity. And therefore, we thought that there are other probable areas and aspects that should be taken into consideration when kind of assessing these other user differences. I think we can get into details about them.

Vivek Viswanathan: Thank you, Daliya. So I would like to piggyback on what Melissa said. Where is the data? So just taking things forward, wanted to tell you guys that we all know that design and development of a generic combination drug development process—it involves multidimensional research. I'm from R&D background. So I've been working in product development area. So I'm sure we all are well aware of—be it a nasal dosage form, an injectable, an inhalation dosage form—there are very specific and tailored CMC guidances which are available for each dosage form. There are certain studies which are put forth by the FDA, which are needed as a part of your ANDA submission. So why can't we leverage some of the data which we generate as a part of development and try and justify these other user differences?

Just to take things ahead, as a part of CMC characterization studies like cleaning, priming, re-priming, robustness studies are done. Now, as a part of cleaning, you have to prove that your device is able to perform equally well even after the cleaning step is done. There are some priming, re-priming requirements of a device which you need to comply to. If you are doing a generic product, of course you have to comply to the priming, re-priming requirements, as mentioned in the guidance.

Then, another interesting point of robustness studies are done wherein we prove that—say there are minor, or there are significant stress conditions which the device may encounter during

its shelf life. In spite of those stresses, your device performance is unhindered. So like that, many other studies are there which we typically perform. In addition, as a part of demonstration of BE, you perform a lot of in vitro as well as in vivo bioequivalence studies. Now in vitro and in vivo studies definitely involve multiple steps, multiple activities which are done. And then suppose the product is meant for emergency use. Then there is a whole lot of reliability study which is done. As a part of reliability study, a lot of in-use studies are performed, and you need to show that your device is really consistent, and there are like the 99.99% study which you need to be doing.

So, talking about the development studies which I briefly touched based upon, I can surely say that all these studies definitely prove that our device or the combination product which we have developed is really robust, and it is comparable in performance to the reference product.

Moving on, there are also some DHF compilation documents. As part of 820.30, we do the hazard analysis. We do the risk analysis, risk mitigation, and I'm sure there are a lot of design validation and design verification studies which are performed. Many times as a part of design history file compilation, we also perform the actual in-use studies which demonstrate the product performance.

So, in a nutshell, what I wanted to say is when we do these development studies and DHF compilation documents, it essentially is an indirect indication of how your device is performing, and eventually in vitro and in vivo tests do demonstrate the device performance.

So whenever our device prototype is finalized, or we are closer to the device finalization, we do perform a threshold analysis. And the sponsor definitely feels that they have moved one step closer in the ANDA approval process. But due to some of the implications which have been laid out, we feel that we are still very far, because there are some other design differences which warrant for CUHF studies.

So by talking about these slides, I just wanted to say that when other design differences are identified, why can't we leverage some of the extensive development data and then verify or check if a CUHF can be waived? If it can be waived using this data, well and good. Otherwise, definitely, the CUHF has to be done. So in the coming few slides, I think that's the interesting part of the presentation. We are going to present some very interesting case studies which would highlight upon what could be the other design differences which typically an ANDA developer would face during their development course, and how we can leverage this existing studies and do the risk mitigation.

Daliya Bharati: So here, we've tried to put forth a case where in the segment of labeling comparison that is done as a part of the comparative threshold analysis—suppose we have a product, a nasal product, that requires more priming shots compared to the reference product. It could be maybe due to a difference in the bottle size, or maybe the dip tube length.

The instructions on the labeling for the reference product say, "Push the bottle with thumb firmly and quickly 5 to 6 times, or until a fine spray appears. Now your pump is primed."

So the instructions for the test product include, "Push bottle with thumb firmly and quickly 6 to 7 times, or until a fine spray appears. Now your pump is primed."

Now this, whether it falls under the other user interface differences, and whether this warrants a human factors studies. Now from the data that is generated during the drug product

development, we do study the equivalence of product performance as a part of the priming, re-priming studies. And this is definitely a part of the ANDA submissions as well.

So if we have clear instructions on the instructions for use as well as maybe the product carton, and have the data that demonstrates drug content from this spray 5, 6, and 7, and a comparison between the test and the reference products, whether that seems adequate to establish substitutability of the test product without actually having to go into the category of doing a human factors studies. Isn't it? This is one example.

The second is of, again, a labeling comparison for a situation where we have an oral liquid product, a test product. It provides lesser number of syringes. It provides lesser number of syringes compared to the reference product. For example, the reference product gives two syringes along with the product, and the test product intends to provide only a single syringe with the oral liquid product.

The instructions on the label for the reference product says, "Use the second syringe for the remaining volume of the medicine to be taken," while the test instructions for use mentions, "Reuse the same, or the first syringe for the remaining volume of the medicine to be taken."

Now, does that kind of a labeling difference warrant any necessary human factors studies, or with the data like comparative dose accuracy studies which are done with both the test and the reference products, is it sufficient to take the filing ahead and get an ANDA approval?

So like we said, for the first case, clear instructions are available on the instructions for use and the product carton. This, supported with the comparative dose accuracy data between the test and the reference product, is it sufficient enough to establish substitutability of the test product against the reference without getting into any CUHF kind of activities, because eventually, that is just going to delay the overall process, maybe getting into controlled correspondences, or maybe getting into a pre-ANDA meeting and just going to delay the overall filing as well as then the approval of the ANDA.

Vivek Viswanathan: Taking ahead, I can just walk you all through a few more case studies with different dosage forms. Now, this is a very typical example. We have an inhaler which has a different mouthpiece design due to IP implications. So, because of the different mouthpiece design, the cleaning step for the test device and the reference device seem to be different. Now, are they really different? If you look at the instructions on the reference product, it says, "Open the mouthpiece by pulling it upwards, and then open the base," whereas due to the design of the test product, the mouthpiece and the base are lifted together.

So what happens is in the real life as well, the type of cleaning or the way in which the device is cleaned is different for the test and reference. So that is why the instructions for use are different. Now, there could be a situation wherein maybe a CUHF is not required in such a case, but this would require extensive to-and-fro communication with the agency calling for more justifications. Instead, what we can propose is, if we do a cleaning study of the T device as per the IFU instructions and do a risk assessment, and we also check the impact of following the R instructions for cleaning the T device, would that not be sufficient, and can that be included as a part of, say, threshold analysis, or a follow-up threshold analysis?

So this is what are just some ways to say that there is no impact on the difference in the cleaning step, and hence there is no risk in the substitutability of T versus R.

Quickly going ahead, this is an emergency use medication. The T device is different in design from the R. The R has a bottom actuation button where the T has a side actuation button. Now, both are essentially of the same color, but the instructions which are given is "press the plunger at the bottom of the device," and the T product says, "press the button at the side of the device."

Considering this is an emergency use medication, we would be doing reliability studies as a part of performance. And we would also be demonstrating bioequivalence in vitro as well as in vivo, and we do the risk management. So would this study be sufficient, because, since this is an emergency use medication, and this is a major difference in the user interface that could call for some obstacles in getting the device approved?

Also, last two case studies, if I can quickly do—this is a device meant for delivering two doses of the medicine. The T product has two actuation buttons, whereas the R has single actuation button. But if you look very closely, the dose indicator for both is essentially giving the same indication. The instructions clearly mentions that after each dose the patient has to look at the dose indicator and check if successful dose delivery is done.

So, since our instructions for use clearly mentions to check the dose delivery, will this design difference call for a human factors studies, or would risk mitigation be sufficient?

And last, but not the least, the last case study is the absence of a safety feature which is present in the R device. But the T device has the same safety feature which is built into the product via packaging. So if you look at it very carefully, the way in which the dose is to be administered is the same, and there are no additional steps which are required for the usage of the T product. So even though the important safety feature is missing from the T product, we need to ask ourselves that, is it really worth considering for additional studies? Or can these studies done on the R device be enough to prove the worthiness of the device?

So in summary, I would just like to conclude by saying that over the years the innovation in device development has evolved a lot leading to a lot of complex, and I would say, novel devices coming into the market. And more often than not, due to the various implications which exists, the sponsors would end up in having an outcome of other design differences. So we propose that there could be certain research done in refining the current FDA threshold analysis guidance that does not provide any detailed classification or any illustrations to other different design differences. Basically what we focus is to justify those other different design differences as much as possible. And if some illustrations can be included in the guidance, then this would ultimately reduce the turnaround time for the generic drug product development, drug combination drug product, at the same time maintaining the substitutability of generic device without any new risk which is administered.

