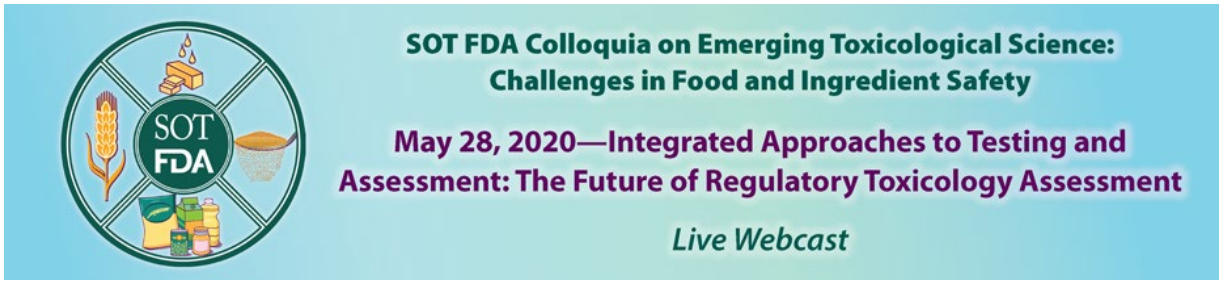

Real-Time Captioning

Note: This is not a transcript.

May 28, 2020

Integrated Approaches to Testing and Assessment (IATA)—The Future of Predictive Toxicology

Schedule

8:30 AM–8:40 AM	Welcome and Speaker Introductions Suzanne Fitzpatrick, Colloquium Co-Chair, US FDA CFSAN, College Park, MD
8:40 AM–8:55 AM	Integrated Approaches to Testing and Assessment (IATA)—An Introduction Lidia Ceriani, Humane Society International, Brussels, Belgium
8:55 AM–9:15 AM	Uncertainty Characterization in IATA for Chemical Safety Assessment: Overview of Available Guidance Andrea-Nicole Richarz, ECHA, Helsinki, Finland
9:15 AM–9:50 AM	Learnings and Recommendations from Four EU-ToxRisk Case Studies on Applying New Approach Methodologies Data to Support Read-Across Susanne Hougaard Bennekou, Technical University of Denmark, Copenhagen, Denmark
9:50 AM–10:00 AM	Break
10:00 AM–10:35 AM	IATA as an Opportunity for Next-Generation Risk Assessment: The Propylparaben Case Study Gladys Ouédraogo, L'Oréal, Paris, France
10:35 AM–11:10 AM	Global Harmonization Efforts for Skin Sensitization IATA Nicole Kleinstreuer, NICEATM, Durham, NC
11:10 AM–11:45 AM	Applications of New Multi-Organ-Chip Tools for Toxicity Assessment Reyk Horland, TissUse GmbH, Berlin, Germany
11:45 AM–12:45 PM	Roundtable Discussion Moderator: Lidia Ceriani All speakers

Welcome and Speaker Introductions
Suzanne Fitzpatrick, Colloquium Co-Chair, US FDA CFSAN, College Park, MD

Suzanne Fitzpatrick: Are we ready to start the meeting?

Lidia Ceriani: Yes. Sounds good, Suzanne.

Fitzpatrick: Are we ready? Can we start the meeting?

Ceriani: I think so.

A-V Staff Member: Yes, you can start.

Fitzpatrick: Welcome, everyone, to the latest SOT FDA Colloquia on Emerging Toxicological Science Challenges in Food and Ingredient Safety. This is a very exciting topic for me, the Integrated Approaches to Testing and Assessment: The Future of Regulatory Toxicology Assessment. I can move the slide, can't I? Next slide.

So, this is our partnership with the U.S. FDA Center for Food Safety and Applied Nutrition and Society of Toxicology. This is our fifth year of quarterly colloquia. We're trying to stimulate a global dialogue, and I think in this topic is a very important for discussion with FDA scientists and for global partners. It is not a regulatory—we focus on science. We're not getting regulatory advice or discussing food or ingredient regulatory issues, but other regulatory issues at FDA. Next slide.

They're open to the public at no cost, and you can see from this symposium today, we have large sector of global science community participating. We're recording; the slides for all of the colloquial are available on the SOT website. Next slide.

Here is just a list of the organizing community, we have wide swath of FDA and SOT and next slide.

And I really want to think today Lidia, who has put a lot of effort into putting this really excellent symposium together and I want to thank all of the speakers. I have seen the slides and they're wonderful and I think they will be informative. Thank you for the Humane Society International and Lidia to put this together and everyone really enjoyed the symposium. We think it's an important topic and we're glad to welcome so many of our colleagues today. With that, Lidia, I will turn it over to you.

Speaker Introduction
Lidia Ceriani, Humane Society International, Brussels, Belgium

Lidia Ceriani: Thank you, Suzanne, for these wonderful introductions. I'm Lidia Ceriani, and I'm working at Humane Society International. Today I am excited to chair this colloquium together with Suzanne Fitzpatrick. Next slide, please? There's a delay in the slides.

I will go ahead. Today we have seven distinguished speakers with extensive knowledge and expertise in the development of integrated approaches to testing and assessment, or IATA. On the schedule of today we start by introducing the IATA concept in order to provide you with the background [inaudible]. Our second speaker is Dr. Andrea Richarz, a scientific officer at European Chemicals Agency. He has previously worked for many years at the European Joint Research Center in Itlay, mainly in the area of toxicology in read-across. He was give us a talk on Uncertainty Characterization in IATA for Chemical Safety Assessment and an Overview of Available Guidance.

Our third speaker is Dr. Susanne Hougaard Bennekou who works as an advisor of the National Food Institute of the Danish Technical University and she is also Vice Chair of the European Food Safety Authority [inaudible] Committee. Previously she worked for almost 20 years as an [inaudible] advisor in regulatory toxicology at the [inaudible] division of the Danish FDA. Her talk with focus on Learnings and Recommendations from Four EU-ToxRisk Case Studies on Applying New Approach Methodologies Data to Support Read-Across.

Then we will have a 10-minute break, and after, we will have another talk given by Dr. Gladys Ouédraogo, who works as a scientific officer at L'Oréal Research Innovation. Dr. Ouédraogo began working in the Safety Research Department of L'Oréal Advanced Research in 2003, where she then became head of the [inaudible]. Now, she leads research collaborations with [inaudible]. Dr. Ouédraogo is going to present on IATA as an Opportunity for Next-Generation Risk Assessment: The Propylparaben Case Study.

Our fifth speaker is Dr. Nicole Kleinstreuer with the National Toxicology Program of the Interagency Center for the Evaluation of Alternative Toxicological Methods, or NICEATM, where she leads NICEATM computational toxicology work. She also [inaudible] NIEHS [inaudible] computational biology branch. Her talk today will focus on Global Harmonization Efforts for Skin Sensitization IATA.

Then our last speaker will be Reyk Horland, who is head of [inaudible] development at TissUse, a Germany-based company that develops [inaudible] platforms. Before joining TissUse, Dr. Horland worked at the German [inaudible] Research Center at the Technical University of Berlin, where he developed tissue models that can mimic human biology *in vitro*. Dr. Horland's presentation will be about Applications of New Multi-Organ-Chip Tools for Toxicity Assessment.

After the talks we will gather for a panel discussion during which you will have an opportunity to ask questions by writing them in the chat box available on your screen. And if you want, you can always submit your questions in the chat box also as people are talking so that, in case we've got enough time to answer one or two questions after each presentation. Instead, if we're running out of time, we're going to [inaudible] all questions [inaudible].

I think this is all for the introduction and I can jump into my presentation, which is the first one of today.

Integrated Approaches to Testing and Assessment (IATA)—An Introduction
Lidia Ceriani, Humane Society International, Brussels, Belgium

Again, [inaudible], I hear some background, if you can please unmute yourself, thank you. Good morning and good afternoon. For those who just connected to this colloquium, I'm Lidia Ceriani and I'm working as a regulatory science advisor for Human Society International. Next slide, please.

There's a delay. I will continue. Before starting the presentation, I declare that I do not have any conflict of interest. It seems that there is a delay, in the slide presentation.

Thank you very much. In the first presentation I am going to provide you with an overview of what IATA are. What is a typical IATA workflow and what is the strategies for discussing the approach? I will conclude with a brief [inaudible] of this project at the global level. At the end of the presentation I will give you with the background and the definition of the main elements and topics that you are going today hear in these presentations.

I cannot see my slides. But I will start by introducing it. As we know, worldwide chemical regulations are challenged by the large number of chemicals that require [inaudible] human health environmental impacts. This is because current testing approaches are too resource intensive in terms of time, money, and animal use to evaluate all chemicals on the development [inaudible] market. There is a growing recognition for the need of more efficient and robust methods and strategies to assess the hazard, exposure, and risk of the wide array of chemicals to which humans are exposed. At the same time there is an increase use of new technologies like in vitro tests, omics approaches, computational methods. And these new approach methodologies [inaudible] between the number of chemicals. [Inaudible] not yet accepted at the regulatory level as a stand-alone prediction. So, we need to combine [inaudible] in a joint approach to achieve a fully acceptable level of protection for human health. In both situations, the Integrated Approaches to Testing and Assessment, or IATA, represent a possible solution to [inaudible]. Next slide.

IATA is structured approaches that integrate and weight difference types of data in order to [inaudible] which we have previously identified. New data can be generated in a [inaudible] way until we are able to answer one question. These testing strategies can be used to perform hazard identification, hazard characterization, and/or safety assessment of a chemical or group of chemicals. Within IATA we can use different types of methodologies and used different mechanisms thanks to the AOP knowledge. Next slide.

In practical [inaudible], I'm waiting for the slide because it's really useful to see it, but I will start because I'm going to say a lot about it. The IATA workflow is composed of three main blocks: existing data, weight of evidence, and generation of new testing data. The first step of an IATA should always be the problem formulation and this step involves the consideration of the regulatory application. So, for instance, are we dealing with a prioritization exercise, or do we want to perform a method assessment? Based on the context of use we also know what is the consequence

level of acceptable uncertainty associated with the decision being made. We also want to consider types of endpoints we want to address and any limitations.

At this point we can start to develop our hypothesis that is basically the explanation of our problem. This hypothesis can be for the revised of the existing data [inaudible] generated information. In general, though hypotheses guide the choice of test methods to be used and the sequence of the testing. Once we have developed our hypothesis, we start to collect existing evidence and all of the information available on our substance or group of substances. And then we can perform a weight of evidence evaluation [inaudible] and relevance and a weight that is assigned to each line of evidence.

And then we combine the lines of evidence to decide if sufficient strands of evidence are available to address the question posed on the hypotheses. If yes, we can conclude our assessment; if not, we generate new data—always keep in mind the question we want to answer. [Inaudible]. We can make use of seven new approach methodologies which is a [inaudible] term to refer to nonanimal methods, like QSARs, *in vitro* tests, *in chemico* tests, high-throughput screening, organoids, and organ-on-chip. In addition to NAMs, we can also use human data in *in vivo* tests. At the same time, toxicokinetic [inaudible] of the chemical or group of chemicals can be included in the test. Afterwards the new information generated will the go back into the framework for the evaluation. And again, if we are sufficiently confident with the outcome, we can conclude our assessment; otherwise we're going to define our assessment through new tests, and at this stage, we can also refine our hypothesis.

As you can grasp from this workflow, and IATA is an interrogative process, and one of its benefits also lies in the potential breadth information that can be used in the assessment. Most of the data gathered at the generation level the adverse outcome pathway, or AOP concept, can facilitate and support the compilation of information in types of testing alongside hypothesis generation. Next slide.

I would like to discuss what is an AOP. It is a sequence of key events that commence with the initial interaction of a stressor with a biomolecule in the [inaudible] tissue and organ; this is the molecular initiating event. It progresses to a series of intermediate key events at different levels of biological organization, cell, tissue, and organ. And it culminates with an adverse outcome at the organism or population level. It provides a connection between mechanistic data sets and [inaudible] outcomes. At the same time, as it is represented on the slide with the blue boxes, AOP allows for the mapping, organization, and integration of different types of information, ranging from *in silico* and *in chemico* data to field study data around the molecular initiating events, key events, and adverse outcomes. Next slide.

I'm really sorry for this lag in the slides. An IATA should ideally be mechanistically informed, meaning it should be based on knowledge of the mechanisms through which [inaudible] toxicity. For instance, if we understand the likelihood of the facts like the [inaudible] at lower levels of biological organization, using, for example, structure-activity relationships or *in vitro* methods, we can also understand [inaudible] at higher levels of biological organization. Therefore, we can better understand which key events of our AOP or network of AOPs we should query to answer our question.

The AOP concept also allows us to evaluate existing information that is available for the chemicals of interest in our structures. And also identify and generate the type of information [inaudible] key event that might be required to increase the level of confidence in the evidence. AOPs are expected to provide insight into the biological relevance, reliability, and uncertainty associated with the results from the different test methods for regulatory use. Bear in mind that when we define the confidence for a given decision [inaudible] assess the weight of evidence of the AOP or AOPs of interest [inaudible] provide guidance for the weight of evidence assessment of AOPs. Next slide.

So, there's a range of IATA, from more flexible, non-formalized judgment-based approaches like grouping and read-across, to more structured, prescriptive, rule-based approaches like the defined approaches. In both cases that can be used on their own or as part of an IATA. The main difference [inaudible] read-across approach [inaudible] expert judgment. For example, in the [inaudible] a conclusion may be open for interpretation. Instead, the defined approaches are standardized approach and the predictions are rule based, so from a regulatory point of view, the defined approaches facilitate the regulatory use of testing of this type of IATA. And let's see why. Next slide.

In a defined approach, mechanistic data generated by *in silico*, *in chemico*, *in vitro* methods that are deemed to be fit for purpose are evaluated by means of a Fixed Data Interpretation Procedure, which can be for instance a mathematical [inaudible] prediction can be used on its own or together with other sources of information within an IATA. Since this is a fixed decision-making process, [inaudible] regulatory consistency and certainty. So, the final approach using an specific set of data test methods, I am going to reach the same conclusion of another person who is using the same defined approach, the same data set, and the same test method, and this way, I've removed the expert judgment component, and my conclusion is not open for interpretation. [Inaudible] skin sensitization assessment, they let in more defined approaches case studies for predicting skin sensitization potential using a combination of *in vitro*, *in chemico*, and *in silico* data. [Inaudible] validation at the OECD level. In Nicole Kleinstreuer's presentation of today, we will focus on these defined approaches, where more details will be provided. Next slide.

Over the last decade many projects were launched at the OECD level focused on the development of IATA. The OECD IATA case study project that was launched in 2015 to increase experience with the use of IATA by developing case studies. So far, 23 case studies have been developed by Member states, industry, and other types of organizations, and they are available on the OECD website. Case studies evaluated by the regulators of Member states at the OECD level [inaudible] identify areas of improvement in uncertainty in the IATA workflow components but also to offer a suitable [inaudible], which can also be challenging. In two presentations of today we are part of the last review cycle of 2019. [inaudible]. There are also two OECD guidance documents on IATA for the prediction of both the skin corrosion irritation and serious eye damage and irritation for chemicals. And based on these IATAs, the respective OECD test guidance has been updated to include reference to predictive testing strategies before performing any animal testing. As I mentioned in

the previous slide there are 12 defined approach case studies for the skin sensitization prediction. Next slide, please.

The US EPA was to develop a defined approach by integrating computational [inaudible] for the prediction of estrogen receptor bioactivity. Together with NICEATM, a [inaudible] defined approach has been developed for predicting the estrogen activity. The primary goal of [inaudible] is to quickly screen large numbers of chemicals and to prioritize the active chemicals for more detailed testing. Use of the estrogen receptor defined approach has been accepted by the EPA as an alternative [inaudible] use in the Endocrine Disruptor Screening Program's [inaudible] battery. Currenlty, EPA is considering whether the androgen receptor approach is a potential replacement for other existing tests regarding the same screening program. Lastly, I wanted to mention the Next Generation Risk Assessment Framework for Cosmetics, which is a sequential hypothesis [inaudible] risk assessment workflow that utilizes IATA concepts, and you will hear more about it during the presentation of Gladys Ouédraogo later on. Next slide.

To conclude the take-home message of this first presentation of today is that IATA are tailored and predictive hypothesis-driven approaches which integrate existing knowledge on chemicals with mechanistic data across different levels of biological organization, together with exposure and other sources of information. It's for targeted testing or derivation of assessment conclusions. In this way, we can optimize resources and go beyond hazard information, also including kinetics and exposure data. With the IATA we can also increase confidence in the use and application of New Approach Methodologies (NAMs). And IATA also plays a key role in shifting emphasis from traditional testing based on apical endpoints to mechanistic-driven testing strategies. Defined approaches can facilitate the regulatory acceptance and use of IATA since they are prescriptive and rule-based approaches and the expert judgment component is therefore eliminated. In conclusion, IATA can potentially be used for prioritization purposes, classification labeling, hazard identification and characterization, and risk assessment. Next slide, please.

All the references that I used by comparing the slides are recorded on this slide, next slide. Last, I would like to thank everyone from HSI, SOT, FDA, and the speakers for the organization of this colloquium. Thank you for your attention.

Now we're on time so will turn it over to two Dr. Andrea Richarz from ECHA, whose presentation title is which is Uncertainty Characterization in IATA for Chemical Safety Assessment: Overview of Available Guidance.

Andrea, the floor is yours. Thank you.

Uncertainty Characterization in IATA for Chemical Safety Assessment: Overview of Available Guidance
Andrea-Nicole Richarz, ECHA, Helsinki, Finland

Thank you, Lidia. Good morning, everyone, or good afternoon from my side of the world. I'm Andrea Richarz. I'm the scientific officer for the European Chemicals Agency in Finland. I'm very happy to build on the introduction from IATA from Lidia

and give an overview of [inaudible] some specific aspects regarding available [inaudible] assessment. Next slide.

I have no conflict of interest [inaudible]. Next slide, please

For the application of IATA [inaudible] guidance will be very useful. We have the same, different [inaudible] levels and aspects for which guidance is needed. We will also see that they are related to different levels of uncertainty.

Lidia Ceriani: Andrea, we've got some [inaudible]. Andrea, I think that you've got some internet issues. Now it seems good but sometimes the voice is deleted. [inaudible].

Andrea-Nicole Richarz: You're breaking up as well.

Staff Member: Would you be able to try calling back in?

Richarz: I can try to change my Internet connection.

So, I am thinking that, is Gladys online? Sorry, is Suzanne online? Suzanne, do you want to step in on with your presentation while Andrea preps, we will check back with her Internet.

Susanne Hougaard Bennekou: I can do that for sure. I hope you can hear me.

Ceriani: Yes, very well. Thank you and sorry for this.

Hougaard Bennekou: No problem.

Ceriani: Jasmine, can you please put on the presentation of Susanne Bennekou?

Staff Member: Yes. Susanne's is up.

Hougaard Bennekou: Yes, I see it. Thanks.

[Change in speakers; technical difficulties]

Learnings and Recommendations from Four EU-ToxRisk Case Studies on Applying New Approach Methodologies Data to Support Read-Across

Susanne Hougaard Bennekou, Technical University of Denmark, Copenhagen, Denmark

Thank you, Lidia. Thank you for the nice introduction and also for inviting me to speak here. I would like to tell you about the learnings we have had by developing case studies on read-across and where we tried to apply new approach methodologies. It is kind of a lot of detail that we have in this case studies and I'm trying to cover all of it in 25 minutes, which is not possible. So, I will try to synthesize

the most important learnings and in terms of how can we facilitate the regulatory recognition of these new methodologies. Next slide.

I also have no conflict of interest to declare. Next slide.

As I said I will explain these case studies. I will not explain the results of them. It is more the approach and how we use new approach methodologies in them. And I will provide learnings and recommendations on how you can report such data, how you make sure that the performance of the NAMs is adequate and it is also documented adequately. And in applying all these different data streams, there is certainly an issue with how to actually carry out the weight of evidence and the uncertainty analysis in the end. Next slide.

These four case studies were developed in the EU-ToxRisk project. I'm not sure every one of you are familiar with it but it is integrated European flagship program driving mechanism-based toxicity testing and risk assessment for the 21st century. And here so you see on this map we have the hypotenuse for all of Europe and it is a Horizon 2020 funded program and we have big companies, small companies, regulators, and of course academia. So, this is the, what I will present is some of the work here and actually it is taken around three years to develop but we did now across all of the different partners.

One of the big objectives of this project was to provide new robust read-across procedures where we would incorporate toxicokinetic data and similarity evaluation in the level of key event activation. And also, we would use various approaches, very high-throughput up to very sophisticated models. This would be in order to try to fill data gaps in a read-across procedure. It is very much driven from the EU regulation because REACH, they actually, it is quite a moderate regulation there. You are allowed to use read-across for pesticide regulation in Europe. It is much more limited to the possibilities there. It had a clear REACH focus, but we also, by providing the case studies, we also wanted to show that in the world of pesticides, which have very prescriptive data requirements, this might also find some use. Next slide.

We have a, we have been working in this matrix where we would develop IATA strategies for two different endpoints. The first is DART, so it’s development and reproductive toxicity. And then RDT, and that is repeated-dose toxicity. Next slide.

And we had this, we are covering four indices. That is liver, kidney, lung and neurotoxicity and DART. And here we have testing possibilities ranging from *in silico* predictions, of course, up to very advanced models and we will hear something about that later today. So, for example for the neurotoxicity we have 2D systems in certain cell lines and then we can actually cover up to 3D human-induced protein cells. This is the battery that we have and next slide.

We built up a hypothesis, the problem formulation in the IATA, we do the testing and then we can go back again and evaluate whether the testing was adequate and whether you really did reduce the uncertainties. And there we need to also integrate biokinetics and bioavailability. Next slide. And the next.

So, in the end it was our approach is iterative and what we wanted to do would be to be able to conduct risk assessment, of course. Next slide.

So, what is the situation today with read-across? There is some experience from ECHA. The experience so far has been that it is rarely accepted by regulatory authorities and this is often, there are several reasons for this. Often it is because the read-across only based on structural and physical chemical data and there were no data provided on biological similarities. And then there was lack of evidence, so it was not possible to or they fail to demonstrate toxicokinetic and toxicodynamic similarities. There was a lack of scientific plausibility, disagreement with the hypothesis, and coupled with just lack of evidence. So, clearly there was actually quite some challenges that needed to be addressed. And this is been described in this paper I have put here to the right.

So, the question was actually how can NAM data fill the data gaps that we have seen so far. This is what we have been working with. Next slide.

So, the basis for our learnings were these four case studies. And I will just in the next few slides I will briefly introduce the case studies and also the NAMs that we used. Of course, I cannot cover all of them. There's a lot of detail. Next slide.

So, have the first case study. Where we, the aim was to predict a 90-day repeated dose toxicity study for 2-ethylbuturic acid using a read-across approach to other branched carboxylic acids.

And here we have all of the chemicals listed and the chemical in blue, 2-EBA, that is the target chemical and then we have a lot of similar chemicals here. Amongst for example the VPA and 2-EHA. These are the chemicals where we have *in vivo* data for a 90-day study, and the others we didn't. Next slide.

And we gathered all of the information and actually the effects that we focused on based on the effects that we had seen in VPA and 2-EHA was liver steatosis. And here, we have, you can barely see it but it is an AOP network indicating that the different MIEs on the left side that could lead to liver steatosis and we have several common key events and others that are more specific for the different AOPs. So, the red one, the red boxes are the MIE key events that were tested in the case study.

After the testing we applied PBPK modeling to make us able to predict actually what would be the human exposures and whether there would be any toxicity also. Next slide.

In this case study there was actually a lot of different data streams. You can imagine that with all of the testing we did, all the testing, all of the chemicals and the key events that we found, and the big challenge is actually how to combine all of these lines of evidence and conduct the uncertainty analysis. And here we applied [inaudible] theory in order to combine the evidence. This is, it is a completely new approach in risk assessment, it hasn't been done yet, it hasn't even really been considered, but at least we did show that it could provide some structured way of doing this. But this is a first step I would say.

Then had the next case study which was on another endpoint but more or less the same chemicals. It was on developmental, developmental toxicity from methyl hexanoic acid. Next slide.

Here we wanted to actually predict the whether the target compound could induce [inaudible], which valproic acid is known to do. So, here we also have a pattern of different data, some chemicals we have a lot of data, and others not. Some of that we do have *in vivo* data and others not, but we still choose to keep them in the study because it is still provides evidence and an understanding of the potencies of the chemicals and it provides more robustness in trend analysis. Next slide. I just told you few slides ago, we have different assays here that we tested the chemicals in. It was not so heavily reliant on an AOP in this case.

And here we have a matrix over the data we found. And again, for the weight of evidence and the uncertainties of that analysis, we applied different approaches. Next slide.

Which was again Dempster-Schaefer theory approach and also Bayesian automatic classification approaches and biological fingerprint classification approaches. So, we have, it was to really to try out these different methods. Next slide.

Both of the previous cases were targeted the REACH regulation, and the next two cases I am briefing you on, they were targeted pesticides. So, the first one was to identify and characterize the parkinsonian hazard liability of deguelin by an AOP-based testing and read across approach.

We only have two chemicals, so we have the target chemical, that’s deguelin, and then we have the source chemical which rotenone. Of course, this is much more difficult. If you only have, you restrict your chemicals base to only an analog approach and you are already, you are limiting yourself a lot. And there will be some uncertainty. But in this case this could be, the uncertainty could be reduced because we based the whole testing strategy on an AOP which has been OECD endorsed now so it is a very robust AOP and you can find it on the OECD website and as you can sue see in the B figure we have assays that covered the MIE and various key events and we humans have, for some of the key events we have two assays and two endpoints that we measured or predicted. The only key event that we didn't have anything on was the key event 5 here. Next slide.

And again, we did PBK modeling. So, here we didn’t do, here we followed the IATA standard approach, very qualitative uncertainty analysis. So, we didn't apply the more sophisticated method described before in the weight of evidence analysis. Next slide.

And the fourth example is where we looked at a group of pesticides. The strobilurins, and whether they could, whether you could by read-across to other strobilurins could predict whether azoxystrobin would mediate a neurotoxicity by complex three inhibition. And here you have also, this is the chemicals and I will not go into details, next slide. The thing is here, so we have, we are not in an analog approach now. Where are in a category. We assembled similar chemicals in the sense that they have this similar mode of action, pesticidal mode of action. Have a similar [inaudible],

but just by looking at the structures here, you can also see that they are, to a large extent, quite dissimilar. And in this case, we didn't have a robust AOP. We came up with a [inaudible] AOP based on the previous AOP with a complex one inhibition, this one is complex three inhibition, leading to neurotoxicity. So, there would be a lot more uncertainty in this regard. Similarly, we had different assays that would target the different key events and those you see in B. Next slide.

And here we also did the PBPK modeling. This is the sphere we are operating in, we had robust AOPs and we had more fugitive AOPs and working within an AOP network and also in a situation where we, it wasn't really explored that deeply. Next slide.

So, this is all very good. We did a lot of very clever scientists worked on this and I am not a web lab person, so I didn't have my fingers in that. But actually we did realize at some point that we, went to seek regulatory [inaudible] touch base with the regulators, and we set up the regulatory advisory board where you can see that we had various regulators from the US and from Europe, authorities, member state authorities, agencies, and so forth. And we got a very clear message that we of course had to get the science right and the actually felt that they will leave that to us. But we need to convince the regulators that we've got the science right and for that to happen we also needed to make sure that we would address specific requirements of the regulatory process. And it was actually, it was tough questions sometimes, but we also had to good learnings from it. Next slide.

So, what the regulatory advisory board came up with was that it was extremely important that we would report the case studies in a way that the regulators could recognize and understand and make sure that we would put in the information that would be needed, so they developed a template for reporting these case studies which were also based on the template from the IATA project. Next slide.

So, with this one thing that we also, this was a focus right from the beginning that we need to make sure that the approaches, the tests that we are using actually are well described and make sure that their performance would be clear to the regulators. And for this purpose, we developed a template for description of cell-based test methods. Next slide.

And it really fulfills all of the requirements of the guidance document when one, from the OECD and it guides the scientists of this new system he or she has to all of the questions that need to be answered and it gives guidance on which information and detail that would be necessary. In this, part of this is to include acceptance criteria for test elements. And this is, to describe the performance of the assays, this cannot be underestimated at all. All of the new test methods, and this is the idea of the IATA, is to apply a new approaches and the regulators are used to having a test guideline which has been validated so this is a new world and it is something that really needs to be bridged well. In this publication, you can look it up, and you can see a detailed table and also will provide a database containing examples on more than 20 cell-based tests and how they were described.

We also seek regulatory feedback on the case study of the workshop one year ago where we also had global participation from regulators in the industry and science academia of course.

And also, we submitted was these case studies to the OECD where we had the feedback and they underwent the review of the IATA project. If I should, the main topics here and which I find the most interesting things is for example the learnings we had on AOP and AOP-based testing strategies. So, one of the questions that came up very early from ourselves and the regulators is, is it necessary to test all of the key events? Do the key events, do they need to be tested with different assays?

If there are no AOPs available, how can they still work, and we found out that you have to make sure that the scope, your problem formulation is much more narrow. You need to do find it very, very well and you need to really provide a very robust scientific rationale for the testing. In regard to whether you need to test all of the key events, it was kind of, the conclusion was that if you see a consistent pattern, this would actually be adequate and maybe it is not necessary either to test the key events with different assays. It would be case-by-case, that's the thing. Next slide.

And then of course so you focus on one AOP or maybe a few AOPS or an AOP network and in the end it narrows down to a few adverse outcomes that you're actually targeting your testing, and this is of course a problem because you never know whether there would be other AOPs or mode of actions that would be relevant. It would be the question that the regulators would ask, and you need to address that somehow, in some way. So, this would be case-by-case that you would look into this, but you cannot skip this. Next slide.

So, the overall learnings that we have, you have a read across case and you have a different source chemicals and you provide the data and it is good to provide data on the reference compound, so you show that the assays you are using are actually working. This is, the regulators found that useful in their understanding. Likewise, PBPK modeling is useful. There is still some skepticism around, this is not, I am giving very general learnings here. It was quite varied what the different regulators thought about these different approaches. It is also useful to include compounds in the category which do not have *in vivo* data. This is actually quite promising because for a lot of the chemicals we are trying to assess, we would not have *in vivo* data at this moment. An analog approach can also be justified on just one compound. If it is robust and you have, you can anchor it to a recognized AOP. It is more difficult to conclude that a category is adequate because you can always find another structural similarity in the numbers. It could be 10 or 20 and where do you stop? And this is a tricky situation so far but that needs to be justified.