So I would say, to conclude, with the increasing number of complex drug-device combination products being available, some elaborate information provided with the guidance may be really helpful and beneficial to the sponsors. Thank you.

William Chong: Thank you, Daliya and Vivek. So our next presenter is going to be Amy Lukau from Kindeva Drug Delivery, who will be providing a presentation called "Industry Perspective: Development of Generic Emergency Use Products."

Amy Lukau (Kindeva Drug Delivery): All right. Thank you, Dr. Chong. And I'd like to thank the agency for inviting us here today to share our perspective on how we think the current guidance regarding generic human factors for drug products should go.

So like Dr. Chong said, my name is Amy Lukau, and I am the senior human factors lead for Kindeva Drug Delivery.

So first things first, a brief disclaimer. The opinions expressed within the content of this presentation are solely my own, and do not reflect those of my employer or the organizations in which I participate.

So I'm going to speak on three tenets that affect emergency use generic drug product development. The first being giving a general overview—what are generic emergency use drug products? The second being context of use—why are they important? And lastly, overcoming challenges of siloed experiences. Well, I represent a new way of looking at emergency use RLD versus the generic drug product.

So first things first, the evolution of human factors guidance for combination drug products has been significant, especially including regulatory guidance documents such as the IEC, the FDA, and the USP. Initially guidance documents primarily focused on safety and efficacy of the product. However, over time they expanded to include overall context of use, drug interaction, in addition to providing safety and effectiveness for use.

So I always like to show this slide—human factors is a critical component of the risk management system. So what we see are three tenants of medical device development used for device user interactions: the use environment, user, and device user interface, eliminating and/or reducing design-related problems that contribute to or may cause unsafe and ineffective use are a critical component of the risk management system.

Emergency use auto-injectors such as those for treating anaphylaxis or inhalers for treating acute asthma attacks are often used by patients, caregivers, or first responders. Because a generic—because the generic emergency auto-injector requires that the user get it right on the first try, the FDA recommends that the emergency use auto-injector include design controlled specifications for successful injection reliability of 99.99% with a 95% level of confidence.

The level of reliability necessary to manufacture safe and effective combination products directly correlates the level of risk with an unreliable emergency use auto-injector. This reliability specification represents the probability that the emergency use auto-injector will perform as intended without failure, given time interval under specified conditions.

So context of use is critical when we discuss the development of emergency use auto-injectors with respect to its reference listed drug. First, we'll assess, identify differences, then they may take into consideration threshold analysis, reliability, the context in which a generic emergency use product is used, and lastly, the use-related risk analysis.

Context of use—so context of use is critical when assessing generic emergency use products against its RLD. The FDA has given guidance in its 2017 paper, specifying what a minor and other design differences. And the reason these errors are interfacing is sometimes we may present to the FDA a generic drug where we think we've made a minor design difference. However, the FDA may come back to us and say, "No, we think this is another design difference. You need to go back and minimize any design differences on the user interface."

So context of use as related with threshold analysis. But one of the questions I like to answer and pose to the audience is: Is our threshold analysis relevant/necessary for emergency use generic drug products?

So like my predecessors have stated, very simple in terms of the threshold analysis, you're going to do a labeling side-by-side, comparative task analysis, and lastly, a physical comparison of delivery devices.

So context of use for reliability is critical for emergency use auto-injectors. Device reliability is a critical component of emergency use products. We do this by employing essential performance requirements. ISO 11608 ensures that a device reliably performs the minimum activation and operation steps to meet safety requirements and to be reliably operated by different users. These often include cap removal, activation force, injection time, dose accuracy, needle length, and lastly, confirmation of the operation by either visual, audible, and tactile means.

A reliability assessment is specific to the combination product's intended use. Risk are likely to be impacted by the conditions being treated, environments for use, emergency use auto-injector technology, drug user characteristics, to name a few.

Reliability analysis, in addition to the traditional development activities, should include identification of the reliability requirements and specifications, risk analysis, design verification and validation of the reliability requirements and specifications, design transfer of reliability specifications to the correct production specifications. The design of an emergency auto-injector should consider various factors to ensure its reliability, such as intended use, associated risk, and user-related risk issues and use conditions.

The use-related risk analysis defines all use cases and subtasks related to the user interface, potential use errors and potential harm to the patient that result from potential use errors and the severity of that harm. A URRA describes which aspects of the user interface and our product design is intended to minimize the identified risks, also known as risk control measures.

The URRA is a powerful tool because it comprehensively considers the overall human factors engineering product development process.

Complexity of combination drug products arises from a diversity of users, environments, delivery devices and the combination products of the drug and the overall user interface, as well as the potential effects of the medication errors. For emergency use generic drug products, the use-related risk analysis comprehensively examines the drug device and combination of the two to contribute to risks and identify opportunities to design the product and to mitigate those risks.

And this is important because all critical—all device tasks within generic emergency use products are critical, whereas for lay devices, not all tasks are critical. So how we effectively design and evaluate minor versus major design differences and overall safe, effective, and context of use as it relates also to essential performance requirements.

Utilizing the use-related risk analysis for a comprehensive consideration of overall safety and reliability of the generic product is essential because tasks are completed by different user groups and are identified in the use-related risk analysis across potential product users and use environments. They also take into consideration the essential performance requirements of the generic product.

Risk control measures identified in the use-related risk analysis for the RLD against the generic drug could provide an alternate means of determination of minor versus other design differences. This method provides a nuanced and more meaningful assessment of emergency generic drug products way of meeting requirements that may be limited by the current threshold analysis in the current guidance.

So that's all I had. Thank you, guys.

William Chong: Thank you, Amy. And then I'd like to invite our last speaker, Heidi Mehrzad, CEO and Human Factors Expert from HFUX Research, for "It's Hip to be Square: Demonstrating Equivalency without Inferiority in CUHF Studies."

Heidi Mehrzad (HFUX Research): Tough crowd. No chuckles. Okay, that's fine. We'll move on.

So hi, everyone! I'm going to start off by giving a little backstory on myself. Hi again. My name's Heidi. I'm autistic, and I'm ADHD, and just to clarify, I'm not sick. I do not need to be cured. It is a neurotype. But I do come with a little bit of a bunch of differences.

So I would like to ask you for your understanding that not all people are neurotypical. Okay, so we don't all act, speak, behave, gesture, inflect like neurotypical people. And because of that, and not behaving in a neurotypical way, I do like to think very much out of the box.

So what neurotypical people consider out-of-the-box thinking for me is an average Tuesday afternoon. And so with that being said, you might have to give it a few seconds or a few minutes more to see where I'm going with a point. And I assure you, just because I am neurodivergent doesn't mean I can't get to the point. I'm just going to take a little longer or take a different route.

And so during the presentation, if you have the urge, and you're begging me to move on and speed up or slow down, just remember that an autistic person on average processes 42% more information of their surroundings. So let's think about that while you're driving a sedan, I'm driving race car. And I already did 42 laps before you even start the car.

So I'm going and going and going. And because of that, environments like these are easily distractible for me. Once I lose my concentration, I'm out, I glitch. And so I would like to kindly ask you to keep your noises from computers, cell phones, email notifications, text messages, anything like that to a minimum. Actually, if you could avoid them at all, that would be great. And I just want to ask you to think of it as a little bit of an experience, because, while we have to conform to neurotypical ways, 24/7, our entire lives to be accepted and not felt awkward and not perceived as different and disturbing, I'm just asking you to do that for 20 minutes.

All right. So with that being said, I'd like to start.

So I'm going to talk about the basics of human factors and then what the problem with ANDAs is, and the elephant in the room, the non-inferiority model, and then also talk about equivalence in drug efficacy versus safe and effective use. And of course, it's so much more hip to be square than to be inferior, and then also talk about at the end the chi-square trifecta, and that innovation and real world needs to be applied to this thinking.

So first off, I'd like to thank everyone for their presentations, and I'm just going to go ahead and say I'm not going to reiterate some of the points in them. And so I'm not going to touch up on the

guidance, the Hatch-Waxman Act, all the different basics. So I hope that you can just bear that in mind, and just follow the presentation as is.

So to start off, why is human factors engineering important to medical devices? So taken from the FDA website, for medical devices, the most important goal of the human factors usability engineering process is to minimize use-related risks, hazards and risks and confirm them, and then confirm that these efforts were successful, and users can use the device safely and effectively.

So remember, safe and effective is going to come up a lot now.

So what is human factors? Well, first things first, human factors is the science we apply to research and design and engineer a product with respect to the end user. Usability is the goal and the measure used when evaluating these efforts with respect to the end user.

So how do we measure usability? Well, typically, we measure usability with the five E's with the goal specifically in med device development to ensure it's safe and effective, focusing on effective, focusing on error tolerance, with a goal that the product is safe and effective to use.

Okay? So in summary, what is human factors? Human factors is the science we applied to research and design and engineer a product with respect to the end user. What is human factors in drug-device combination product development? Human factors is a risk-based approach. I think that gets often forgotten a little bit. And what are its goals? It's to minimize use-related hazards and risks related to the use of the product and mitigate such through the design of the device UI user interface.

So how do we measure it for the usability evaluation of the UI of the device with focus on its effectiveness, efficiency and user engagement, its error tolerance, and its learnability. And why do we do it? Well, to ensure that users can safely and effectively use a user interface design.

Now let's take a deeper dive into the current ANDA guidance, its approach, its non-inferiority model, and its hurdles as well as the challenges it poses for industry.

All right. So Brandon touched on this. I'm not going to repeat all of this again, but as he already walked us through the Hatch-Waxman amendments, I don't need to rehash it, but let's just remind ourselves of some of the other guidance highlights.

The one thing I'd like to touch on is that drug products that are approved in ANDAs are generally considered by FDA to be therapeutically equivalent to their RLD products, classified as therapeutically equivalent, can be substituted with the full expectation that the generic product will produce the same clinical effect and safety profile as the RLD under the conditions specified in the labeling. We've heard that a bunch of times now. Why is that important? Why am I bringing that up again?

Well, this does not, however, mean that the proposed generic combination product and its RLD need to be identical in all respects. FDA recognizes that an identical design may not always be feasible.

And I think one key point to make here is that it's not required to be the same as RLD, and it can differ from it, and FDA recognizes that. But while FDA recognizes the possibility of the design difference, industry, however, needs guidance to support the fact that they almost are inevitable.

So now let's look at industry challenges specifically with the CUHF. Of course, this has been brought up a bunch of times. So I don't want to bore you. But let's just touch on a couple of facts.

There's a lack of public data in the absence of available market data to support estimate of true error rates for RLD. This requires us to establish and then setting the NI margin arbitrarily. Well, this actually requires us to establish an error rate on our own for a product we did not develop. This can be done by doing comparative use HF testing. However, this will still lack the ability to set the NI margin, because this now must be set arbitrarily due to the fact that we do not have standardized HF margin rates we apply to each product safe and effective.
So product individuality has to be taken into account, considering critical tasks related to safety and risks like dosing, disease state, etc. The setting an arbitrary margin, like the delta, can be 10%, is not applicable to optimal HF practice. I think often we forget humans are not machines, and humans are not drugs. They do not have a guaranteed XYZ percentage they perform at at all times.

There's also the question of large sample sizes and the lack of resources, expertise and drug efficacy statistics. And then there's my favorite points, the rigidity and study design, and that we're actually losing sight of true HF risk-based approach where the CUHF human factors performance versus efficacy mindset has led to the losing sight of true HF and its risk-based approach. Because where does root cause analysis fall into this? Residual risk analysis? And why are we applying a statistical model that is derived as a standard model for clinical testing? More specifically, pharmacology to a question meaning to be solved with human factors testing.

So for that, I would like to discuss how drug efficacy does not equal safe and effective use a little further.

So just to remind ourselves, FDA does not consider the comparative use human factors studies to be clinical investigations, rather intended to confirm that the differences in device and labeling are acceptable, and that the proposed generic can be substituted with a full expectation to produce the same clinical effect and safety profile as the RLD under the condition specified in the labeling. Okay?

So we know human factors testing here comparative is not considered clinical testing. Instead, it's intended to be used to confirm that the difference in the design are acceptable, so that if the user's given the generic or vice versa, the RLD—but we'll get to that later—they can use the product safely and effectively with respect to the use and all its human factors, with the same safety profile as the other product, which would mean administering the product just as safely and effectively as the other product in order to enable the same conditions in which it can yield the same clinical effect.

That is a big difference to what we're currently saying, and that is that we are using human factors to ensure the same clinical effect or prove the same clinical effect.

And to elaborate on that, I would like to take a closer look at the NI model itself.

So the goal of comparative human factors study within NI design intended to support the approval of a generic combination product is to demonstrate that for each critical task, impacted by a change in critical external design attribute, the error rate of generic or the RLD is no greater than the error rate of the RLD plus delta, where delta is some acceptable deviance above.

Now let's also not forget, in contrast, a comparative use human factors study with an NI design is intended to help confirm one aspect of the substitutability of a proposed generic combination product for its RLD, and that is its usage from a human factors point of view, and not for determining differences relative to a drug or placebo.

That means we need to distinguish drug efficacy from safe and effective use when you ask the question of is a drug as effective as the other or a placebo. You have to consider that the NI model is addressing that. Okay, so let me just be very clear, because the NI model looks at the error rates and comes from commonly statistics in clinical investigation. You can rely on the NI model, but the NI model focuses on use errors. It does not focus on the actual behaviors that are behind safe and effective use.

So what's the behavior associated to safe and effective use? Well, you have to look at the actual root causes of these use errors. So study design and statistical model should be identified and selected, based on the question you're trying to answer. And we should focus more on the study design demonstrating equivalence in safe and effective use and then select a model most fitting to that, instead of using an NI model that focuses on a pass/fail mindset of tasks, counting just the use errors.

So the NI model does not look at the behavior associated with safe and effective use. Instead, it limits its focus on use errors relating to the design differences only, and that's not a realistic HF, usability measure or performance standard.

So what I'm trying to explain with that and to illustrate that a little bit more, I'm going to use the Simpson's paradox because we need to understand that right now, we are conflating drug efficacy between RLD and generic with safe and effective use of a product between two products, yielding the same conditions that allow the drug to produce the same clinical effect. And to illustrate that, that's why I'm pulling up the Simpson's paradox, because the Simpson's paradox is a phenomenon in probability and statistics in which a trend appears in several groups of data, but disappears, or even reverses when the groups are combined.

This result is often encountered in social science and medical science statistics, and is particularly problematic when frequency data are unduly given causal interpretations.

So, as you can see, there's the positive trends on both for two separate groups, whereas a negative trend appears when these groups are combined to give a little bit of a more resembling visual. The Simpson's paradox has been used to illustrate the kind of misleading results that the misuse of statistics can generate. You see, when you see this negative trend, and you see all the data points, you actually aren't seeing that every group has a positive trend. You're only seeing the overall negative trend.

And you see, when you use use errors, use error rates of critical tasks and compare them, you are losing individual root causes and other factors, such as close calls, operational difficulties or assistant requests, and only focusing on the actual pass/fail rate, ergo error rate. You actually can have this phenomenon. Where, for example, here's the error rate for a task A.

And now it looks like it's showing a negative trend, meaning a less than desirable performance of the pass/fail by the users. However, if we looked at each user individually, we see that they show that each is showing a positive trend in the task. Now, imagine this is your overall use

error rate showing a negative trend only to discover that the data points underneath show a positive trend.

Okay? So now, we could just pick that up that everybody always likes to bring up. Why not just conduct a human factors validation study?

So let's just toss all this ANDA guidance out. And let's just do a validation study. Well, while it is hard to disagree with that point, I would also like to remind you that HF validation testing would only focus on the overall safety and effectiveness of the products tested, and it would not allow you to make any type of statements with respect to substitutability, or even comparison between RLD and generic. And while that is okay, if you consider that you have shown it to be safe and effective either way, this prevents us from being able to substitute RLDs with generics, and vice versa, which is actually a more likely real-world scenario that the current guidance in its state right now does not address.

So FDA would generally accept a proposed generic combination product that had the same error rates as the RLD as demonstrated by an adequately designed comparative use human factors study, or studies. Now, what should we do to address this conundrum of wanting an ANDA pathway filing, preferably using human factors, comparative testing to ensure safety profiles show equivalent, and hence allow substitutability between the RLD and the generics between the two products without affecting clinical effect negatively? Well, we need to begin.

We need to begin with rethinking the study design of CUHF studies, and with that select a better statistical model that focuses on human factors, social sciences and its measures instead of drug efficacy.