To some extent, and this is what the strobilurins case study showed us, that you can also perform a biological read across. You have compounds actually that were quite dissimilar but this, in the situation we had a lot of *in vivo* data because there were pesticides, but that uncertainty was mitigated by other data, *in vivo* data. And then the reporting template, everybody recognizes that it is absolutely mandatory that we report well but there is not really a consensus at this moment at what level of reporting. This is something we need to work more on. In these situations, where we have current *in vivo* models to have their limitations, this is the case for what we try

to predict the parkinsonian liability. It is much easier to waive a study based on a clear mechanistic hypothesis with the use of NAM data. This was quite promising, I think. And of course, the validity of the NAM guideline needs to be detailed a lot.

We also pointed towards other where we need the guidance from the OECD and they are already working hard on it. For example, the question what is come if you try to predict no toxicity or low toxicity how much space with there be around the negative prediction. This is a very tricky question. You can always ask for more data and more and more and more. This is reiterating again that we need more guidance on whether to test more key events in the AOP or not. And also, we, it would be nice or helpful with more guidance on how to report the cases. And on how to integrate many lines of evidence. We have tried to showcase different methods here, but this is certainly not the final solution and I think there is a lot to learn.

What we, in the end we have been trying to integrate and we came up with a read across assessment scheme or the steps that would be needed. Have published this. I will not go through it now, but you can find it in this publication. Go to the next slide and next and next.

What we actually do provide here is [inaudible]. Is it okay now? When you do a read across and you start generating data, we have provided a scheme where we show where experimental data can be used and where *in silico* tools can be used and it depends on different scenarios or whether you know the mode of action, the AOP and whether you don't, or whether you are, but you know the specific effect or there is actually the situation which we don't have a case study yet to present but where we don't actually know the adverse outcome. So, look into this publication. There is a lot of detail and explains how these types of data could be used. Next slide.

So, in summary for a successful regulatory read across you have to do some kind of testing and the thing is the regulatory authorities they often demand standard OECD guideline testing. So, when we do not provide that type of data, there is a bar that we need to reach and that is the ensure that transparency and quality of the data and document that. We had to explain very well the methodology and the justifications. And we have to integrate the hazard and kinetic data and access the limitations as such. So, it is more work. As opposed to when you have a defined approach and it is you get an Ames test, you get a result in you know how to assess it. Next slide. Next and next.

Coming up, next, so we're now working on an advisory document on read across based on data from NAMs. This is based on the learnings that we have here, so there will be a lot more detail provided in this and the four case studies that I have mentioned here, they will be published on the OECD website this fall. So, I hope you at least got a bit curious here. It was a lot of data to cover in one short presentation. Next slide.

And this is the references. Next.

And I just wanted to acknowledge all of the partners of the EU-ToxRisk Consortium and the Regulatory Advisory Board and then we have the case study leaders here,

Sylvia Escher from Germany, Bob van der Water from Leiden University, he is the coordinator of the project, Dinant Kroese from the Netherlands and Hennicke Kamp from BASF. And of course, the funder, the European Union, this would not have happened without.

That is all for me.

Ceriani: Thank you very much, Susanne, for the really interesting presentation. We are a few minutes behind schedule, and we have to move on with the presentations and now we should have Dr. Andrea Richarz on the phone. Andrea, can you hear?

Uncertainty Characterization in IATA for Chemical Safety Assessment: Overview of Available Guidance
Andrea-Nicole Richarz, ECHA, Helsinki, Finland

Richarz: Hello again. Can you hear me? I hope the phone connection is better than the Internet line. I will use my headset as well. Let's start again and sorry about the technical issues. In eight weeks of teleworking, I didn't have any problems, but that's life.

I am Andrea Richarz, I am scientific officer at the European Chemicals Agency in Helsinki. I will take up from Lidia's introduction on the IATA and discuss some specific aspects regarding the available guidance and especially uncertainty assessment. Next slide.

I have no conflict of interest and that the overview is based mostly on a project carried out by the OECD working party and the assessment has been involved. Next slide.

They are saying that the IATA complex and therefore the application guidance will be very useful. We have seen different building blocks of IATA; therefore, there are different levels and aspects which are needed. You will see that are related different levels of uncertainty and the mentioned project the OECD Working Party on Hazard Assessment has carried out the mapping of available guidance related to IATA and we will give some examples especially considering inclusion of guidance and uncertainty and then summarize some of the conclusions from the project on gaps, overlaps, divergence, and further needs. Next slide.

As we have seen in previous presentations, the IATA framework comprising many building blocks that ideally builds on a mechanistic basis such as knowledge from adverse outcome pathways and take into account the results from multiple information sources involving different types of studies that can be combined in different ways. Overall the weight of evidence assessment is performed to conclusion on potential hazard or risk which can be used to inform regulatory decision-making. Next slide.

So, we have three different layers related to IATA. The overarching concept, the IATA components, the different methods and building blocks, some cross-cutting

issues and principles, quite importantly at the end, the integration and assessment of information sources and the weight of evidence. Next slide.

For every step in the different layers of the IATA, there are some uncertainties involved. This can relate to the input data information sources, for example, data quality, methodology quality, the reliability of the data and to uncertainties of the extrapolation or interpretation of the data. The integration and assumptions made for the use of the specific methods and innovative evidence. Next slide.

These different levels of uncertainty sometimes they are compared to the [inaudible] from the basis of data, the information instructed with this further channeled and integrated to reach a conclusion at the end of the assessment. Click please.

The uncertainty can be presented in the scheme of different layers with very basic considerations at the core, that is related to the data format or reporting format, related to the methods and to the integration of the information. Next slide.

It is important to characterize all of the different uncertainties related to IATA and to transparently document these uncertainties. And last but not least it is important to communicate the uncertainties appropriately. So, how to do that? What guidance is available to help with this? This was the aim of the scoping exercise carried out by the European Commission Joint Research Centre band further developed into a WPHA project. Next slide.

Generally, there are many international initiatives and organizations discussing integrated approaches from chemical and uncertainty assessment. For example, the World Health Organization International Programme on Chemical Safety and the IATA case study project which discusses the new use case studies and summarizes learning especially of your regulatory application for different national registrations and risk assessment proposals. There is also the International Cooperation on Cosmetic Regulation, which has a Working Group on Integrated Strategies for Safety Assessment focusing on the use of new approach methodologies and this is especially driven for cosmetics for the animal testing in several countries. This working group has summarized my principles on the integration of newer methods and basis and [inaudible] is mainly the sources [inaudible] they should be characterized and documented. You will hear more about that later. Another example is the GRADE framework which comes originally from the [inaudible] of public health and has been taking forward to the [inaudible] toxicology and chemical safety assessment. It looks into the integration of evidence from the *in vivo*, *in vitro*, *in silico* studies. Similarly, there is the *In Silico* Toxicology Protocol Consortium initiative and includes over 50 members of regulatory agencies in different parts of the world and companies, academic groups and stakeholders. It is developing protocol for integrating existing *in vitro* data and *in silico* approaches and effects. They develop protocols and resemble IATA framework endpoints. They also consider confidence scores for the different information. Next slide.

From the following I will present results from the mentioned the IATA guidance project from the OECD Working Party on Hazard Assessment. The first part of the project is aimed at summarizing key concepts and terminology related to IATA. The main part of the project was compiling existing guidance. The overarching principles

of the consideration of uncertainty, the aim is to show what guidance is available and to provide either access to the results as well. Sources were mainly documents from the international organizations or national authorities. The outcome was a large Excel table with details on each of the documents mapped containing photographic information of the [inaudible] but also indicating with the practical tools, templates or checklist are included in the guidance document, and then uncertainty assessment was taken into account in the guidance. And how it was taken into account more qualitative or quantitative way. The document was led by the European Commission Joint Research Center and included Canada, Germany and the Netherlands. There are several discussions generally within the IATA case studies group and contributions of guidance for the mapping was also received a feedback from the WPHA. Next slide.

Just a note that the list of guidance doesn't constitute any recommendation over practical guidance, but it was intended to give an overview or landscape of existing guidance. And that includes about possible gaps or duplications.

Corresponding to the different layers of the IATA framework. The different layers of uncertainty with different levels of guidance needed which might be available for guidance of the overarching concept guidance related to a IATA input and method building blocks, guidance for crosscutting issues and characterization of uncertainties, and guidance for integration of the data, for example, in the weight of evidence. We are taking of this color coding of this IATA scheme to show the guidance level as examples. Next slide.

These are some examples of general guidance on uncertainty assessment that doesn't necessarily are not IATA related. WHO guidance which includes the structure tool. There are examples from national organizations, for example, the Environmental Protection Agency and European Chemicals Agency, and the most recent very comprehensive guidance and uncertainty analysis and scientific assessment from the European Food Safety Authority, which gives a very thorough overview of available methods. That includes the second document which explains more of the background. Next slide.

Looking more specifically at the IATA guidance. The OECD guidance on the reporting of defined approaches which contains a general introduction of IATA. There are principles for IATA listed and principles [inaudible] the consideration of uncertainties. In the other case studies project the template was developed for the case studies and the project and uncertainty assessment is an important part of the template. Looking at some specific IATA building blocks, for example, for read across the ECHA read across framework is all about confidence in the read across prediction and guides through a structural assessment and that also includes some assessment score of confidence. In other words, the uncertainties. And at the end there is a recent new OECD document with guiding principles for establishing a weight of evidence of chemical assessment which is emphasized the role of uncertainty assessment including the risk of bias. Next slide.

We're looking for at the example for the methodology of read across. There are several guidance documents available from several organizations and regarding the characterization and the reporting of uncertainties there have been many

discussions and suggestions in scientific literature with different types of practical tools. Some examples are shown here ranging from templates and words to decision trees scores of uncertainty. These examples for scientific discussions on issues and suggest guidance that exist on many topics related to the methods of hazard assessment and consideration of uncertainty. From these discussions usually takes some time to take up the suggestion and to be formalized into official guidance documents to be agreed on how to harmonize and endorsed international authorities Next slide.

This is about *in silico* models, quantitative structure-activity relationships, QSAR and guidance identified. There is available guidance on the OECD principles for validation of QSAR models and at a general high-level. Some more concrete guidance and harmonized practical recommendations for example for good modeling practice would be very useful and again there has been a very long discussion of the scientific literature and here an example is shown of a recent description of uncertainty, variability and bias for QSAR predictions. Next slide.

This is another example of new methodologies for which international agreement and harmonized guidance is still needed. This was 10 years ago the standardized reporting of mixed data for example. This is shown the Miami guidelines, minimal information about microarray experiments. There have been a few papers on the workshop for [inaudible] for chemical assessment which is organized by the European Center for Ecotoxicology and Toxicology of Chemicals. This paper discusses good laboratory practices for omics technologies. Actual at the OECD level the reporting framework for omics technologies is under development and this encompasses the importance of the characterization of uncertainty. Next slide.

The previous examples have shown that [inaudible] play a very important role. At the guidance of many of these aspects are available generally for example for reporting studies as seen the MIAMI of omics and there is also guidance, how to report guidance of *in vivo* experiments or how to plan them in the first place. And also important to mention the organization of the harmonized templates many performing to OECD test guidelines and the OHT 201 for intermediate effects, which can be used in the effect of toxicology studies. Regarding methodology description [inaudible] specific guidance from OECD as we mentioned here and how to report nonguideline regulatory methods focusing good on *in vitro* practice of the given. Just to say the nonguideline [inaudible] 2011 has been the basis to build the database an alternative methods of [inaudible] the URL is included on the slide. This is a nice, publicly accessible collection of new approach methods and protocols, and is a practical source to look at new methods. Regarding data reliability there are many approaches suggested in the literature. The classic reference, of course, are the [inaudible] score [inaudible] tool that helps to develop [inaudible] score. But other initiatives have been published in the meantime. The systematic review approach is increasingly been used to consistently identify relevant studies and their quality, and the risk of bias plays an important there. Next slide.

To say again that these are very basic concepts and considerations within the complex layers of the IATA framework, but they are present the basics of uncertainty and evaluation. The good study and data quality and data reliability is crucial to improve the quality of the IATA and conclusions drawn from the IATA. So, these

aspects are very important to reduce overall uncertainty and increase the confidence in the outcome of the IATA. As shown on the slides, these aspects include standardization of data performance and reporting performance, consensus on what to report from the studies and what metadata to report, such as a grade and many more information requirements, harmonization terminology is very important so everybody is talking about the same thing can be done with the harmonization of ontology, an agreement on how to evaluate and rate stydu quality. Next slide.

Looking at guidance for weight of evidence, there are many documents available from several national authorities and international organizations. Some examples are shown here. The most recent is overarching document from OECD on the guiding principles for a weight of evidence. Next slide.

We have looked so far at examples of different IATA building blocks. What guidance is actually available apart from the mentioned weight of evidence guidance. The world concept and framework of IATA. There are not so many, not as many for the detailed aspects. Will only found the OECD guidance document on reporting of defined approaches. Summarizes IATA concept and that is also for OECD guidance and how to use adverse outcome pathways in developing IATA and the IATA case study project has refined the template and discussing experience with the application of the IATA and [inaudible] the case studies. So far for this project mainly non-endpoint-specific overarching guidance was considered, but it should be noted that there also guidance documents available for the development of specific endpoints, endpoint IATA, such as skin corrosion, development and neurotoxicity, and we will hear more about that later. Next slide.

In conclusion we are seeing that the available of guidance available for IATA, the building blocks the guidance landscape is very complex and different levels of details can be very confusing for this potential user. In general, there is more guidance available on the aspect of data generation and reporting than integration of data and for the overall assessment level and they have historically grown separate by different scientific communities. For example, these *in vivo*, *in vitro*, *in silico* methods, however the distinctions, for example, between *in vitro* and *in silico*, are [inaudible] nowadays. Computational methods are not only used to generate data, to predict results, but they are also used for analyzing and interpreting extra data in general so it is all quite mixed in a complex interaction between interaction and computational methods. Overall guidance strategy is missing to put all of the different guidance into perspective. And we are seeing already there is a challenge to keep up with the pace of new emerging methods. Next slide.

In terms of guidance needed to support IATA, there is especially need for additional guidance and new approach methods, where many uses not much experience with should be used to report or show how to report new methods and also how to integrate the new methods within the IATA and how to take into account the uncertainty propagation together with the other methods used. Overall guidance and uncertainty characterization and documentation not only for [inaudible] methods but for the overall IATA level that is needed. To support the application and use more practical tools such as templates would be very helpful. Next slide.

There's some overlap and multitude of guidance documents available on the same topics such as weight of evidence, but generally there based on the same principles and their mostly complementary. But still some attention should be paid to the development evolving in this scientific community and they should be eventually harmonized. In terms of divergence there is no obvious inconsistency in the overlap. There are differences which are due to different regulatory requirements, procedures and practices. For example, either quantitative or qualitative uncertainty assessment might be preferred. And even if the same guidance is used, interpretation of the results, are the result of decisions might be different depending on the context. And the search for harmonized definitions for this project have shown that there is a different understanding and some inconsistent use of some key terms and this was most apparent for the definition of new approach methodologies. So, there is a need to harmonize. And almost last slide to show the future needs.