And for that I would like to discuss the chi-square model because the chi-square allows you to look at each individual critical task that is associated with safe and effectiveness, and if you find that one does not show equivalence and/or superiority meaning, they are equivalent, you can look beyond and perform post-hoc tests to identify root causes. The NI model requires you to use a use error ratio only of all tasks which then actually dilutes your results by giving you only an overall picture without identifying the actual behaviors or root causes that led to its results. So a chi-square test is a statistical hypothesis test in the analysis of contingency tables. The test is primarily used to examine whether two categorical variables are independent in influencing the test statistic.

So why use the chi-square test? Because, unlike the NI model, the chi-square uses actual observed data, not arbitrary NI margins from estimated error rates, meaning the chi-square test is used to determine whether there is a statistically significant difference between the expected frequencies and the observed frequencies in one or more categories of a contingency table.

Not only would the chi-square address most of industry challenges, such as lack of public data, setting the NI margin, lack of resources and expertise in biostatistics, it would also allow more flexibility in study design to refocus on true HF risk-based approach, and as a bonus, it could reduce requirements of large sample sizes, and with that reduced challenges in recruitment and the need for surrogacy, especially when thinking of more rare and orphan diseases.

So how do we start with this? Well, first, we would design the study to optimally be supported by the chi-square, and actually focus on the behaviors and the use problems overall of each task that is affected by design differences experienced by users. Additionally, with this rethinking of the comparative human factors study design, we can now implement an additional aspect, and

that is to ensure that substitutability isn't just confirmed from RLD to generic, but also from generic to RLD.

So you would typically set up this study design with your four user groups focusing on the RLD experienced who is exposed to the RLD, the RLD experience exposed to the generic, the RLD and generic naive exposed to the RLD, and the RLD and generic naive exposed to the generic.

And then you could apply a simple sample size or distinct user group, as we're already doing. Fifteen, for example. You have all participants perform three doses and using the chi-square test to see if there's a relationship between those two categorical variables, use errors, and, for example, number of attempts of critical tasks. And if no significance is found, product can be said to show no differences in use problems. And if it is significant, it can be said that the product can be said to show differences in use problems. So the chi-square allows for more flexibility in study design, and greater diversity and subsequent data analyses in correlating statistics.

And just to comfort your variables would still be the same. You're still using the same data points that you're currently using in HF validation testing. They're just performing the statistics a little bit different.

And to give you an example, you would conduct the test where you test experienced RLD users with RLD and generic, and then you use users, indicated, but not received, either generic or RLD. You randomize those, of course, and you have them do multiple treatments. For example, we're talking often injection devices. So let's say, perform three injections. You use the chi-square test to see if there's a relationship between the groups in observed versus expected for things such as use errors, number of attempts, close calls, operational difficulties.

And then you evaluate whether it's significant or not. If not significant, no differences in use errors between categories. You can move on. And if it is significant, you can say that there's differences in use errors between categories. And now you could delve deeper.

Now, one thing but I do have to bring up that is a weakness of the chi-square is that if your cell value is below five, what now? Because in that situation the chi-square is not optimal. Well, that's the beauty of stats. It's got the solution for everything. And so you would use the Fisher's exact test. Now I have to say, I want to remark one big thing for me, the chi-square you can do manually in your head. You don't need a calculator. You don't need software. You don't need a statistician. You just need somebody to be able to follow the basic math here.

I will say the Fisher exact test, you will need software, and that is not a manually computed in your head kind of model. It is only for the case where your cell value is below five.

And now that we understand what human factors is and how we can't apply NI models to human factors testing like we can in clinical testing, and we understand how the chi-square would serve as much better in confirming RLD and generic equivalents and substitutability with respect to design differences, I would like to address a final point I'd like to make. That was actually mentioned more than once during our March discussions.

And for everybody who wasn't there, this kind of focuses on the fact that they don't have to be identical, right, the product. And so the FDA may accept such design differences if they are adequately analyzed, scientifically justified, and do not preclude approval in an ANDA.

And so, while everybody says, and the filings do not allow for innovation, I say, why on earth, not? If you expand CUHF to include all tasks which already are being performed anyway, one could go as far and state that you could potentially compare the RLD and generic safe and effective use aspects on all tasks and analyze, use errors and use problems overall, and the differences of each to enable robust statements of RLD and generic equivalence and substitutability from a usability performance aspect. And that is what human factors is supposed to address.

So if we move away from the current CUHF study design and corresponding statistical model that is rooted in the principles of drug efficacy evaluations, and instead focus on comparing actual usability performance with respect to safe and effective use, we could allow for outdated RLD UIs to be reimagined in generics and mitigate RLD existing known use problems, adapt generic UI designs to today's technology standards, address RLD existing design issues, not accounting for neurodivergent and disabled end users, and address RLD existing design issues, not accounting for real-life use scenarios in today's healthcare system and its health insurance and prescription issuing and filling challenges.

Because one another scene that came up in March, and for everybody who wasn't there was, also, let's not forget the patient impact statements where we've learned that in the real world, the change from RLD to generic isn't often just that change. In fact, one patient stated that they didn't even start with the RLD. They started with the generic and then were switched to the RLD because of insurance coverage issues. So how do we account for that?

Well, again, that's why I like the chi-square so much because this model allows you the interchangeability of use errors and use problems and the differences by tasks between group categories, meaning if you set it up as a four-group study design and you focus on overall use problems and not just use error frequencies, and with that ratios, and you actually look at individual differences from the severity and criticality level, from the risks and hazards associated with those tasks, you can do RLD versus generic. So you could be making the statement for patients and users using RLD first and switching to generic, and then you could make the reverse statement as the other way around. Meaning this model allows statements safe and effective use to account for patients and users, not only switching from RLD to generic, but also from generic to RLD.

And this would finally allow us to get back to the true essence of the human factors, risk-based approach where we look at actual root causes of use problems and factored those in when evaluating a product from a human factors standpoint of safe and effective use conditions, focusing also on—I'm sure you're all going to go home with the trauma of having heard that sentence too many times—the root cause analysis. So we would focus on the root cause analysis of all use problems encountered, and with that move away from a model that only considers a pass/fail mindset, meaning, only identifying, evaluating, and then analyzing, use problems or use errors leading to failures towards a model that focuses on all use problems encountered during all critical tasks to ensure a comprehensive comparison of actual usability performance between RLD and generic to enable a robust statement of substitutability from an HF safe and effective criteria perspective.

The end. Just kidding, to be continued. And that was it. Thank you.

William Chong: Thank you, Heidi, and a special thank you to all of our speakers. Those are really informative presentations provided a lot of things for us to think about. Our schedule has us taking a short coffee break now. We'd like everybody to be back around 3:35, where we'll

start our panel discussion. And as a reminder to our panelists, we're intending to start with questions to the speakers that anyone has based on what we've heard today. And then we also want to hear from everybody on this, on the priorities that you identified from what was discussed today. So thank you again

Session 4: Drug-Device Combination Products – Part 2

Panel Discussion

William Chong: Welcome back, everybody, from your coffee break. Hopefully you were able to stretch your legs a little bit. We're going to pivot into our panel discussion for this session. As I mentioned, we're hopeful that we might have our public commenter online. Do we know if he was able to join?
Yes, very good. So Dr. Feldman is here as well. If anyone has questions for Dr. Feldman, I'd like to start with any questions from the rest of our panelists to our speakers, to our public commenter on anything they heard today—ideas that they'd like to learn a little bit more about, or questions about anything that was said. So we'll start there.

Don't everybody put your hands up at once. And if we don't have any questions, we'll move to some questions that I have.

Panel member: But I have a question—yeah, just not entirely relevant, but for Heidi. I guess I missed the implications of citing Simpson's paradox in this area. And then the other thing I would mention to Jason over break is that Simpson himself was a civil servant in England, so you know, I'm proud of him.

Heidi Mehrzad: I have no relations to him.

Panel member: No, but you described the paradox, but I'm wondering why,

Heidi Mehrzad: because when we use something like an NI margin, we're using arbitrary measures, and we could be conflating the actual root cause of a use problem by just counting use error frequencies to set the NI margin as opposed to actually looking at use problems overall—meaning close calls, meaning operational difficulty, meaning severity of critical tasks, meaning the actual hazard and risk associated with it—and then look at the root cause, like kind of how Brandon was showing that there could be many reasons why you have a use error. They're not all related to the actual design difference. So if we just count all use errors under the assumption that they are, then we're conflating the actual purpose of it.