These have been found to be the need for harmonization of definitions of some commonly used terms, and the overarching roadmap of guidance to have us navigate to the complex guidance landscape, and overarching uncertainty assessment framework for IATA. This uncertainty assessment framework can draw on existing guidance and should reconcile evaluation frameworks for different communities and pay attention to integrating emerging the uncertainty assessment. Last but not least, as you've also seen already, the uncertainty templates are very helpful and if all of these points are taken into account, this will help to further [inaudible] the use and acceptance of IATA in the context of chemical safety assessment. Next slide.

I would like to acknowledge the project team of the OECD secretariat of the guidance method overview and everybody who contributed for the comments and discussion and review. You very much.

Ceriani: Thank you very much, Andrea. That was an excellent presentation, we are behind schedule, but we can still take a 10-minute break. Everyone please be back at 10 past 4 PM European time or 10 past 10 AM US time.

Break

Ceriani: Welcome back to the colloquium, everyone. I hope you had time to grab a cup of coffee or tea or have a quick break. And now, we start. I am pleased to welcome our next speaker, Dr. Gladys Ouédraogo. Gladys, the floor is yours.

IATA as an Opportunity for Next-Generation Risk Assessment: The Propylparaben Case Study

Gladys Ouédraogo, L'Oréal, Paris, France

Thank you, Lidia, and thank you all for inviting me to this colloquium today. I am Gladys Ouédraogo, scientific officer working at L'Oreal in the area of alternative methods to animal testing. L'Oreal is actually marketing some of the products containing some of the compounds I'll be taking about today. And L'Oreal is also funding the research led at the Cosmetics Europe science program the work I will be

presenting here is part of a European Trade Association Science Program which aims at developing nonanimal based approach for safety assessment of cosmetics.

Basically, I will walk through the practice of safety assessment of cosmetics in Europe because this is about the European context today. And the case study I will be presenting in this IATA session is mostly about the methodology and not about the compounds because these compounds are pretty data rich and they do have well-known safety opinions for safe use in humans and it's just a theoretical case to illustrate the methodology on how IATA could be applied to make safety decisions. Next slide. Next slide. A lot of typos. Two more clicks, please. Thank you.

Basically, in the past in Europe when animal testing for cosmetic ingredients were possible because [indiscernible] was based on cosmetic ingredients, the animal data we use to derive the point of departure that was used to derive safe dose in humans. One more click please. So, this was basically the process how to develop a new ingredient and to develop the safe use in humans. Next slide, please. Basically [inaudible] started before 2013 but it fully entered in place since 2013. So, we no longer can use new animal data but [inaudible] data is actually to be used. And we can have a couple more clicks on the slide. So basically, what we have after you perform the human safety assessment approaches, toxicological concern, read across to kind of derive the point of departure and use it to derive the safe dose and develop new ingredients, basically old approaches haven't changed this much. Now that we kind of use the NAMs, we can have one more click, we use the NAMs to build confidence in the way we do the safety assessment. So, another click, please. So, the NAMs are the way to help us build confidence in the way we perform the safety assessment and I would like to introduce the concept of next-generational read-across, which is a [inaudible] risk assessment approach [inaudible] new approach methodology to play an important role. Because now that we are facing a situation where we have data for compounds, how to we make sure we have the information we need to make the risk assessment? That's where the new approach methods to play an important role in this approach.

Where we are aiming for the future is to be able to rely mostly on these nonanimal-based approaches to make the decisions for the safety assessment. Next slide. This is part of the exchange where we would like to rely more on the NAMs to make the safety decisions. So, it's moving from something based mostly on observation in the apical endpoints on the animal to something that more relies on mechanistic understanding because the NAMs would be digging into the mechanistic understanding, so to move to this nonanimal-based approach is actually to rely on the NAMs to make the decisions. Next slide.

The traditional read-across approach is actually basically one with structure. Most of the information is derived from the structure, to be able to derive a point of departure. So, it's mostly have an assessment, while the next-generation read-across is mostly, you know, we still rely on properties related to structure, but not only we use the NAMs kind of to help in identifying analogues and also to predict some internal exposure and use that actually to make safety decisions. We go up to the full risk assessment. We don't just stop at the hazard assessment level.

So, here is an IATA that has been published by the OECD IATA case studies that we have used for this case study, and this was conceived in a European Framework Project called Seurat-1. And it's impressive because it's tiered so it's flexible and it's actually designed to help structure the knowledge and the data in a way to help and have enough information to make decisions. There are different [inaudible] you can see on the righthand side, TTC, read across, internal TTC, or ab initio, so different safety could be made based on that. And actually, it's work framework actually for risk assessment, which is exposure led, and hypothesis driven based whenever possible on mechanistic understanding. So, that's the framework I will be using across this case study. Next slide.

These are clicks to show that we will be talking about read across. I would like to think the previous speakers because they actually facilitated some things for me here. From previous experience, if we go from left to right, grouping was mostly based on the structure and actually we have already observed that the NAMs could really add value in terms of strengthening the similarity and at that time in 2016 we observed that most of the uncertainty the read-across was stemming from the toxicokinetic side. Basically, the similarity in chemistry is not enough to sustain the full read across prediction. And we have seen from Susanne's presentation that this has evolved within the EU ToxRisk project to dig further into the toxicodynamic side with different test AOPs, what we know about defined AOPs, the key events that could be tested to strengthen the toxicodynamic data and also toxicokinetic to help derive the part of departure that could be used for the risk assessment. Basically, when mechanistic knowledge was added into this project, and within the cosmetic, you know, the Cosmetics Europe [inaudible] LRSS program, we are aiming at pushing further to next-generational read across where we [inaudible] previous deliverables but there we are kind of using the NAMs also to predict some internal exposure estimates and also relative potency to use that to make human decisions actually and to support the safety assessment. We can move to the next slide. Next slide, again.

The challenges we are facing here are how do we build confidence in using the NAMs to support, to inform, this type of read across? There are different questions we are facing here. How can the NAMs actually help strengthen the identification of the analogs? How can they inform the similarities/differences when we're talking about toxicokinetic properties, toxicodynamic properties, and what are the roles of NAMs in the safety assessment actually to really define a margin of safety? And how do we assess confidence, we heard Andrea's talk about the different guidances out there, how do we actually assess that confidence in overall read across?

For this case study, the problem formulation is key for IATAs in general. In this case, we are using a cosmetic ingredient, a preservative, propylparaben, which is actually used in civil sectors but in food, in agrochemical, and it is widespread in cosmetics and is used via the dermal route. The key here is whether we could use this preservative as a maximum load, which is .19%, could we use it safely in a body lotion used twice a day? Since it's a preservative, we have many sources. We did consider aggregated exposure. We pretended that we did not have reproductive toxicity information on this kind of compound. So, for the dermal purpose, we tried to see how we could assess the safety in terms of reproductive toxicity, it was repeated-dose toxicity in general, including reproductive toxicity here that we are

facing. Of course, when we're doing the safety assessment of cosmetics, we need to assess for everything, including acute issues, but this wasn't actually the topic of this theoretical case today. We were actually trying to address the repeated toxicity endpoints here. And the point is, for these compounds, the parabens actually are based on the existing *in vivo* studies, we don't know about specific critical issues based on the animal studies. So, here we don't have the APOs like Susanne presented from the EU ToxRisk initiative. So, here we are facing what we call low toxicity of compounds.

So, once we know about the problems, can we safely use preservatives in cosmetics at the maximum concentration load? And we identify the structure. And then we knew from the exposure maximum load concentration that we could not exceed [inaudible] so next stop was to move to read across to try and find analogs. So, 58 analogs were identified, setting the threshold of structural similarity at 70%. We have identified three closely related analogs that have *in vivo* data. It wasn't the only criteria that we use to define the number of analogs we were using here. We used all the properties related to reactivity, the biotransformation, in these properties as well. We ended up with these closely related analogs which also yielded the same metabolite hydroxybenzoic acid before. We tried and worked with that. Excuse me. It's not working. I'm sorry. Basically, these are really three compounds that's similar because we collected the phys-chem properties of these compounds. So, it was basically a category because we kind of observed a trend like increasing LogT with increasing [inaudible] length, and we decided to actually work with compounds that have the highest Tanimoto coefficient related to our target compound, which was [inaudible] read across. We can move to the next slide.

So, basically, looking at these categories we could see that these are all actually penetrating the skin. And they were pretty much cleared readily, the esters were readily cleared to a metabolite, the 4-hydroxybenzoic acid for all of them, actually, by the carboxylesterases that were present in the skin. And liver cells experienced the same thing, actually. In liver, the clearance was most extensive done in skin, but [inaudible] compounds still had similar behavior in these properties. We can move to the next slide.

So, we looked into existing data, we did some *in silico* screening, and based on [inaudible] wea estrogenicity, estrogen receptor binding. And from the existing ToxCast data as well, they had actually not many hits, and number of hits from ToxCast assays tend to increase with increasing [inaudible] length confirming still the category here some trend in terms of potency. And in addition to that, we have generated some extra data, the transcriptomics data, to look into this category and it confirmed the category was the trend. So, most [inaudible] observed when we increased [inaudible] length, and basically, has the same profile hitting estrogen receptor pathway basically. And fewer genes actually for the hydroxybenzoic acid, and in terms of ToxCast no hits, no hits would have warning for the acid, showing that if there were something here, it was triggered by the parent comound, by the esters, basically. We can move to the next slide.

So, we decided to [inaudible] the compounds based on the ToxCast estrogen activity here, so we can see that based on the numbers here, they are weak estrogens in terms of activity, because compared to [inaudible] we can see that different activity is

low. And this is a model developed by the US EPA. It's one of the defined approaches on estrogen we heard about previously by Lidia. So, butylparaben was right at the most potent, so it attributed a potency of 1 and [inaudible] butylparaben. So, the second was our target compound, propylparaben. Next slide.

So, for us, since we are doing a safety assessment, the exposure component is very important. So, we have considerable use of propylparaben in cosmetics. You have to consider coming from all the sectors, and we considered the deterministic, theoretical value where propylparaben would be a conservative preservative used in all cosmetic products used by the consumer in a maximum concentration, which is very conservative because it's very unlikely. In this case we ended up with an estimate of .48 milligram per kilogram per day of amount of paraben that was applied in the whole-body lotion twice a day. The 17.5 g/day you see is actually the total amount of cosmetic product that can be applied for the whole body for the day by costumer, the maximum volume.

These were then used to derive an internal exposure value here. We use the ADME properties. We have measured actually make use this model using a PBPK model here to estimate the exposure. And what we did is on one side based on human exposure and we also actually use for that compound and things that were related to number of and we actually had a study that we know of [inaudible] we had a study with NOEL that we used to convert it to an internal concentration using the PBPK model. And the rat study was a study on neonatal rats by subcutaneous route, so the PBPK model was already published. It's the model published in 2015, and it was at that time adding subcutaneous compartment to be able to do the simulation for the rat study here. In this case, internal concentration was, the max basically is 20 nanomolar that was estimated from that. We can move to the next slide.

And once we have one side actually the external exposure from humans and the point of departure from the animal study from our source compound that we used to read-across our target compounds, as we moved actually to calculate the margin of exposure. And usually for cosmetic ingredients, we used to calculate a margin but of external exposure, which is actually the ration of the point of departure measured from the animal study divided by the human exposure. That's external dose. But in this case, we pretended that we did not have that. We moved to do the calculation the internal exposure, and that we used to calculate a margin of internal exposure.

And when we are doing so, since we are dealing with internal exposure, we are using chemicals-specific information for the chemical, for the ingredients. So, we don't need to account for interspecies extrapolation factor or uncertainty factor. So, a margin of safety of 25 would be equivalent to external margin of safety of 100 and more. So, we can move to the next slide.

And when we apply this equation, we end up here with the margin of 290. Which is far more than the 25 that we need to make sure we are on the safe side. Basically, this shows here, basing the reasoning on the NAMs, using the NAMs to sustain the relative potency, to calculate the internal exposure estimates and using that to do the margin of safety calculation, we have sufficient conservative margin of safety, which is in line with the known position on this compound, actually. We can move to the next slide, please.

So, this is actually uncertainty. So, uncertainty needs to be assessed for every piece of information we need to perform the assessment. I won't go into detail. But what you can see, the ++, +-, they all mean, the ++ means that we are overconservative in terms of assumption. +- means we don't really have an impact in terms of weighing, being more or less conservative. So, basically, overall this assessment here is on the conservative side. We don't really have an impact in terms of, so this material and it is protective for human health. Some of the pieces you can see in the middle column here were used either for weight of evidence or in the read-across, or also in the risk assessment assay to drive the margin of safety. We can move to the next slide.

So, talking about the confidence assessment of this read-across, actually informed by NAMs, actually there are different questions that we wanted to go through to make sure that the assessment was robust. We can move to the next slide.

This is derived from a paper by Schultz and Andrea as we heard and Mark Cronin, and we have adapted it based on an event we had two years ago with the European Scientific Committee for Consumer Safety here for read-across. And we had this topic and we decided take up this question with them. So, that is where we kind of decided to also use these questions for the next-generation read-across because it made sense. We can move to the next slide to see what the answers were in the specific case.

The type of category that we formed here is one too many because we had one target and several source compounds, and the uncertainty was low because we had a compound with the source compound that all yielded the same metabolites pretty widely, although we observed the trend within the category, *in silico*, *in vitro*, and validity information all pointed to similar behavior for all these compounds. And how well was [indiscernible]. It pretty much was okay. As we told you based on the legacy *in vivo* data, these compounds do not have a defined target organ or specific thing happening. The only thing that was reported was estrogenicity based on the uterotrophic assays and some of the studies, and that is where we have actually built on this rationale to look into the estrogenicity with the *in silico*, *in vitro* assays to derive also the potency to use that for the safety decision. And the rationale that was used to select the NAM was this exact rationale actually based on analyzing, observing the results from all these pieces of evidence actually to make this rationale, so the uncertainty was pretty low to medium there. And the mechanism of action to support the assessment here, it's overall on the estrogen receptor pathway, and here it was low to medium because all the evidence we had from *in vivo* legacy, *in silico*, *in vitro* pointed to the same thing here as well. How were those similarities defined and assessed? They were actually converted readily by esterases to the same metabolite. And all the properties used to characterize the compounds were all consistent as well. So, the uncertainty was low to medium. We can move to the next slide.

The uncertainties on the toxicological data here were, for the source compound we used a study that was based on neonatal rats by the subcutaneous route. And we didn't have dose-response, so it was [inaudible] for this reason, so it was very conservative point of departure, so medium uncertainty based on that. And how were the NAMs applied and how did they assist in reducing uncertainty? In several ways

here because in terms of PK, it was helpful actually understand metabolism and to understand plasma protein binding and also to help do the PBPK modeling and derive the concentration estimate internal estimate there. And also, to derive the relative potency that was used for that for the marginal safety calculations. The NAMs were useful in this case study. And the over uncertainty will put that low to medium. And the trends in these cases, because one thing we use in IATA, the workflow that helps structure the process, organize the data, and integrate the data in the end as well. We considered exposure, realistic exposure here. So, this was very important, and some of the limitations we already heard about, we had the limitation related to the PBPK modeling estimates here, and the verification because we didn't have TK data from the rat study either. We did not do some *in vitro* biokinetics to actually see whether it was better to use something else than the nominal concentrations. We did not do that. One thing we heard about that is how we report the information. That's still a challenge. We moved to the next slide. Next slide.

Basically, what can we actually have when we look at the IATA, we use in this case study. We had consideration of exposure, external, internal exposure of *in vitro* biokinetics, we didn't dig into that, but this could be valuable to refine the exposure. And in terms of mode of action we actually look at different things, *in silico* tools, structure activity relationships, the kinetics, the bioavailability, we did some *in vitro* testing using human cells to do that, some targeted testing like estrogenicity, digging into these assays there, and the transcriptomic there, and all were key pieces of evidence to confirm that the hypothesis around using this estrogen activity as a proxy for biological reactivity there to kind of help inform this read-across case were pretty strong. And for the risk assessment, we were able to calculate a margin of internal exposure and also to assess the uncertainty. And then we can move to the next slide.

So, just to summarize. Read-across, when we use the chemical, the structural properties, it's ok that it can have limitations and using the NAMs made the assessment more robust. We were able to test the hypothesis, testing some of the things to kind of help strengthen the reasoning on this read-across case, looking into the TK and TD to confirm similarity across this category with a trend, a potency trend here. This approach, this kind of next-generation read-across, is pretty incremental. It doesn't mean that we need to step away from traditional read-across. Whenever we're able to make the decision with the traditional read-across, it's still ok. But in another case, that's where we need to make use of NAMs to see how it can help with your uncertainty and hopefully make safe decisions in the end. And right now, we're trusting an overarching framework to disquiet that, and also we believe that it can some guidances we've heard about, the wealth of guidances that is out there, but still, we still actually need practical guidances to put that into practice. And that's what we heard from the Scientific Committee for Consumer Safety here in Europe when we presented there. And this is something, and I'd like actually to call out here that we need to do in cooperation, in collaboration. It's not just a one-way initiative here. The industry is working. We are not walking alone. The Science Program of Cosmetics Europe is led with different partnerships with academia, with contract research labs, and other sectors. And we are trying to dialogue with regulators and is important to have them on board and quite early. Not in the end when we have some things that are finalized. We need to have a dialogue because confidence-building is

actually what we need here to translate this into reality. We can move to the next slide.

So, several people contributed to this work, I'm talking on behalf of all of them, and I will like to thank you for Catherine Mahony, who was very helpful in drafting these slides as well, and all of them for fruitful discussion. And then you have some references, it's not comprehensive. But I'd like to acknowledge the recent paper by the US FDA pointing to the consideration of NAMs [inaudible]. At ease. You can hear some background noise. I'll end here pointing to the recent paper by the US FDA pointing to the consideration of NAMs for nonchemical studies. I would like to thank you all for your attention.

Ceriani: Thank you very much, Gladys, for this wonderful presentation. We are perfectly on time. There are a few questions in the chat box, but I'd prefer to discuss them during the roundtable discussion so that we can stay on schedule.

Global Harmonization Efforts for Skin Sensitization IATA
Nicole Kleinstreuer, NICEATM, Durham, NC

Ceriani: Now, we can move on to our next speaker. I'm really pleased to welcome Dr. Nicole Kleinstreuer. I was her first address that so we can stay on schedule. Now we can move onto the next speaker. And she will give us a presentation on Global Harmonization Efforts for Skin Sensitization IATA. Nicole, the floor is yours.

Nicole Kleinstreuer Thank you very much, Lidia. I trust that you can hear me well. If not, I will be notified in the chat window shortly. So, thank you to the organizers for giving me the opportunity to present our work today. And thank you to the other speakers, I think we have seen really excellent presentations and case studies so far. I am looking forward to the rest of the session and the roundtable discussion to follow.

I have no conflict of interest to declare.

This is an outline of what I will cover today. I am going to introduce or refresh for some of you the role of the Interagency Coordinating Committee for the Validation of Alternative Methods and then the National Strategic Roadmap that we published two years ago and then I will talk through some of the concepts that have been introduced earlier by Lidia, and I will build off of what she said about the adverse outcome framework and how that feed into develop of integrated approaches to testing and assessment in the context of work being done at the OECD to provide guidance and ultimately to create internationally harmonized test guidelines for defined approaches. And then I will discuss skin sensitization as a case that he case study and then go over some of the details for defined approaches for hazard and potency characterization and then discuss some ongoing progress towards regulatory implementation both here in the US and globally. Next slide, please.

The Interagency Coordinating Committee for the Validation of Alternative Methods was established by a congressional authorization act in 2000. This year we are delighted to be celebrating the 20th anniversary. ICCVAM is comprised of 16 federal agencies that require or consider chemical safety testing data, and you can see them

listed on the slide here. They are evenly divided between regulatory agencies and research agencies, including a number of institutes that are part of the National Institutes of Health. Other participants in ICCVAM include the Tox21 consortium and we have already heard a fair amount about Tox21 and the ToxCast high-throughput screening efforts and how those are informing things like read-across and NAM case studies. My group, NICEATM, is the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, and we provide scientific and administrative support to ICCVAM. So, all the work that NICEATM does is really driven by the priority of those ICCVAM member agencies. In particular our regular partners like FDA and EPA. These are the agencies that tend to rely most heavily on traditional toxicology data and have incredibly active programs in moving towards more human relevance and alternatives to chemical safety testing.

In 2018 we published this strategic roadmap for establishing new approaches to evaluate the safety of chemicals and medical products in the US. If you have not read, this I would encourage you to follow this link and download; it is only 13 pages and really provides a very nice overview of the shifting paradigm toward how we are trying to help end users and stakeholders in the process guide the development of new tools and alternative approaches needed to support the decision context that they deal with. So, the concepts that are covered in the roadmap are nicely exemplified by this circular diagram on the front cover that shows a feedback loop between the users and those that are responsible for establishing confidence and then new technology developers. So, the objective of this roadmap is really to foster the use of efficient and flexible and very robust practices to establish confidence in new methods that will then really encourage the adoption and use of those methods by federal agencies and regulated industries. This is all driven by the concepts of protecting public health and improving human relevance. You will see that these high-level objectives are paired with detailed implementation plans that are available on NICEATM's website and are continuously updated and are tracked and publicly reported and are targeted towards specific toxicological endpoints, such as acute systemic toxicity, or in the case of today's example, skin sensitization.

So, we have already seen a version of this image. So, I won't belabor the point by talking through it. I think it is worth revisiting and understanding how the adverse outcome pathway framework is really being utilized to inform integrated approaches to testing and assessment. And I'm grateful to my colleague Patience Browne at the OECD for providing this figure for my use. As Lidia articulated earlier in our session, the initial problem and initial weight of evidence assessment relies on the existence of information on the chemical or the problem at hand. And if that information is not adequate for decision-making, then the additional information is generated. And more and more, the generation of this additional information is really being driven by the availability of new approach methodologies or NAMs that are mapped to key events or molecular initiating events in an adverse outcome pathway framework. So, the development an AOP for the endpoint of concern is really critical for being able to identify test methods and testing strategies that will allow you to generate relevant information that will ultimately give you a sufficient amount of evidence to be able to inform your decision-making process. Next slide, please.

A lot of these concepts are very nicely explained in guidance documents that are available from OECD and there are two guidance documents. GD255 and GD265

that talk about this with respect specifically to the skin sensitization, so guidance documents on the reporting of defined approaches to be used within IATAs, and also on the reporting of IATAs for skin sensitization in particular. And then a broader guidance document number 260 discusses the use of adverse outcome pathways in developing IATAs. These are all incredibly helpful resources for understanding how the concepts are being really implemented in practice. Next slide, please.

This table intends to delineate the differences between IATA and defined approaches. IATA are really fit for the purpose at hand, so they are tailored in response to a particular problem formulation and they really incorporate a large degree of expert judgment to interpret and combine and understand all of the data that is available. Versus a defined approach could be thought of as more like a standardized test method, so they have defined information sources, a fixed data interpretation procedure, and there is really no subjectivity and no room for expert judgment. And that inherently means that inflexibility makes them much more suitable to international harmonization and ultimately towards the development of guidelines for defined approaches that would be applicable under the mutual acceptance of data clause. OECD currently have the first proposal for that defined approach, and that's defined approaches to assessing skin sensitization potential that would be covered under mutual acceptance of data. At the moment any data that is generated under an OECD test guideline is covered under that mutual acceptance of data which avoids a number of redundant testing across all of the OECD member countries.

Several years ago, the International Cooperation on Alternative Test Methods held the workshop to discuss the concept of defined approaches and how they could best be evaluated in a standardized and a rigorous way to establish scientific confidence in defined approaches for regulatory use. This is the position paper that was put out and the end of 2017. It discusses the standardization of defined approaches for skin sensitive in particular but outlines an evaluation framework that is generalizable for any defined approach and not just specific to skin sensitization. The categories for this evaluation framework are the structure of the defined approach, so what are the information sources and what types of information to they provide? What is irrelevance in term of the mechanistic coverage of an adverse outcome pathway or the particular toxicity endpoint at hand? What is the predictive capacity of the defined approach? Similar how you would validate an individual test method by looking at the performance compared to reference data, taking the same approach with the DA. What is the reliability and terms of the reproducibility of the defined approach and its information sources? How to characterize the applicability domain of a DA in terms of the chemical space that has been tested but also technical limitations of individual sources? What is the degree of complexity in terms of the data interpretation procedure? Is it a simple rule-based DA or does it rely upon more complex algorithms like machine learning? And what is the transparency in terms of the open-source nature or availability of each of the elements of the defined approach. Next slide.

So, shifting now towards a specific example of skin sensitization, this is the adverse outcome pathway framework and as you click through this if you click four times, we will show an animation that demonstrates the biology of the adverse outcome pathway for skin sensitization as initiated by covalent binding with skin protein. That

covalent interaction is molecular initiating event, followed by the response of keratinocytes and activation of inflammatory cytokines and induction of cytoprotective genes, followed by dendritic cells. mobilizing and inducing inflammatory cytokines and [inaudible] compatibility complexes which, are then presented by the dendritic cells to the T cells which are activated and began to proliferate. All of that leading to the adverse outcome, which is inflammatory response that worsens upon repeat exposure to the allergen. Next slide, please.

Taking those key events of the AOP and mapping test methods to them, you can see that the traditional animal-based methods measure those downstream apical outcomes. So, with the guinea pig maximization test, you're measuring the actual adverse outcome in the organism. With the local lymph node assay test, that uses a mouse model, you're measuring that key event 4. But now we have a number of *in vitro* or *in chemico* test methods that have been included in OECD test guidelines that map to the first three key events of this adverse outcome pathway. And you can see some of the specific test methods that are incorporated in the key event guidelines listed on the slide here. And none of these individual methods are approved for use as a standalone replacement for the animal test, but the thought is that if you integrate multiple of these methods then you can provide enough coverage of the biology of the adverse outcome pathway that you can achieve the performance equivalent to or possibly superior to that of animal test. That has been the basis for the tremendous amount of work that has been done over the last decade to develop integrated testing strategies and defined approaches that combine these different test methods and other aspects such as chemical structure, structural features and properties, into the defined approaches that are intended to replace animal test. Next slide, please.

About five years ago a project was initiated in collaboration between NICEATM and Cosmetics Europe to analyze the available nonanimal defined approached that had been submitted as case studies to the OECD and here we look at around 128 chemicals that had generated data in the *in vitro* assays that were mapped to the three key events that I showed you on the previous slide. The specific assays that were used here were the direct peptide reactivity assay, the keratinosens assay, and the HCLAT assay, so data was either curated or generated for all 128 chemicals. Also, these chemicals were chosen because they had existing reference data not only in the animal test, so the LLNA assay, but also had human clinical data with them. So, there were tremendous efforts to assemble the human repeat insult patch test data and other types of information such as prevalence of the sensitization, the population and use case to provide a weight of evidence assessment for the true skin sensitization potency in human for each of these chemicals. It was a very unique and rich data set in that we had both animal and human data to anchor the predictive performance of the defined approaches and base our comparisons on. We also generated *in silico* computer predictions and incorporated chemical structural features and properties where they were incorporated as information sources in specific designed approaches.

So, wherever possible we took a transparent approach in reproducing these defined approaches. So, in some cases they were constructed using things like proprietary software. So, wherever we could we reproduced that using open source code. And then evaluated the performance of many different defined approaches against this

LLNA and human reference data. With the downstream long-term objective ultimately trying to develop a harmonized guideline for defined approaches to skin sensitization. You can see those here and that summarizes the database analysis of the defined space.

So, this project incorporates a large number of types of defined approaches. And they range from very simple like the two out of three. If two of them are negative it is classified as a nonsensitizer. Two more complex approaches that would rely on things like support network machine models or Bayesian networks artificial and neural networks so like machine learning driven defined approaches. Click once.

The four that I will be talking about in the next few slides are those that are currently being used in the OECD defined approach guideline project in terms of those that are circled in blue. And they were under consideration for the first round and then I will talk a little bit about that that provides a more quantitative prediction that is currently being used and a draft risk assessment that the EPA is doing. Next slide, please.

So, these are three examples of defined approaches that provide hazard and/or potency categorization. These are by no means the only defined approaches that are able to provide these and they are just the first batch in the pipeline that was submitted many years ago as case studies at the OECD and then that group agreed would be a good starting point because they are simple rule-based defined approaches that are easy to interpret. So, as I mentioned, they are based on concordance of multiple test methods and this testing strategy begins with the third key event. And based on the results of that compound can be classified as a strong or weak sensitizer. If it is negative and based on the results it can be classified as [inaudible] or potentially negative and the and a graded testing strategy also submitted by that incorporates both that and also [inaudible] prediction.

So, the original version of that incorporated the Derek software. And this has been updated to have an alternate version where we have used that toolbox to provide that prediction. And based on the quantitative results *in vitro* test method and then existence of alert or not and that method a total score is provided for the compound and then the bins are there are the compound is a strong or weak or not classified. Next slide, please.

This model is an artificial model. This was published in 2015 and it was built using a black box proprietary software, so we reproduce this model using the open source and it just shows you the reproduction of one of the figures to ensure that we were incorporating all the correct parameters and reproduce the model and there are actual multiple different configurations of this that rely on either that in the deeper and the NCLAT or the deeper and the HCLAT and the keratinosens to provide a continuous EC3 prediction. It is the effective concentration that result in an index of three in the mouse model. That is the quantitative potency value that you would obtain if you were having to run the LLNA or the animal test. So, this is eventually one that can reproduce that value that you would get. That prediction can of course also be translated into potency categories as well. Next slide, please.

But these four approaches and a number of defined approaches that I show on a few slides [inaudible] we evaluated their performance against the existing reference data. And as I mentioned we had a very robust reference data. And so, we were able to use this against the human data as a benchmark for comparison to see how well the defined approaches were. And rather being confined by the [inaudible]. With hazard performance ranges from 70 to 80 percent and potencies in three potency categories, so random would be 33% percent. Potency performance of 55 to 60 percent. They are driven by the fact that there are a number of different meta-analyses that have been published that we had tried to incorporate. And our analysis in the [inaudible] we found that the defined approach depending on which we were looking at performs quite well against the human clinical patch test data. So, the hazard predictive performance ranged from 77 to 85 percent, and the three-class potency prediction ranged from 63 to 69 percent. Next like, please.