Panel member: So it's more the qualitative aspects of a paradox than the specific groups

Heidi Mehrzad: yeah, it was merely an example to visualize what can happen when you misinterpret data.

William Chong: I think, Markham.

Markham Luke (FDA): Hi, I just wanted to thank all the presenters for wonderful presentations. They were very clear and approachable. I also want to especially thank our colleagues who have participated in our GDUFA research enterprise, and you can see some of the results from Melissa's presentation that you worked on together with Megan on some of the work that was done from using the user fee funding to assist with the thinking. And also, some of that thinking then leads to some collaborative interaction and networking. You can see industry starting to think a little bit more about this as well. So, Brandon, thank you for your presentation as well. I think how we put these together and move forward—this is, as was pointed out, a work in progress that will continue to think about this space. But I think one thing is to start putting some of this into action. And so how do you think we should do that, Brandon? Do you have any thoughts? Are you going to send that in with the application proposal, maybe, or something like that? So just some thoughts about how you would implement some of the discussion points that you raised. Appreciate it.

Brandon Wood: Yeah, no, absolutely. Thank you for the acknowledgment, Markham, and for the question. So I think, you know, there are a couple of different considerations that need to be taken here. And just being a regulatory professional, I think about how, if I were to get an alternative consideration, how I'd actually go about this, right? And then the other is the practical implementation. So I'm lucky to say that I have a very talented engineering and human factors group that I work with. So in theory, I think, you know, their position is that really, at the end of the day, we need to adjust our expectations for sameness based on the tools that we have available to us to evaluate sameness.

And when you're talking about human factors sciences, it's largely a subjective and psychological science, right? So in that sense, more qualitative considerations along with root cause analysis and benefit-risk analysis may be more suited to support such proposals.

So I know we are already running certain case studies and have run certain case studies, as were presented in my presentation, employing this alternative IEC 62366 evolution design, where the main changes—it's basic human factors principles. But you're changing your specification to align specifically with RLD device users, and then also scaling the sample size appropriately and executing to demonstrate successful use.

Practically, that's how we would move forward. But then, from a regulatory perspective—and this is one of the things that I think Vivek had touched on in his presentation—you need to have alignment on the other differences first. You don't want to get too far down the line and then realize that you have another difference that is identified that you didn't necessarily identify, because then your whole study downstream is kind of null and void, and you're back to the beginning. So my regulatory recommendations back to my engineering and human factors groups would be to get that buy-in up front on those other differences and then kind of move forward into discussions via product development meetings, controlled correspondence, etc., putting that more traditional human factors-based study design in front of the agency with specific questions and then move forward to execution. So thank you.

Markham Luke: Thank you, Brandon. And also I want to point out that this is the interface between engineering and drug medicine, so the whole enterprise that we're embarking on is looking at how we can regulate these device engineering components that help deliver drugs more effectively and efficiently. And notice he brought in IEC and ISO. So standards are the

lifeblood of device regulation, and so we wanted to make sure that we try to introduce those so that we don't reinvent the wheel as much as possible. So thank you.

Panel Member: Can I—sorry? Go ahead. Thanks, Melissa. All right, thanks. Sorry, Melissa. So I actually had two questions. I'm going to ask Brandon one, and then I'll turn it over to Melissa.

So you mentioned scaling sample size to severity, and I'm wondering how you—and then, how are you thinking about that?

Brandon Wood: Yeah, absolutely. So, you know, a lot of basic human factors principles go back to Faulkner, right? And everybody talks about 15, which I think, you know, what we've acknowledged is there are diminishing returns when we are increasing sample size. You know, for us in the alternative study design, we're ultimately looking at, you know, can we—what sample size is appropriate to detect new use errors that can be attributed to negative transfer? And I don't think it's so much that the increase in sample size would increase the probability of detection of new use errors, but would give further confidence in the study that was designed according to the product risk. So it's kind of a give and take there.

Panel Member: Okay, yeah, thanks. So yeah, Faulkner had a lot to do with likelihood of the error occurring as well, and so I would have to probably think through—I'd like to talk more about that. But think through that, Melissa. Thanks.

Melissa Lemke: So I just want to kind of piggyback on Brandon. So when I come at human factors, I actually look at our science as objective and how we assess our user interfaces with task performance. So while we do include subjective data from the users that feeds into our root cause analysis, it's still objective. They complete the task correctly, and that leads to correct administration or not. It is really black and white in my mind.

That being said, there are nuances, and we do get into a qualitative approach, even with human factors validation. But a lot of times, I hear, you know, residual risk arguments—it's very subjective. And I haven't seen that be successful with submissions. You really have to get down to: is the user interface optimized for human factors validation? In this space, we're doing that kind of extra step of asking, is it considered the same as, or just as safe as—it could be safer? But then, kind of where a nuance comes in: is this unique feature introducing potential harm? New use-related risks?

And that question typically is answered by yes, because if you do human factors testing—home health care and even clinical environments now, I mean, they're very complex. It's a very high-risk system. So I think that's where every difference is scrutinized, and it's very hard to say that it's not going to introduce new risk. And I'm proud to say that in our grant, we're really looking at a new measure where we can look at users' close calls, difficulties, and then looking at the study minimum—kind of reducing sample size. And Megan can certainly jump in here, too. But we're looking at still doing statistics, because if you have four users and four difficulties or close calls on one product, and ten difficulties or close calls on another, how can we say that one is safer than the other or just as safe? And getting into those nuances is really where this space is really challenging.

William Chong: Thanks. I had one question—just I know we spent a lot of time talking today about comparative human factors studies. But I feel like at some point in the conversation today, we also heard a little bit that it's still challenging to do the comparative analyses and like

designating the differences. I was curious: is that an area that any of our public panelists see as a continued area of need for, you know, more research into how to do it, more standardized, provide clarity on what is a other minor—no difference? I mean, hopefully, we can identify no difference, but differencing between minor and others—is that an area that there's still need for research?

Melissa Lemke: I can answer again, just kind of seeing what goes on in industry and hearing what's taught in our survey. People say that they get it, and it's straightforward, and when you read the draft guidance, it seems straightforward. But kind of was on that slide that I presented: the goal is not to have design differences. So I think a lot of times, if you don't have human factors expertise especially, or you go to design standards—AG 75—how much button force is different? Well, that might be a measure or a factor depending on your user group. How much injection time—holding the injection—is a millisecond going to matter? Typically, we may not know until we test.

And so that's where getting at, you know, some of the bench studies aren't necessarily going to give us that user data, and especially looking at RLD compared to the new product, and assuming, you know, they may switch back and forth or get prescribed or dispensed back and forth, and thinking about the end users who don't have a lot of context of being a healthcare provider—if they're a patient, those types of things. A lot of times, we need data. So, you know, we have heard from the agency, and you can look at doing other studies other than comparative use human factors. And I think that's an area that could use research to look at what types of studies other than these comparative use studies might get us at the same answer.

Panel Member: I'd add on—we've been looking a little bit about ways to identify minor and other design differences through categorizing design features using the taxonomy. But also, I think, where we're arriving is that to be considered a minor design difference, you likely need to meet a lot of different criteria—like maybe not being associated with a critical task, not changing the way that the user interacts with the device, not being maybe for emergency use. So there's likely, you know, a checklist of things that you need to meet to either fall into the category of minor or other.

Brandon Wood: Yeah, the only additional point I would make there—you know, I do think additional research, and at least making those ground rules available, would be useful because we've tried to make it, you know, at least within Teva, we tried to make it as quantitative as possible with the incorporation of these related risk assessments and hazard analyses, and trying to draw some quantitative correlations to what tasks actually are presenting the most risk. But, you know, we've still been surprised with certain classifications of changes after submission, and that's why, you know, again, as a regulatory professional, I'm so keen to make sure that we put this in front of the agency first and foremost, to get buy-in up front before we start more of our downstream activities. Because as much as we've tried to make it a quantitative exercise, there's still some level of subjectivity. So if additional, you know, research or publication of those ground rules could be made available, it would greatly help us out, at least earlier on. Even though we'll still work for that confirmation, it would just kind of skip us in front of the line a little bit. So...

William Chong: All right, not seeing any additional questions coming from the table. Maybe we'll follow last session's model on the panel discussion and take a moment to touch base with everybody on the panel. I'd like to start with the public panelists who didn't get a chance to

speak yet, to see what they heard today, if it aligns with what their experiences are, their priorities are, and really, to hear from each panel member sort of what is the area that you think we should be focusing our research efforts on to support generic combination products.