So, the take-home message is actually that most of that testing strategy so essentially all of those approaches to find better or perform better than the animal test at predicting that and potency. And if we ran that thought experience where we did not have that human data as a basis for comparison [inaudible] if we use the reaper's availability [inaudible]. We found that the defined approaches were equivalent to the animal test app heard acting themselves. Next slide, please.

This work has been the basis of a number of really exciting regulatory policy developments, so 2018, the US EPA released this draft interim science policy announcing the use of alternative approaches for skin sensitization as a replacement for the animal test. This is the joint policy. And two of the approaches that I have already discussed in the presentation are included in this. Two out of three and the sequential testing strategy mapping to key events 3 and 1 and the policy includes all of the assays that are covered by the respective key event-based OECD test guidelines. For example, direct peptide reactivity assay is what we use in the Cosmetics Europe project, but they is now also the amino acid direct reactivity assay, the ADRA is also incorporated and it would be considered a suitable drop in replacement in the defined approaches that EPA is accepting. This policy will continue to be updated as more defined approaches are accepted and validated under that work.

Something that recently happened in the last two weeks, it is not on the slides because it is very recent. EPA has released a draft risk assessment on a group of compounds called isothiazolinones. And this risk assessment relies upon the artificial neural network model that I mentioned a few slides ago. It uses the quantitative EC3 potency prediction from that model as a basis for deriving that points of departure for quantitative risk assessment for skin sensitization. I think this is really the first time a risk assessment from EPA has been fully based on alternative approaches for this endpoint. Next slide.

Our partners at the FDA have also released some forward-thinking guidance on the nonclinical safety evaluation of the immunotoxic potential of drugs and biologics. There is language around the fact that FDA will consider a batter of *in silico*, *in chemico*, and *in vitro* studies, so that is a synonym for defined approaches shown to predict skin sensitization with an accuracy similar to *in vivo* methods.

In the last few minutes, I'll talk through the progress that's ongoing with the OECD skin sensitization guideline project. This was one of the intended products of the NICEATM Cosmetics Europe collaboration to try to develop the first internationally harmonized guideline for defined approaches. And substitute the need for animal testing for skin sensitization. The objective is that the defined approach would be amenable to the mutual acceptance of data agreement and would be considered on equivalent footing to the anima test guideline. For the first stage, we are looking at the regulatory requirements for hazard characterization. Discriminating sensitizers and nonsensitizers and trying to discriminate strong from moderate and weak sensitizers and those that are not classified, so essentially the GHS potency categories. The next stage will focus on defined approaches that address the regulatory needs of quantitative risk assessment, similar to how the EPA has begun to utilize defined approaches in that regard.

We have an expert group of 68 members that cover not only national coordinators form the OECD test guideline program but also the test method developers, experts on validation of alternative approaches, regulatory authorities, animal welfare and other NGO representatives, so we have a diverse and robust group of experts advising us. At the moment, we are really in the end stages of this three-year process where we are finalizing the reference data, this is an expanded dataset, so we have a large list of reference chemicals with LLNA and human data and we hope that these databases will become valuable resources for the scientific community and future defined approach developers. And then we are looking at some scientific issues around how best to characterize things like confidence, uncertainty in the defined approaches and applicability domain. We received some excellent feedback at last month's meeting and next month we have a virtual meeting to discuss these issues and work toward finalizing the guideline. Our goal is to have a final defined approach guideline submitted. For written approval by the WNT by the end of this year.

The last thing is how the NTP and the Immunotox testing group is supporting expanding and assessing the coverage of chemical space of these *in vitro* methods and the defined approaches that rely upon those *in vitro* methods to understand whether they could be more broadly applicable beyond the monoconstituent substances and be potentially useful for assessing sensitization potential for substances like agrochemical formulations or personal care products. So, in working with our ICCVAM agency partners and our ICATM partners, we received nominations from multiple stakeholders, so we have around 125 substances that have been tested so far and we're nearing the finish line of this project, so we're about 90% done with the testing and we're starting to get all the finalized data and get all the reports together. I'm excited about the aspects of this project, to be able to understand and better characterize the applicability domain of these methods and the defined approaches.

There were a number of people that contributed to this work. I'm honored to be speaking on behalf of all of them. I'd like to thank the NICEATM group. None of this work would be possible without them. Thank you very much.

Ceriani: Thank you very much, Nicole, for this great presentation. All the questions will be addressed in the roundtable so that now we can move on to our last speaker,

Dr. Reyk Horland. Reyk is going to give us a presentation on Applications of New Multi-Organ-Chip Tools for Toxicity Assessment. Reyk, the floor is yours.

Applications of New Multi-Organ-Chip Tools for Toxicity Assessment
Reyk Horland, TissUse GmbH, Berlin, Germany

Thank you, Lidia. Welcome, everybody. It's a pleasure to be part of this webinar. What I'd like to do is give you an overview of these upcoming tools called microphysiological systems or organ-on-a-chip tools and give you an introduction on how these systems could be used with regards to risk assessment.

Of course, as I work for TissUse and it is manufacturing and commercializing one of these systems, there is a certain conflict of interest here.

As I mentioned, I would like to start my talk with an overview on what are microphysiological systems, what kinds of microphysiological systems are out there, there are a lot of systems out there. Then move on to introduce the system that we have, which is focused on multi-organ-chip systems so systems which are capable of interconnecting different organs to study organ interaction and organ communication. Then give you an introduction to the basics of the skin liver model, combination of a skin model on our platform and how we use that in risk assessment. Then moving to the ongoing Cosmetics Europe case study which we are doing to evaluate the use of our system in risk assessment.

Starting with what are organ-on-a-chip or microphysiological systems? I just picked a fire from an overview paper on the landscape of microphysiological systems and here the authors used two parameters: lots of systems that try to emulate the function of a single organ on a microfluidic device and then there are systems that try to emulate several organ functions in a common circulatory system. And the second parameter is what kind of hardware is used. There are models which use [inaudible] with some added microfluidic component into this plate-based system in order to generate some fluid flow, and then there are so-called chip-based systems, in this case chip has nothing to do with a computer chip. This was derived from [inaudible] systems which are microfluidic devices used in the clinics and that term stuck and was transferred, so any kind of more complex microfluidic device is dubbed a chip device. Using these parameters, you can divide all the available systems into certain categories, going from plate-based single-organ system to chip-based single-organ system to plate-based multi-organ system to chip-based multi-organ systems. And the in the future, people and TissUse belongs to this group, working on systems where everyone to have some kind of minimal essential organism on a chip ina combined circulatory system.

Now, regardless of which system you are looking at, all have two major components. First, the [inaudible]. So, how is the microfluidic device filled up, how is it structures, what kind of cells, tissues, biopsies can it accommodate? In our case, we have a couple of different chip systems available. They are generically branded because the number and name always implies how many different organ models you can combine in a common circulatory system with each other. Meaning in a chip 2 you can combine two organ models and study interaction. Which doesn't necessarily

mean that you can't use a single organ in these chips. It's just the maximum number here is depicted in the name. The second big part is the biology. So, what can you implement in these systems?

If you see here, it's that we have a quite broad portfolio available. We have not yet to reinvent the wheel, so especially also this is true for the skin model. If we see some best-in-class skin models or in some cases OECD-validated skin models, the system in our hands is capable of using these models in a kind of pluck and play fashion. We also have models that are internally developed and established. A lot of people are working with their own models, which they have developed in their labs and our systems and you can also work with biopsy models in the systems. In a way, the system is very flexible when it comes to the setup of the organ models and the tissues in the system. Most of the organ models we are using are three-dimensional models. They are [inaudible] in the system. There are only two for the current kidney model. The other models are three-dimensional models which not only incorporate [inaudible] incorporate also other important cell types for these organisms if you go with the liver. The liver not only includes hepatocytes but also [inaudible].

So, in order to study this organ interaction and communication, we have decided to work with an on-chip pump. And the function of this pump you see here, this pump basically works a bit like the human heart, just with three chambers instead of four. On top of these chambers are flexible membranes and by [inaudible] these membranes with pressurized air and vacuum, you generate a recirculating flow in the system. So, [inaudible] is recirculated in the circulation, and you have in this case the chip 2, you have the first organ compartment here, and the second organ compartment here, and the media is transported continuously form the first organ compartment to the second one and then all the way to the pump again and then it's recirculated. What you can also do for the skin and lung model and area model, you can use [inaudible] so they stay 100 microliters above the fluid flow, which also means that you can cultivate models on the air-liquid interface, which is important for the skin or the lung.

We have the biology and the heart rate available, the next big step is to say how can you build an assay on these platforms. And I already introduced you to the two major components such as the heartfelt which would qualify for qualified assay. We tried to adhere to the European IQ and [inaudible] standards and to be qualified biologic models. This has been easy if you work with models which already very well established like a case of the skin, there are couple of skin models that are available that are highly qualifies and used in the industry but it can be also that when you build up a model from scratch that you have to do qualification internally and requalify that also with partners externally in order to get some kind of global qualification procedures of this model.

The first step is if you combine a couple of different of these organ models together is to from a tissue engineering approach to show that these organ combinations are viable and functional over prolonged periods of time because ideally as these micro-physiological systems or organ-on-a-chip systems are miniature bioreactors, you want to use the advantages of such bioreactor systems which are perfusion-based systems which means you have a better nutrients and oxygen supply leading to complex tissue or cell models compared to standard [inaudible] models, and this also

allows you to go for long-term experiment. In our case we always qualify each and every model to be stable for a period of up to 28 days in the system.

And when you have the qualification from a tissue engineering level achieved then you go on to build your assay around these models systems. The assay then, there couple of parameters where you can model the assay according to your specific needs across the group so we're seeing different numbers of chips being used depending on the end user. Naturally industry has always going for a higher throughput than when you are looking at the academic research. You can go for different administration routes and this is going to be one of the main topics of the case study I will show you. So, you can mimic different application routes in the system and decide on different exposure regimens, so do I go for single-dose exposure or do I go for repeated dose exposure regimen? And you can also play around with dosing. Do I stick with a dosing which is giving a stable concentration in the system or do I do some dose escalation studies in the system?

If you run such a long-term assay you are going to have the option of [inaudible] out of the system so basically you take out [inaudible] media out of the system that you can analyze for cytokine expression but also for metabolite profile. And because the chips are fully microscopable we can also use a lot of imaging methods in order to understand tissue behavior and tissue morphology in the system.

At the end of the experiments you have a lot of viability when it comes to endpoint control so you can remove the intact tissue models out of the system and go for the immuno histology and immunofluorescence staining in order to understand pathophysiology or structure of the tissues. You can go for gene expression analysis, and genome microwaves for a lot of different options available for you. Next slide.

I already talked a little bit about the throughput and the throughput because the systems are definitely more complex that standard *in vitro* systems can be limited. That is why we develop an automated system which is capable of each of these automates that you can see in the picture which can run up to 24 chips in parallel. You combine several of these automated systems in order to increase the throughput of the experiment. These robots would be capable of running up to 96 chips in parallel. The chips are being staged here in some kind of [inaudible]. The automated system and robotic system is capable of performing the substance education and media extraction and the feeding of the chips, below the chips you have a microscopic unit which is capable of light microscopic imaging and fluorescent and imaging and can take pictures of any kind of spot on the chip at any time. What is also important and especially also when we potentially talk in the discussion about regulatory acceptance of these new models, they are more complex so this means they are also more operator dependent then the standard *in vitro* [inaudible] and of course such as automated system removes the operator dependency from such a complex system. Next slide.

I will give you a little bit of overview of what people are using this technology for. I will not go into detail with each of these models, but there's abroad variety of different applications in pharma, biotech, cosmetics and chemical industry where these models are already used or are being investigated for certain uses. Next slide.

Coming to the skin-liver coculture this is actually an approach that began around 2013. Next slide. We first started with our colleagues from [inaudible] with a very small feasibility study. Here we wanted to understand basically for the first time is the system capable of correctly emulating different application routes. What we did is we took a chip two. We cultivated in one organ compartment, a skin model, on the [inaudible] interface and the second compartment we cultivated a liver model, steroid-based liver model. Then we compared topical application on the skin model for the systemic applications so an application into the compartment of the liver where the liver is directly exposed to the substance. We used a couple of different substances here in this example that is shown for retinoic acid. First of all, what you can see here in this diagram on the lower right part of the slide is basically what it is to be expected. You see a different reaction in the skin model depending on how you apply the retinoic acid to the system so retinoic acid binding protein two is differently expressed in comparison from topical application to systemic application. Next slide.

However [inaudible] was also interested in basically do we see a different metabolite profile when we compare the different application routes. This is here to basically show you that we thought the most striking difference. Next slide.

And this is the result of the full metabolite analysis. On the left side used a repeated dose systemic exposure so in the repeated dose format once per day, system was exposed or the liver to the retinoid acid and the repeated topical exposure the system was exposed via the topical via the skin model to the retinoic acid. As you can see the metabolite profile was different. If you compare the concentration on the Y axis to see that there is a general difference and that the profile has a different metabolized most regularly for [inaudible] is different if you compare the different application routes. Next slide.

This is basically the initial start and we started to look at this in a larger consortium and we were very glad for Cosmetics Europe to join into this project and it make it bigger. Next slide.

The aim of the larger project with Cosmetics Europe to basically tackle a couple of different questions. In general we wanted to evaluate this organ on chip technology could contribute to the safety assessment for subacute repeated dose systemic toxicity and we wanted to see if these micro organ chip models would be able to really depict the interaction of skin and the liver in order to understand the specific metabolism of cosmetics and chemicals after single and repeated and donor and systemic exposure routes. Next slide.

With regard to the other models which we selected for the skin model we selected epidermis model because this is one of the best qualified models in the industry. This is inter-exchangeable so it is also possible instead of a epidermis model to work with other epidermis models it has been easy and to also work with other models, you could also go instead of a single epidermis model you could go for a full-thickness model so including also dermis or you could also go directly instead of the model with biopsies in the system and the same holds true for the liver model. In our case we decided on the HepaRG cell line and as I mentioned previously, we wanted to make a little bit more physiologically relevant and at a Stellate cell which are mixed

in a ratio of 24 to 1 which is approximately also the individual ratio of hepatocytes to Stellate cells in human.

For the experimental design, there were two experiments in order to compare a single application versus repeated application. For single application you have the application at day zero and basically adjusted media exchange and sampling on each and every day until after six experiments were [inaudible], while for repeated dose application we started the first application at day zero and then also we took samples out of the system each and every day until four, you exposed the system to a new dose after compounds. In total this goes back to the throughput of the experiment we used 46 to 54 circuit, with special thing is for, each chip has two circuits are your running approximately 23 to 27 chips in parallel. There are 3 to 5 circuits per condition chosen in order to get statistically relevant data, and we also see in the next slides, we included the negative controls. Okay go back. For the endpoint controls we actually checked for substance application after 24 hours and 48 hours and five days and the readouts, metabolic analysis, parent compounds, metabolite quantification, histology, qPCR, and RNA sequencing.

This project is divided into a couple of phases. With the first phase we basically selected these two compounds which do have some relevance for cosmetics but are not able to call them prime compounds so the first one is permethrin, which can be included in ointment against scabies, and the other one is hyperforin, which is included in cosmetics and dermatics. Next slide.

First of all what we wanted to test in order to understand is it possible to actually model repeated dose exposure of the compounds into solvents which we are using on the skin models of choice or to test it on the skin models, so do these skin models keep their barrier function over the repeated dose exposure over repeated dose exposure of five days and as you can see for the permethrin and hyperforin although we see some thickening of the stratum corneum, comparing to the normal control that the barrier stayed intact, we saw some slight deviation in T value, but principle in general the barrier stayed intact. Next slide.

For permethrin, I will focus on the data that we got for the metabolites. Next slide. What we did is with regard to the metabolites similar to, the previous experiment was to get an understanding of do single application versus repeated applications and do topical versus systemic applications make a difference in the system? As you can see on these graphs there is a single systemic single application. The systemic means to directly models by topic means it is applied first to the skin model, on top of the skin model and to penetrate through the skin before it reaches the circulation system and is then distributed by the circulation system to the liver model. That these kinetics are a single topical application are different to the single systemic application which are shown here. But for the repeated dose applications we only saw differences for two metabolites which are shown here in, the M2 and the M16 [inaudible] metabolites. Next slide.

For the hyperforin to give you some ideas on the XME gene induction modulation. Next slide. So, here again for the hyperforin where we first did was to do and this was true almost all of the experiments we did a lot of the experiments together the [inaudible] where we also look at these models and characterized their condition in

order to be able to compared these and these are actually results from the study conditions that we, where we checked for cytotoxicity of hyperforin. The darker blue is the ADH values at the different concentrations of hyperforin after 24 hours of incubation. Light blue is ADH after 48 hours and you see that after around 0.64 micromolar but at least at 1 micromolar you get an increase in the ADH level in these liver static liver models. The green box, dark green is 24 hours after albumin production and light green is 48 hours albumin production. You also see that the higher the concentration is increased we see a stinking decrease in the albumin production meaning that hyperforin at higher concentrations has a pronounced effect on the liver cell activity.

We did the same experiment actually with a higher concentration of around 2.5 micromolar, no 1.25 micromolar. Because of the differences of the static to the dynamic perfusion-based system, the model is reacting differently to the exposure of hyperforin, so we were able to work with higher concentrations compared to the static. Next slide.

In terms of dene modulation, we looked at quite a lot of different genes. I just took a couple and again it was important for us to see that depending on the application root there is a difference in the gene modulation. The systemic application shown in red and the topical application route shown in blue and you can see for a couple of enzymes you see significant difference in the gene expression. Next slide.

What you may have seen in the previous slide is that a part of the whole study was to do some inter-laboratory comparison. These experiments run in parallel at [inaudible] and at TissUse in order to understand are these findings reproducible by another lab if the application or the assay is transferred and persons who [inaudible] a lot of experience working with our system are capable of reproducing the results which we found internally. This is to give you some additional idea on the comparison here, on the left side is the [inaudible] results and on the right is the results for our internal testing at TissUse. The trend is the same. Now we have some at TissUse we did some more repeats and that is why some are more statistically significant changes are upcoming here, but if you compare the data you will see that the trend is the same. It was very encouraging to see that other labs that are not total experts in microphysiological systems were able to reproduce the results we got internally.

Summing it up this was the project phase 1. We can complete that project phase successfully which then, next slide, which then led to the project phase two where we took some more input from the Cosmetics Europe task force in order to determine which chemical to select. Especially for their parameters to alter XME metabolism, then we wanted to have some *in vivo* data available [inaudible] this is available for human data because [inaudible] is present in soy-derived food products so you also have plasma levels in humans available. For AHT which is used in hair dyes and hair colorations. This is only available in red. But both are, genistein is also used in cosmetics, it has antioxidant and anti-inflammatory products, so both are used in cosmetic products. Topical exposure is relevant. We know that they have quite fast dermal penetration to previous ones for permethrin and hyperforin. They slowed penetrators through the skin, so we wanted to have different attributes for the chemicals here. As I mentioned *in vivo* application route we had that available and

the cost is always an issue, so we try to select compounds which wouldn't cost too much when we are performing experiments with them. Next slide.

For AHT the first pass effect in skin was very crucial in defining what kind of metabolites you will find for instance in the [inaudible] shown in the small table. This is a table from a rat study where the compared the oral uptake and topical application and depending on the application route we see quite a difference in the composition of the metabolites and the plasma of the red so and this is why we picked this compound. The skin has the capability to immediately metabolize to AHT which can also later be produced by the liver and AHT can also be further delivered and circulated or glucuronidated while the parent compound that is also being able to be circulated or glucuronidated by the liver. Next slide.

We found some quite interesting things and again trying to walk you through the slide. Not shown we did some preliminary experiments [inaudible]. I will come back to that because [inaudible]. First of all, we compared the concentration levels of the parent compound and the metabolite into different compartments at different computations [inaudible] and you see that you find the parent compound in the same compartment [inaudible] so it’s penetrating quite quickly through the skin, so after 15 minute you see a high degree of the parent compound in the skin compartment but you also see an increase in time in the metabolite. What you see is basically a decrease of the parent compound and you find more parent compound over time in the liver compartment because the pump is turned on and basically the pump is distributing the parent compound to the liver. For the small concentration, this is what we expected to find.

For the high concentration we saw that the first pass effect was not as pronounced as it was pronounced as it was for the small concentration, as you can see here. Only a very minor fraction of the parent compound was actually metabolized. This was a stark difference or contrast to the static experiment. We believe this is because this high concentration is oversaturated in the system and it is passing through and then as shown again here and is going to the liver compartment where it can be detected. In the static [inaudible] system, however, because the parent compound stays underneath the skin model is because of the path of diffusion has the possibility to interact with keratinocytes in the skin model and then over time because it cannot be distributed to any place in the well-plate-based system, it is slowly metabolized. So, you do see a difference here depending on the [inaudible] the microfluidic multiorgan system, how the system and the penetration kinetics are in case of the AHT. This graph basically sums it up, took a look at the circuit. Again, showing you that you have higher concentrations of the parent compound in the system when you increase the concentration of the parent compound. This is for the skin model only. In order to understand the first pass [inaudible] effect, we tested the combination of skin and liver, and as you can see here again, depending on the application route and then took a look at the other compounds which are metabolize which are produced by the liver, we got for a lot of this compound different metabolite profiles depending on the application route. Next slide.

For genistein, this is a snapshot of the data. For those of you who are Cosmetics Europe, the full report should be available. We took a look at the stability of the XME gene modulation, the XME genes over time, you see here that the system is quite

stable so the genes are steadily expressed over time and therefore also able to produce the relevant metabolites over the experimental timeframe in a stable [inaudible]. This holds true for both the two main metabolite pathways of genistein which is glucuronidation and sulfation. Next slide.

Again, summing it up, I did that type two times excellent intra-laboratory reproducibility, so it was you could recapitulate it internally, we also did these experiments in two laboratories in parallel, so this was also nicely done. The important point here is basically for both compounds we could show that the route and frequency of application results in different effects on the gene expression in the epidermis and the liver organoids, and as I said, for anybody who has access, I'm not sure who has access exactly to that report, I invite you to read the full report on this case study. Next slide. Thank you.

I will finish my talk with basically using a recent application from Thomas and colleagues from the U.S. EPA where they already tried to include the new systems, the microphysiological systems, into a tiered testing framework where there is a clear potential use of these systems at the tier 3 stage especially if you have identified no adverse outcome pathway, but I'm also looking forward to discussing this with all of you in the roundtable discussion which should be coming up after my talk. Next slide.

I want to think the people who are involved in the study. So, from [inaudible] Katrin Brandmair, Thamee Rings, Slike Gerlach, Andreas Schepky, Nicky Hewitt from Cosmetics Europe and Thi and Ilka from TissUse and of course Pharmacelsus for the invaluable metabolite analysis because for these great analytics from Pharmacelsus we couldn't have done most of the case study work. Next slide.

And of course, as mentioned previously this project was financially funded but also very much scientifically supported by the Cosmetics Europe ADME task force so very big thank you to all of the people involved here from the different companies. This was really helpful and the discussions we had were very great, and I think that the ongoing project, we are on a very good way to further characterize and qualify the system for later potential use in risk assessment. With that, I'm at the end of my talk. Thank you very much for listening to me, and I am very happy to later take questions. Thank you very much.

Ceriani: Thank you, Reyk, for your amazing presentation on this technique and its applications.

Roundtable Discussion
Moderator: Lidia Ceriani
All speakers

Lidia Ceriani: I think that now we can jump right into the roundtable discussion. I am going to read the questions as they have been entered into the chat box, so we will start with the first one and it is for Gladys. Since [inaudible] is just a similarity metric that can be based on any input parameters, does [inaudible] similarity in the context of the currently discussed case studies refer to similarity based on a genetic fingerprint or the different types of information like reactivity that is referred to?

Gladys Ouédraogo: The search of analog was not based on the [inaudible] similarity but based on features, some phys-chem properties, also reactivity, indeed metabolism, all these were considered, the functional aspect was considered in selecting the analogs.

Ceriani: There is another question for you. Is the transcriptomic data also from ToxCast data?

Ouédraogo: The transcriptomics data were generated within the Cosmetics Europe science program, so experiments were done in the lab at Procter & Gamble and they were analyzed based on the connectivity map approach. So, these were generated specifically within the specific initiatives here. We did have existing data to work with.

Ceriani: There is another question. Regarding the need for consistence in terminology, does this raise questions about the use of different notions in read-across? I may have missed something, but it seems that the basic kind of read-across could be considered a component of IATA whereas the next generation framework would be considered an IATA in and of itself. I can break the question into pieces. The first question, I think that Andrea is the correct person to answer it. Does this raise questions about the use of different notions of read across? Andrea?

Andrea-Nicole Richarz: The assumption here is correct. While the traditional read-across would be considered a subset of an IATA while the next-generational read-across is a whole IATA because we do next-generational risk assessment. That's correct.

Ceriani: The second part is [inaudible] and for the first part, I'm going back to Andrea.

Richarz: I think it is a good question and read-across might be a good example for a term that might use some more harmonization of terminology. Actually a similar question came up in the discussion when we were looking at IATA in the discussions for the overview document and there was no consistent view whether the read across is a NAM to start with, so is it a method or methodology or is it a weight of evidence on its own and as pointed out [inaudible] range of complexity indeed, from very straightforward extrapolation from one analog to another, to quite complex read-across, but I would say it can still fall under the framework of IATA which can comprise quite complex approaches of a [inaudible] component is considered to be quite broad. This could still be further discussed indeed.

Ceriani: There is another question for Nicole. In the two out of three approach, wouldn't a positive result for [inaudible] key events be more significant? Is that taken into account?

Nicole Kleinstreuer: Thanks for the question. It is a good point and it is something that has been discussed at length in the expert group on defined approaches to skin sensitization. We ended up doing is a permutation analysis to look at the impact of testing order when looking at the various key events-based test methods in the two out of three defined approach.

So, looking at it objectively, does the performance of the two out of three defined approach change based on whether or not you are looking at key events two and three or key events one and three or key events one and two and what order you preform those tests in? The reference data has yet to be finalized and we will be redoing these analyses, but from a preliminary perspective the answer is no. Within the confidence intervals and the uncertainty defined by the reproducibility of the reference data, the order of testing as far as the key events that are used in that two out of three approach does not seem to make a big difference.

There is some concern that in fact the key event three-based test methods may actually be a little bit more prone to false positives and they are more sensitive that the other key upstream key event-based methods. Of course, that is potentially positive or negative based on your intended use. Whether or not you are trying to optimize compound lead development or whether you're looking from a regulatory perspective or you're interested in maximizing sensitivity or specificity, that may have more less benefit to have a sensitive test method.

Ceriani: Thank you very much, Nicole. There are a few more questions for you. The second one is, is there a ballpark estimate of when the EPA defined approach guidance will be updated to include additional more complex defined approaches such as AEN?

Kleinstreuer: I couldn't comment on when the actual draft policy on the acceptance of defined approaches as an alternative to the animal test for skin sensitization, when that document will be updated. That is obviously at the discretion of the EPA. So I'm not sure about that, but I can say that the draft risk assessment of the six isothiazolinone compounds that relies exclusively on the artificial neural network defined approach, that draft risk assessment was published on May 14th in the *Federal Register* notice and it is open for public comment right now. I was certainly encourage folks to take a look at that because regardless of the policy document, that's is an example, direct example of the EPA relying upon a more complex defined approach that provides that quantitative potency production value as a basis for deriving points of departure for their risk assessment.

Ceriani: Thank you. Could you elaborate on the size of the [inaudible] dataset number of compounds that come along with human data?

Kleinstreuer: The dataset is still being finalized. This is a joint effort being led by the German Institute for Risk Assessment, so the BFR, and NICEATM. At the moment we have compiled a database of over 2500 human predictive patch test results. Some of those are replicate results on the same compound although I would say that the vast majority of compounds in the database have only one predictive patch test results associated with them. At the moment it looks like the database is going to end up being having around 1200 unique substances with human predictive patch test results. Those are not all mono-constituent substances. Some of those are natural products. Some of them are mixtures with undefined structures. They have all, all of the studies are being characterized with respect to their relative reliability, so we developed a system for categorizing those predictive patch test results. We really, in the support of the OECD project, we focused on applying that UN GHS classification criteria to the human predictive patch test results to try and come up

with classifications for the reference chemicals that also have *in vitro* data defined approaches being applied to them and have *in vivo* LLNA data. That is a much smaller subset and more in the order to 80 to 100 chemicals. As far as the overall size of the database, it will cover about 1200 substances; we plan to make it publicly available. We have a series of manuscripts that we're developing to describe the database to characterize their ability and uncertainty in the data and I think this will be a tremendous resource for the community once we can get it out there.

Ceriani: Great. That was clear. Thank you, Nicole. So, we've got three questions for Reyk. With regards to the [inaudible] chips, can we request for customization for these? Please let me know if I need to repeat the question.

Reyk Horland: Basically, the chips are customizable. The [inaudible] is customizable and the complete design of the chips are customizable, so we do actually have quite a lot of variations already in the portfolios of these chips to chip systems. We're constantly working on other versions of these chips, increasing the number of organs which can be combined, so this is definitely possible.

Ceriani: The second one is, depending on the endpoints, can we request to develop it to mimic specific organ physiology?

Horland: Yes. This is definitely also possible. The organ and endpoints can be defined and also with regards to organ physiology and the functions of the organ, this can be defined and basically emulated in the chip system. This sometimes can be done with the chips that you already have available. In other cases, like if you want to have electrical coupling of neurons or cardiomyocytes, there may be adaptations of the chips necessary, but in general this is possible.

Ceriani: Thank you. Are there contracting services available to conduct experiments?

Horland: Yes. There are services available and this is also what we are discussing currently with parts of Cosmetics Europe because I also tried to mention that in my presentation that we shouldn't be the only lab who is capable of working with this technology, so our philosophy here is to say we want to enable as many possible to work with this technology and we do have transferred this technology to a couple of labs. Either we can perform this contract testing services but also other labs are available to do that for you.