So I will start with Tim, because he's closest to me. So Tim, I'll let you have first word.

Tim Briggs (Viatris): Thanks, Bill. Thanks to the agency for the opportunity to participate. Compliments to all the speakers, and yeah, I think they did a great job of framing the challenges that we're all experiencing.

Certainly be interested in research into alternative designs—alternative risk-based approaches which consider the cause of observed errors, whether that's an evolution of the HF validation study design, such as Brandon spoke to, or an alternative statistical approach, like ideas presented.

Considering the current NI study design, I'd be very interested in research if we could consider standardizing the approach of NI margin selection, whether that's dependent on therapy type. So we've had discussions over whether there are specific NI margins, considering the therapy type—whether it's chronic, acute, or emergency use.

And any data that can help us to inform those inputs into the statistical analysis plan. So another aspect to that is calculation of the use error and the delta—trying to obtain that published data can be tricky, so potentially research into what data is available and making that available to everybody.

Considerations in a comparative use study designed for use scenarios with opportunity for learning. So how can that be introduced to a study type without risk of potentially introducing study artifact?

And then also looking at the recruitment challenges, and whether there are opportunities for when we need to recruit these large numbers, whether we can—surrogate users or representative users could be given a level of training on the reference product and then subsequently evaluated on the test product, and how effective that could be to meet that challenge.

William Chong: Great! Thank you, Tim. I'd like to go to Manoj next.

Manoj Pananchukunnath (Biocon): Good, thanks, Bill. So good, and thanks to all the speakers for very interesting presentations. It was opening new lines of thought in terms of approaches.

Just with just one public comment, I think one concern is that, you know, is there a level of confusion out there which prevents people from putting more comments in? Because we had 30 comments in the last session, 23 in the last session, and just one here. So it's probably telling me the confusion we see out in the circles in the industry is not allowing people to come to a coherent thought process about what are really the issues we can distill and try and put it up for consideration with the agency.

So I'm going back about 15 years. I'm sure you are aware most generic manufacturers don't manufacture the devices. They get them designed by shops outside and various types of design shops there. Being with the largest generic companies in the world, we worked with the largest device manufacturers 15 years back, and what we realized at huge expense was they were

overlaying a brand development process onto the ANDA development process. And the biggest hit was that by the time we could be ready for threshold analysis, we already spent a lot of money to develop the API, get the formulation done, and, you know, kind of fill-finish it, also start putting up devices on stability. And then all of a sudden, we realize that there is more stuff to be evaluated and potentially, and eventually it led to some programs being shelved—a costly learning for all of us in those phases.

And I think over the years we have evolved to a very large degree in terms of how we've addressed these problems with both the agency and our internal development processes. Teams have been built up within pharmaceutical companies for design verification or human factors studies and regulatory purposes.

So we reach a stage where we're putting in applications, and we see questions back from the agency. And now we're thinking about how can we better this. And some of the things which come to my mind—I think Brandon touched about it a little bit—there is an evolution in the RLD, and there are RLDs in the market today who moved the devices from what it was five years back down the line to at least three or four different design devices. And that's giving us an opportunity to understand that those devices, I presume, have been approved with a certain level of understanding on studies being performed with them. And can we look at it from that fashion? You know, the agency has a better ability to understand that we can only point out the marketed differences.

And we had an example recently where we were talking about—it's a simple thing—the dose extension knob in a pen. Now, steadily the thought process has been to move the extension of the dose knob to as little as possible. And if we have a pen which, say, it extends up to 14 and the RLD is at 10, now there could be a discussion on whether it's a minor change, or it's another change, or what it is. But I can show you five other products in the market which have those extensions up to 30 mm. So the concern that somebody would not have the span with the thumb to exercise the dosing is already demonstrated, and that goes back to what Professor Aaron was talking about—that, you know, real-life evidence is maybe not in the manner he described, which is quite a long way to go in terms of collecting that all information. But I'm sure we have market complaints. We see it every day in our products. It comes back to us about various issues in administering the product. Is there any indication or a signal of those kinds of issues is something we can think about as approaches, and, you know, try and bring them to the table to address it?

The other aspect which I urge everybody to think carefully about is access. And we've reached a point where we know as a generic company, the day we launch the first wave of the product, we're already working on cost reductions. And I've seen it in the last 30 years I've been in the industry that it used to be API first, where mechanisms were not very clear about how can you switch an API supplier? Now it's pretty standard. We are pretty clear on data requirements, and we actually plan for these things well in advance.

Manufacturing sites—how do we switch a manufacturing site? This has been an evolution over the years. It may be routine today, but there was a time when, you know, we used to talk about how we could do these things because we have a problem with the site, either capacity or quality issues. How do we switch sites?

So what I'm coming to is definitely in generic devices, we're going to be seriously talking about switching a device either due to supply constraints or the current pricing. Because what's priced

today—we have contracts with suppliers for components and they're volume commitments. But if the prices need to go down for access, then obviously we need to have a mechanism to see: do we go and do the full ANDA again, which is prohibitively costly for us to do again? Or is there a middle path where we can talk about what are the critical pieces of data which the agency would like to see from the industry to allow a switch from a current device to a similar device which has a better cost positioning and availability?

Third, which is a Holy Grail for us, and it's been debated quite a bit—we have indications or avenues available in different markets across the world, where, if the brand is a disposable device, we have been able to talk about a reusable device. What it does is it kind of helps you from a cost positioning where a single device can be used for three years. The cartridges could be provided, or the primary container can be provided, and there are examples in the market where such solutions have been already put in place. But the issue is, the brand has done it—easy to follow on. But if the generic wants to do it, is there a way we can talk about it? Because it has huge implications. If you talk about the amount of plastic we are going to be putting out with disposable devices, it's humongous. It's not just a device. They arrive packaged in trays at our sites, and they're all plastic, and there's no way we can do anything with that plastic except get rid of it in some form. But it overall adds to the quotient of ESG for us that we're now putting more and more plastic into the market. And that's one of the things which, you know, maybe it's not a low-hanging fruit, it's not a medium-hanging fruit, but it's something to talk about with somewhere. It's going to hit us in some form in the future.

That's all I had. Thank you.

William Chong: Thank you, Manoj. I'll circle back here to Megan, who is, I think, our last public panelist who didn't have the chance to come to the podium. So Megan.

Megan Conrad (University of Detroit Mercy): All right, thank you. So I'm really interested in the different ideas for both statistical and study designs for the comparative use human factors process. I think we talk a lot about how to determine the sample size, and I think we need to remember that our user interface usability sample sizes came from a different place than some of the statistical design studies, right, that we were talking about. How many participants do we need to identify all potential use errors versus how many participants do we need to compare these two aspects and say that we have statistical significance? So also along those lines, what our metric is.

So the chi-squared is an interesting idea, because there's some flexibility in that model. We don't need to assume that we have a normal distribution of our data. And I think certainly, if we're counting these errors, and I would expect also with user rates, however we're computing them, that that would be a skewed distribution. So there's some validity with that chi-squared model. But I'm also interested in new metrics. So Melissa mentioned incorporating risk, incorporating close calls, difficulties, incorporating using your observational data to say, what's our probability of a user—what's the probability that user leads to harm, and how we can use that—some other additional metrics to further the studies and get more information. So they're my ideas.

William Chong: Good. Thank you. All right. So we'll circle back. We'll start with Brandon, work our way this way so everybody can start thinking about what they want to say.

Brandon Wood: No, absolutely, thank you, Dr. Chong. So I think these were all really great presentations and comments from the agency. I also think we'd be remiss without saying it's been a great collaboration with the agency, especially recently. You know, here, the DDCP conferences, just the individual product-specific feedback. I think the collaborative effort from the agency and industry side has been on full show, you know, for a couple of years now, and I think that's greatly appreciated.

Just going back to the presentations, though, I kind of see this from both a short-term and long-term perspective, because we're not going to turn the coin on its head, you know, overnight, right? So we have the current paradigm that we live in that might have some quick wins to help us out with the current framework, and then some longer-term considerations that might change the landscape of what we're doing.

So I already spoke about research to assist in early identification of other differences and just kind of leaning on what was mentioned earlier—research to clarify statistical expectations as it relates to determination of non-inferiority margin, utilization of real-world evidence to facilitate regulatory flexibility. I'm aware that within the agency's fiscal year 2025 budget, there's some indications to allow Congressional authorization for flexibility with drug-device combination products. I'm really interested to understand, you know, what the agency has going on on that front.