Ceriani: Thank you very much, Reyk. Then we've got another few questions about PBPK modeling. I think that Susanne and Gladys are the best persons to answer them. So, I will read them now. Are there any disadvantages in employing PBPK modeling? It does allow reduced uncertainty in relation to PK aspects when extrapolating from animal to human, however, being based on the modeling tool, wouldn't there be some kind of uncertainty associated? I don't know who wants to go first. Perhaps Susanne?

Susanne Hougaard Bennekou: If I can answer for my regulatory background, I think one of the biggest challenges there is actually to explain or to increase the confidence by the regulators. I sense quite a lot of skepticism still there and I think

we need to—we are underestimating what is needed for the recognition for them to really trust that type of data. And that, then you can start discussing the uncertainties but if they think it that if they don't completely understand it and it is not well explained then there is a lot to bridge, I think.

Ceriani: I agree with you, Susanne. Gladys, from your point of view with your case studies?

Ouédraogo: The key about confidence in these tools is important. I think it is like any other tool actually. There is some uncertainty no matter which model you are using. Even if it is an animal models. Here with the PBPK modeling, it is the case but there are verifications that are done when PBPK simulations are exactly performed to see which type of parameters are most influential as input parameters. Because there are physiological parameters depending on the system we are considering. Either it is an animal, human physiological parameters, and a chemical specific parameter is also a measure of model and imported into a model to do the simulation. Right now, actually there is an effort ongoing at the OECD level to define standards for PBPK especially when no *in vivo* input its actual available. To be able to do the simulation still and an estimate the uncertainty. Is mostly a model of estimated uncertainty because let's say we have some benchmark information and we know that we have the twofold or tenfold difference between the prediction and the actual measurement if we have a benchmark. What does it mean for the decision we are trying to make? That is where we need to see whether it makes a difference. So, there are ways to assess that and see whether it is enough in terms of confidence to make the decision we are trying to make.

Ceriani: In your experience, I'm sorry.

Kleinstreuer: This is Nicole. I wondered if I could add a thought to the discussion here. I completely agree that the characterization of the uncertainty is really a necessary precursor to being able to establish confidence in regulatory application of PBPK modeling. Another effort to that we have try to undertake is to do as much of an apples-to-apples comparison as possible when looking at both the biology that is being measured on the *in vitro* side and on the *in vivo* side. And then iterating through various pharmacokinetic approaches to do that *in vitro* to *in vivo* extrapolation in the middle. And looking at the impact of using models of various complexity, using, adjusting using math balance equations to estimate intracellular exposure. Adjusting for a fraction of unbound chemical in the plasma with respect to steady-state model calculation to approximate bioavailability. To do that we actually, I will put the link in the chat here, we published a paper we looked at the U.S. EPA ToxCast estrogen receptor pathway model and compared it to *in vivo* uterotrophic guidelines studies for a range of reference estrogen receptor agonists, so the ER pathway model based on the ToxCast data has been validated as a replacement for the uterotrophic assay in the endocrine-disruptor screening program from the EPA and is also considered by ECHA and EFSA as well for indicating estrogenic modality for things like pesticides. It is pretty well-established that pathway model is really measuring the same biology as the uterotrophic animal test so using those two anchoring endpoints on the *in vitro* side and the *in vivo* side that are measuring the same biological pathway, we were then able to iterate through the methods that were used to do the PK modeling and the *in vitro* to *in vivo* extrapolation, do the source of

the parameters whether the intrinsic clearance and the fraction unbound were measured using experimental values or whether they were predicted using *in silico* platforms and putting you through the different approaches and combinations to look at the impact of that *in vitro* to *in vivo* extrapolation based on the different approaches and different complexities. So, I think that approaches like that are what is needed to try and help establish confidence in being able to use PBPK modeling to translate *in vitro* activity concentrations to estimated equivalent administered doses in a way that will allow regulators to have confidence in using those.

Hougaard Bennekou: If I can just add, it is a really good initiative you mentioned, Andrea, with the good modeling practice. This is also something that could help here. That coming from the OECD.

Richarz: Yes, indeed. That has been discussed and it will be we discussed at the OECD.

Kleinstreuer: Andrea, that guidance document is in preparation right now. Is that right?

Richarz: The overview document, yes. It has been reviewed and commented and will be presented at the next WPHA meeting. So, hopefully it will be published, the document and the Excel sheet with all of the documents in it.

Kleinstreuer: That is great.

Ceriani: And Andrea, will there be a follow-up on this project perhaps under future activities?

Richarz: There are some needs found in the project and recommendations have been made. I cannot really talk on behalf of the OECD, but I think this issue will be discussed in the OECD and there might be some upcoming activities to be decided upon. There are more discussions needed and some more follow-up is needed. This is up to the OECD working groups to decide.

Ceriani: Sure. So, thank you for these really comprehensive answers. Does anyone have ant freely available PBPK tools to recommend? Based on your experience or on the case studies that you.

Kleinstreuer: This in Nicole. I can start. The EPA Center for Computational Toxicology and Exposure has a really excellent R package called the HTTK, so high-throughput toxicokinetics R package. It does require a certain amount of familiarity with the programming language but there is really excellent documentation for folks that are interested in learning that. For those that would prefer to be able to access the models in the HTTK package through graphical user interface, they are welcome to visit the integrated chemical environments page that is the National Toxicology Program, so I put the link in the chat box. Under our tools section, we have an IVIVE tool that is essentially a front end to the EPA's HTTK package. So, you can go and choose what type of *in vitro* endpoint and choose all of the ToxCast and Tox21 data are available in [inaudible], so if you want to start with that as the basis for doing the *in vitro* to *in vivo* extrapolation that is already uploaded in the database but there is

also the option for users to upload their own *in vitro* data and perform the IVIVE and you can choose the parameterize based on the species, you can choose multiple different models, exposure routes, various simulation parameters. And then you can choose which chemicals you would like to run it on either by user-supplied cast numbers or based on various quick list that we have in the [inaudible] database. We're working constantly with EPA to continue to update the models that are available on the back end as EPA updates their HTTK package. I would say for more gaining familiarity with the tool, this is a great interface; for more power users that really want to dive into a deeper analysis, I would recommend the HTTK package.

Ceriani: Thank you, Nicole. I think there is another question that you are the best person to answer. What is the importance of computational toxicology in integrated toxicology assessment?

Kleinstreuer: I think that as we have seen from a number of the presentations today, computational toxicology, which really is an extraordinarily broad term, is becoming a fundamental cornerstone of a lot of the integrated approaches so depending on how you define computational toxicology, we're all using computational resources in every facet of our day-to-day lives to perform our jobs and do our research. So, in the case of things like developing defined approaches and IATAs, a number of them rely on tools like QSAR models or *in silico* predictions or parameters that are integrated with *in vitro* data. Other defined approaches rely on machine learning algorithms that are built off of training data sets and use computational approaches to make predictions on new chemicals. But then even more broadly some of the work that Reyk showed us on the microphysiological systems require precise computing in terms of the protocols and the setup of those from a physical perspective and from a measurement precision perspective. So, I think that it is really fundamental to every aspect of toxicology, now, the integration of computational systems and software.

Ouédraogo: Gladys here. If I may add, that is definitely the case. I think it is difficult actually to do toxicology nowadays without some computational tools and when it comes to IATA, one of the ways to facilitate the uptake of these IATAs, because we can see that they are quite flexible. The defined approaches, it's straightforward, actually. But when it comes to flexibility and things like that, we talked about the challenges of documentation. There guidance documents out there, but still, when it comes to putting things into practice, if we were actually to have computational tools, we will still have to justify the rationale for picking some type of data, some *in silico* or *in vitro* assays and things like that, but if we were to have this tool, and I know some of them exist, some computational workflows, it's really helpful in terms of bringing transparency and practice of putting IATA into practice. I believe all of these computational tools are key here for IATAs.

Ceriani: I agree with you. We have received a few questions from the FDA by email. The first one is. If the goal is to evaluate risk versus showing safety, how does the [inaudible]development of IATA in the process of using NAMs? I think Gladys and Susanne are the best persons to answer this question, but anyone can give their input. Gladys, what do you think?

Ouédraogo: In terms of IATA per se, it wouldn't be any different. Unless there is actually a guidance related to a specific need where there would be some prescriptions in terms of types of information that needed to be considered, because we have seen for the IATAs, we have different blocks and within these blocks, depending on the decision, it depends on the problem formulation. Depending on how you formulate your problem. You will actually call in specific pieces of information, and maybe not all of them. The flexibility lies there as well. If it comes to risk assessment and where you have actually the knowledge to be able to dig into a mechanism and to document everything and to state where you have a concern, I think you'll call the information needed and I am thinking were AOPs are strong, robust, and can be used, you could use that to show that the likelihood of triggering some adverse outcome there. But where you don't have defined adverse outcomes, specific target organs to work with, you need to go with the weight of evidence. In terms of structure, it wouldn't change the IATA per se. It would change the type of information you would use to make your decision. So, that is where I think the differences maybe lie, but in terms of IATA, it wouldn't change the structure of IATA. It would change the specific guidance related to specific questions you are posing. Susanne, I don't know if you want to add something.

Hougaard Bennekou: This is what I had in mind. I think I can only echo you. Gladys, that you have to be really careful in your problem formulation and make sure that you really specify which regulatory requirements are you going to address. And then it would be case specific. And depending on the endpoint and what other kind of data you have around it, so but it would, that is within the flexibility of the IATAs. If you want to do hazard characterization versus risk assessment, it really depends.

Ceriani: Thank you to both. I agree; we have to ask the right question upfront. Then there is another FDA question. How are IATAs modified for post-market evaluation versus pre-market evaluation? Gladys, do you want to have a go with this one?

Ouédraogo: I can give it a try. The same here. IATA would not be modified. If it is about post-marketing [inaudible] we would like to call from epidemiological evidence. We need to collect whatever is available for us to start formulating the hypotheses and the exposure issue would be key as well here. So, it really depends on the type, the type of information you will be collecting would depend on the type of issue you are facing. And how you want to solve it. It could be something, something that is quite huge, you don't have two years to solve it. You may to solve it within the month or weeks. All of these considerations need to be taken into account. So, it would be part of a problem formulation of what is yet context, what decision needs to be made there and what tools you have. What means do you have to make your decisions? So, based on that, you use your IATA to structure how you will actually proceed to collect the existing information and generate new ones if you have means to do so within the given timeframe as well to make your decision.

Ceriani: Thank you very much.

Hougaard Bennekou: Just to supplement, you will use different tools, so in the post-market situation, you probably have more data and compared to the pre-market situation. And this of course would guide you where you need to fill gaps and guide you in what kind of questions you are asking on the first place.

Ceriani: Thank you both. We have an SOT question by email. How do we move beyond skin sensitization? I would pose this question to Nicole.

Kleinstreuer: I can take that first pass at it, but the other panelists are just as well informed and able to give their opinions on this as well. I think the advantage of skin sensitization is how far along we are in terms of the development of the AOP, and obviously the existence of human data as an added benefit to have that anchoring endpoint. It provides a really nice proof of concept for using the AOP framework to develop defined approaches and IATA, and it shows that when you are driven by knowledge of mechanisms and adverse outcome pathways, and when you're developing testing strategies that adequately covered that biology, then you can really exceed the performance of the existing animal test because you're using human molecular targets and human cell-based systems that are relevant to your endpoint of concern. And as the others have said, it is all about having the right problem formulation as you begin to tackle the specific endpoint that you're looking for.

So, moving past skin sensitization but using skin sensitization as a guiding principle that have been developed and have been shown to be really successful in that area, we have, the broader we, globally, many groups are working on this, have a number of ongoing efforts to address other topical toxicity endpoints, such as eye irritation and skin irritation. There are models for defined approaches, adverse outcome pathways that are being proposed, reference data that's being curated, lots of ongoing work to try to validate some of those defined approaches as replacements for the animal test.

Those acute topical endpoints are certainly low-hanging fruit. But we heard some really nice examples today of how more complex endpoints are being addressed by things like the EU ToxRisk case studies and how the AOP frameworks are actually being leveraged to establish testing strategies that really might help inform on things like developmental neurotoxicity and other types of more repeat dose complex chronic toxicity endpoints. So, I think am very optimistic as far as this really robust scientific basis that is being used to apply the adverse outcome pathway framework to developing these testing strategies and how we're moving together in lockstep to make incremental but very substantial progress in areas that are maybe not quite so straightforward as skin sensitization. I welcome comments from other folks on this that are more informed with other endpoints.

Hougaard Bennekou: It's such a nice case and it is so well-defined, and I think what also makes it or should make it acceptable is that you are using data from test guidelines that has already been validated. But the next leap is really to do the same, but where you don't have validated test guidelines, because this will be probably where we can really progress it more; otherwise it will be quite a bit slow, maybe, to be blunt. I think we need to make, take the next step and try to see how we can create that confidence in other types of systems that have not been validated formally. It is quite a step, I think.

Ouédraogo: It is actually. At the same time, I want to say there is effort ongoing, and I think we won't have guidelines or guidance documents for everything. I think we

need to be creative and find ways to build confidence. That is why think collaboration is key. Because when we think about more complex endpoints and it is actually about the case study that I have shared here and the case study that you shared, Susanne, from the EU ToxRisk side, we kind of have already demonstrated, and we still have more coming to demonstrate that things are pretty possible, optimistic there, but we need to be creative in terms of building confidence. How will we work together with all stakeholders to be confidence in these approaches?

Ceriani: Thanks to everyone. There are few more questions on the chat box. And thank you, Gladys, for answering them. Second from the last one, I know that there are some private consortiums developing AOP repository databases. Are there any public ones available? Yes. There's the AOP Wiki, so thank you, Gladys, for posting the link.

And the other one, unfortunately, worldwide regulatory agency takes so long in accepting new concepts. Tell me why we will be conducting a huge number of animal studies, to show I know that [inaudible] because it takes one in a million to have a read-across assessment accepted by [inaudible]. How will [inaudible] adopt these new approaches to evaluate and approve more read-across assessments [inaudible] compliance. I would like to remind that we cannot provide any regulatory feedback or advice in this regard. So, thank you, Gladys, for your response. So, yes, regulatory acceptance of these non-animal-based approaches is a challenge. We need to work directly with the agencies as we develop them to get feedback and collaborate with them.

Thank you all. We're at a quarter to seven here in Italy, so I think that we're all set. I want to thank everyone who listened. A really big thank you to our speakers for your super job. Now we turn it over briefly to Dr. Jason Aungst, who is the FDA liaison for the SOT Colloquia, for the concluding remarks. Jason, are you online?

Jason Aungst: Yes. Can you hear me?

Ceriani: Yes, we can. I guess we'll have to move to the next slide.

Aungst: Next slide. Next slide. Thanks, everybody, for attending today. I want to give you a notice that we would have two more colloquia next year. In December, well, next fiscal year. In December of 2020, we will have a session on new plant-based foods and proteins from novel sources, and in April 2021, we will be talking about the toxicology of nanoparticles. Please keep an eye out for those advertisements. Next slide.

As always, information, recording slides, captioned text, will be available on www.toxicology.org, that's SOT's website, so you can find all the information there. Give us a few days to get that uploaded from today's colloquium. Next slide.

Thank you for participation. Thank you, speakers, for a great session today. And for all of the good questions we received and discussions we had. Thank you and enjoy the rest of your day.

Ceriani: Thank you. Bye.