But then long-term, obviously the alternative considerations—and the common theme with both what I presented with IEC 62366 fit-for-purpose and with Heidi's chi-squared method is the incorporation of root cause analysis. You know, root cause analysis is paramount to successful human factors principles. And, you know, that's—I don't think it's coincidence that within both of those alternative proposals, root cause analysis is a big consideration that plays into the consideration. So just a couple of thoughts from my end.

William Chong: Thank you, Brandon. Vivek?

Vivek Viswanathan (Rubicon): For that, yeah, I would like to thank the agency and all the speakers for such wonderful talks. So resonating with what has been discussed in the multiple presentations, I think one area which is very clear we would need more guidance and more clarity is the specific area between the other UI differences and beyond. So there I would admit that there exists some amount of, I would say, ambiguity or lack of information which would definitely support ANDA sponsors and companies who are developing generic medicines to facilitate generic drug product development.

Because, as I mentioned, we have evolved a lot, and with the device becoming more and more complex, more and more complicated over the years, there is a lot of IP which is associated with the device. So we need to have definitely some options like alternative designs and other statistical outcomes which could promote the entry of generic drug-device combination products in the market. Yes.

William Chong: Thank you, Vivek. Manoj we're going to skip you and we'll come to Melissa.

Melissa Lemke: Thank you. I'm going to kind of go a little bit out there with this one. So actually kind of go back to the source of truth and the regulation itself in this pathway. And I know for human factors there's a big—there's a lot of issue with not being able to innovate. And that's really, you know, this pathway is not intended for that from the regulation standpoint itself. And

so we look at, you know, the high risk of negative transfer, and that the regulation itself really drives the research question that we're trying to answer and provide evidence to the agency for.

But I'm really intrigued by this real-world evidence, and I think there could be some collaborative efforts happening between human factors science and the clinical efforts that are happening with real-world evidence. And then kind of looking at what is actually representative use out there, because I think the regulation misses some potential use-related risk. You know, we look at assuming the RLD is safe and effective, and only looking at the new generic. Heidi talked about going the other way. You know, what is the agency really expecting, and how prescriptive does the agency want us to be when human factors is really focused on representative use? And this, I think, kind of narrows because of the regulation itself. And I don't know if you guys work on changing regulations, but I don't know. Maybe...

William Chong: I do not. But thank you for raising that.

Heidi, I think you're up next.

Heidi Mehrzad: Did you say probably not?

William Chong: I said I personally do not.

Heidi Mehrzad: Got it. Just to clarify, the chi-square model is not Heidi's model. I did not invent that. I'm just trying to clarify it's just an idea that I'd like to explore. I really made all my points in my presentation. I think we're far off from using human factors for what it is in this pathway. I think we lost sight a little bit of what human factors is and what the role it's supposed to fill in this. And we're trying to make a statement that allows us to say if you started with the generic, you can use the RLD, or if you have the RLD, we can rely on our design to make sure that you can use the generic. And I think when we think of it that way, we come back to the ultimate route, and that is the human using this.

And that's why I have such an aversion towards just using user frequencies. And I think the idea that I presented is more an explanation of a model—what could a model look like if we thought more of it from a social science aspect and focused on actual behaviors leading to unsafe, ineffective use? And I think that's more of where human factors should be heading.

I have a strong regulatory background, but I'm not an expert, so I don't really go into depth of thinking that this is all a question of regulatory pathway and legislation. I think it's looking at a problem that we overall have in human factors.

And just not to go too much off on a tangent, but I'm okay with saying this publicly, but I also don't think industry takes human factors very seriously. And with that comes a big struggle in trying to do something that has validity and has robustness behind it, like a comparative use human factors study that is leaning on actually using quantitative and qualitative data and not just one or the other. And I think that that gripe within industry—one crowd yelling qualitative data, human factors needs to be qualitative data, and the other side yelling, well, but if you don't know how many, then why—quantitative is just as important. I think that that is one of my original gripes when we look at stuff like this.

Because, yes, is the NI model not optimal for this? Probably not. But can we make the statement that the stats are too complex? I don't know. We work in medical device development. I think we all need to be able to handle statistics and math in science. And so I think sometimes

when I see presentations or when we discuss human factors, sometimes we lose sight of what it is, and that is ensuring that the human using your product is using it safely, effectively, and without causing any harm to anybody else or the environment around them. And I think when you look at it that way, it's easy to understand where Brandon was going, it's easy to understand where Melissa was going, where Amy was going, where your presentation with the case studies was going, right? Like you look at the human and not just at the frequencies of what occurred or how many use errors occur during one specific time period, and then make a statement based on that. And I think when you divorce yourself from that, then other alternative models are easily discerned.

William Chong: Thank you, Heidi. Amy.

Amy Lukau: Yeah, I think for me, one of the critical parameters and difficulty that I have of the current design—the current guidance—is that it doesn't necessarily encapsulate emergency use auto-injectors and our emergency use products. I primarily work with emergency use products for all—critical or all tasks are critical. So it's not enough to say, oh, the user rates—well, if the user rates are high, then all your patients are dead, potentially, and are significantly harmed. Not only that, but you're potentially looking at recalls of the respective product.

So I think that we need to consider possibly adding a corollary to the existing comparative generic drug guidance that was issued in 2017, as it relates to generic drug emergency use products and its RLD. And more importantly, just looking at also what happens if there's no context in relation to—if their product—if there's no information on the RLD listed and/or it's discontinued. We do see this a lot with respect to emergency use products.

So that's one of my recommendations that I would give to the agency. And also we do have to rely specifically on use-related risk analysis, because I do have to identify user interactions to assess and mitigate use errors and are potential indications of where harm may be inflicted on the patient. They only have one try with emergency use generic drug products, and we have to get it right the first time, unfortunately, and the agency holds us to a higher standard. Hence the reliability of the five nines, you know, 99.999 with a 95% confidence level. So I think this is one of my major considerations, and more importantly, to how we assess minor and major design differences. I mean, for human factors, This is a very shade of gray, because we do also have to consider essential performance requirements which also could be construed as design inputs. So that's just my take.

William Chong: Thank you, Amy. Daliya.

Daliya Bharati: We had really wonderful discussions today. All the presenters came up with wonderful presentations. One important thing that comes out here is that the industry requires a lot of clarity, be it when one discusses the human factors studies, or be it at the earlier stage of whether all other user differences really warrant human factors studies.

So basically from the industry perspective, we look for more clarity in the guidances, maybe describing illustrations on how certain differences should—can be or should be taken further, because with increasing innovative drug product-device combination products available now, the IP barrier also, you know, is surmounting. And with that, if there is less clarity in how to go ahead, or if we have to invest a lot of time in cycles of controlled correspondences or the ANDA meeting requests, and then additionally, also, during the ANDA reviews, that can probably, you

know, not help the generics actually take up those products for development. And in fact, you know, generic entry is then compromised.

So the important part is the agency coming up with maybe more clarity and more details on how things can be taken ahead.

William Chong: Thank you, Daliya. I'm going to ask some of our FDA panelists to share what they've heard as well. We'll ask you to try to keep it a little bit brief where we realize we're a little bit over time. So maybe also start from the end. Stella, would you like to go first?

Stella Grosser (FDA): Sure, I mean, I don't want to go first, but here I am.

I think what I've heard is that we need to have a better definition of a success or a successful use or what's causing an unsuccessful use. You know, we've broken it down into little tasks, and I know in our reviews we always have an argument: do we do it task by task, or do we do an overall? But I think that overall needs a lot more thinking. I also agree that the difference between chronic use and emergency use products needs to be thought through—both, you know, the standards, the criteria, what we would accept.

And the other thing I'm not sure I heard, but I'm interested in going back to the regulations to see what we need to do for the device that's connected to a generic drug in the first place. So thank you.

William Chong: Thank you, Stella. I think, Edna, that puts you next.

Edna Termilus (FDA): Yeah, I think just to—I think the high-level takeaway for me from this is that we do need to spend some time and some space thinking about how to incorporate qualitative assessments into, you know, our assessments of comparative use human factors studies. And the other big takeaway for me was that there is sort of a little bit confusion still in our assessments of minor differences versus other differences and how we can provide, you know, more guidance and how we can develop more research to sort of delineate that a little bit more that will make doing future assessments of, you know, other differences and design differences for applicants that are submitting—making that road a little bit clearer and having that clarity before any additional studies or any additional information is being submitted.

Markham Luke: Brandon, you called out the—it's the 40th anniversary for Hatch-Waxman. That was 1984, and 1987 FDA established the 510(k) program for review of devices. And so the device review process has always been risk-based. It's always been: let's look at what risk categorization is—is a Class I, Class II, Class III, risk III being the highest risk. And throughout that process, there have been numerous attempts at changing how 510(k)s are reviewed via the substantial equivalence process, etc. But it's always fallen back to looking at what the predicate devices are and what the risk-based assessment of the changes to the predicate or the new devices coming in. And so that paradigm has served FDA and the American public successfully for many decades. And so they're thinking along those lines: hey, maybe there's some wisdom there that can be imparted from the historical view of devices.

And the 2000s, FDA started looking at combination products and pulled some of the device review aspects and combined them into CDER for our NDAs and now ANDAs. And these products have taken off. They become really important, as has been pointed out by many talks during this meeting. And how we regulate the device or constituent of these products is going to be—we should give it very careful thought and be adaptable to potential changes that will make

it better. And some of it may come full circle, and we may find ourselves regulating these devices very similar to how CDRH is regulating devices. We have the authority in FDA to do that in that context.

The combination product statutes are constructed on top of historical precedents for how we review medical products. So it's really interesting from a historical perspective, and history can portend the future potentially. So we just want to think about how we do these things in a historical perspective.

William Chong: Thank you Markham. I think we do need to kind of go quickly. So if you could focus on what you think would be areas for research, that would be really helpful.

Kyran Gibson (FDA): I promise I'll be quick. So as you can see on the screen, I am a biomedical engineer, not a human factors engineer. So there was a lot of information kind of to absorb throughout the day between Session 3 and Session 4. But I will quickly just reiterate what both Brandon and Markham have said about the collaboration from a device standpoint. The collaboration that I've seen today and previously, as I have started to become involved in these discussions, it's really admirable, and to be able to identify pain points and from both perspectives say, "Okay, there's something that we need to work on." I think that's admirable. I also wanted to thank Amy for her emergency use details included in her presentation. I thought that got to the crux of what I deal with on a daily basis. But yeah. That's it.

William Chong: Jason

Jason Flint (FDA): I'll try to be—I'll try to be quick for once. So lots of great discussion today. Lots of great presentations. Thank everyone for presenting those. I did hear a lot about like we want to do CUHF differently, or would have alternatives to conducting—to how do I conduct CUHF studies. I actually think that—and I think we touched on this a little bit at CRCG a couple of months ago. But, you know, what can we do to demonstrate that? Okay, we've identified that these differences exist, right? But demonstrate that they're not—that they may not be impactful, right? And it's to sort of the approval of a generic.

So I heard some comments about—in the presentations about how much data is available, but I think that what we rarely see, or at least what I've rarely seen in CUHF, or even in comparative analyses, probably more so in comparative analyses, that data is either presented as data, right? Not as information. That's an important distinction. But also there's no story, right? How do I connect that information story? How do I connect that to this idea that yes, there is a difference. We've thought about that. We have all of this other evidence that supports that it may not be an issue. And so I think research in that—sorry, research in that area would be of interest to me. And I also think, and I didn't get a chance to ask the question, but the real-world, real-world evidence is a very interesting concept to me, and I'd like to see more on that. Thanks.

William Chong: Thanks, Jason. Ariane, do you have anything to add?

Ariane O. Conrad (FDA): Hi! I don't really have anything to add. I'm going to move quickly. Just thank all the panelists and everyone for the presentation today. Thank you.

William Chong: Thank you. Yeah. Same question to you, Bob.

Robert Berendt (FDA): Yeah, Bob, from OPQ. So I just wanted to thank everyone for their feedback and also—although manufacturing and, you know, product quality wasn't necessarily

touched on directly, I certainly appreciate some of the feedback that was received with respect to trying to make sure that our expectations, as far as risk-based analysis for supply change and material changes, especially during the long-term product lifecycle—that's certainly something that we consider, and we try to be as outward as possible with everyone. So thank you very much.

Heidi Mehrzad: Is it okay if I can say one more thing, because I'd like to address what Jason said with the cohesive story? This isn't something that was in my presentation, and I don't bring it up as a topic, because I can go really, really off tangent. But in human factors, device validation studies, right, when you prepare your submission, you have to submit a human factors engineering report. In comparative use in ANDA filings, you just have the comparative use. The cohesive story is gone. So you're just focusing on this one story.

And I think, to Jason's point, I think one thing that could really be helpful if we thought of this as something that we could supplement our comparative use studies with or in general, an ANDA filing human factors work with an actual report of the overall development from beginning to end in human factors, just like we do it with device where we tell the story from the beginning to the end and not just focus on the difference that's in relation to the three or four design differences that affect that particular product. I think that would be supplemental and really helpful and easy to implement.

William Chong: Thank you, Heidi, and thank you again to all of our panelists and our speakers. That ends our session. I'd like to welcome Dr. Lionberger to the podium to close us out for this workshop.

Robert Lionberger (FDA): Hi, thank you, Bill. Thanks. Thanks everyone for participating in this workshop, you know. I think I attended almost every session, almost every minute of every session. And, you know, I think it shows the breadth of scientific issues that affect the generic drug program. And, you know, through our research program, we're really committed to using the resources that we get through the user fee program to address those issues and make the review process and the development of generic products better, increase access to that. So really appreciate all of the people who came as panelists to provide the generic industry perspective. I think that's really, really important to FDA staff as we try to plan a research program, right? Hearing what are the actual things that you struggle with from the industry. So, you know, really value the industry participants who have come here and provided that perspective. And, you know, I think that's really valuable to our FDA participants who have to take this back and go back and say, "All right, now that we've—as we've heard these products, we understand what our challenges in the review, but we also understand some of the challenges that are going on in development before things come into FDA to really formulate a research program that's as effective as possible at accelerating access to generic products." And so appreciate people across all of the sessions who did that. All of the people who made the public comments providing their input and thought into that. That's very helpful to us, really says, you know, we're open to looking for innovative ideas in any different place that we come from. We try to create this—this year we try to create more opportunities to hear from as many different people as possible in that context. So with that, this workshop, I think, is a fantastic success. But it's only a success because there's a huge amount of people that are working behind the scenes to get all of the logistics right. Get all of the people in the right place, get the questions done, get the agenda, get the publicity, you know. Probably, you know, one of the biggest challenges is actually getting this room and locking down the access to it. So, you know,

we want to acknowledge like a lot of the people who are involved in this—a few groups. They want to—the great room staff who provides the logistics, also the—getting this out to everyone who's watching online—important part that there's a Zoom webinar administrator team that helps with that from the FDA. We're actually helpful this year. We actually had a little coffee room there. If you were here last year—no coffee this year. Coffee, thanks to like our Sodexo staff, who were able to get that open for us.

Think you've heard about this meeting because of the OGD and OPQ communication staff—Fallon Small, Ruth Augusta, who helped make sure it gets out through all the different channels, and people are aware of that.

I think within OGD, a lot of the project managers help keep this on track. Karen, Shivangi, and Calliope to make this work through this. There are a lot of groups internally that are involved in the planning this. Internally, we have a GDUFA III research program committee that, you know, helps develop the program—helps develop the programming. Want to especially acknowledge Sarah Rogstad for her work in organizing the workshop and getting all the faculty. Want to acknowledge Maria Monroy Osorio for our key project manager who's been running the material back there.

Want to acknowledge Jessie Floura and the ORS research team, Jessica, Christina, Sara, Eileen, and Sarah for doing all of the logistics, being at the desk and the organizing part, getting all of the people in place here. Want to recognize again, there's a lot of other ORS fellows from the OS who participated in manning the tables there and working on the logistics, getting all the signage up to get you here in the right places.

On the content side as well, we work closely with the AAM and the other stakeholders through the GDUFA regulatory science program and the meetings we have with our stakeholder group, biannually to help to provide input and planning into this. We work closely, as you've seen, with the Center for Research on Complex Generics. And Anna Schwendeman and Jim Polli both were here. That group does a fantastic job of trying to also get industry input into this program and helps with different workshop planning and activities around generic drugs that are very valuable. We also appreciate the contribution of the USP group who provided in—another group that sort of can aggregate information from different generic drug stakeholders and appreciate their contribution and providing talk at this workshop as well. So really looking for wide ranges of inputs.

And with that, bring this workshop to a close again. Thank you. Everyone who participated in all the different ways to make this a very successful two-day event. Thank you all.