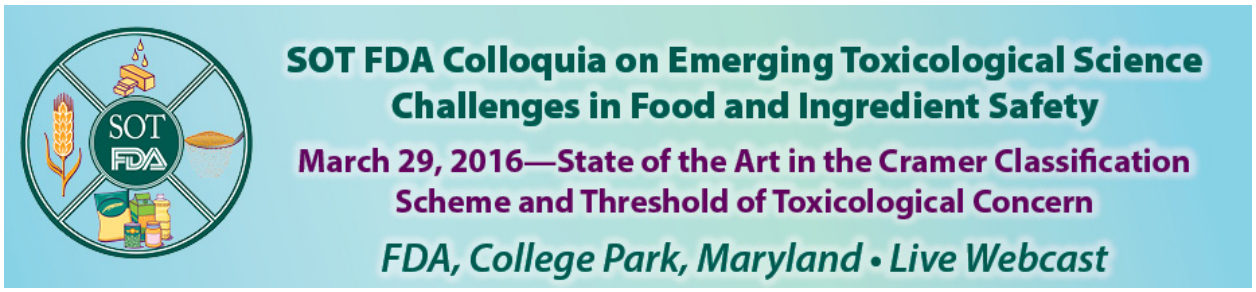

Real-Time Captioning

Note: This is not a transcript.

March 29, 2016

State of the Art in the Cramer Classification Scheme and Threshold of Toxicological Concern

Chair: Ivan Rusyn, Texas A&M, College Station, TX
Co-chair: Timothy Adams, US FDA, College Park, MD

8:00 AM–8:30 AM	Badge Pick Up
8:30 AM–8:40 AM	US FDA Welcome and Overview, Mary Torrence, Director, Office of Applied Research and Safety Assessment Laboratories (OARSA), US FDA, College Park, MD
8:40 AM–8:50 AM	Welcome from SOT and Introductions, Peter Goering, PhD, SOT President, US FDA, Silver Spring, MD Speaker Introductions, Timothy Adams, US FDA, College Park, MD
8:50 AM–9:30 AM	Threshold of Toxicological Concern Approach in Regulatory Decision-Making: The Past, Present, and Future, Grace Patlewicz (Tier), US EPA, Research Triangle Park, NC
9:30 AM–10:10 AM	Advancements in Food Ingredient Safety Prioritization: An Expanded Cramer Decision Tree Schema, Timothy Adams, US FDA, College Park, MD
10:10 AM–10:30 AM	Break
10:30 AM–11:10 AM	*In Silico* Methods for TTC Assessment, Andrew Worth, European Commission, Joint Research Centre, Ispra, Italy
11:10 AM–11:50 AM	Quantitative Prediction of Continuous Toxicity Values using Chemical Structure Information, Jessica Wignall, ICF International, Arlington, VA

11:50 AM–12:50 PM Roundtable Discussion
Ivan Rusyn, Moderator
Kristi Jacobs, US FDA, College Park, MD
All speakers

US FDA Welcome and Overview, Mary Torrence, Director, Office of Applied Research and Safety Assessment Laboratories (OARSA), US FDA, College Park, MD

We are going to get started. So we can keep everyone on time.

First of all, welcome to the third in the series of four SOT FDA colloquia on Emerging Toxicologic Science Challenges in Food and Ingredient Safety. [inaudible]

I am the current director of the Office of Applied Research and Safety Assessment, OARSA, and I really want to thank the organizing committee especially [inaudible] and Alan for inviting me to introduce the session. I think these colloquia are great models for all of us to look at emerging science and to promote collaboration to cross centers across disciplines and they are fun to attend.

Just three brief slides about OARSA and the Division of Applied Toxicology. If you look at the slide, the Division of Applied Toxicology's strategic plan is really what this is all about. What it does is it provides us with a flexible and responsive range of approaches for emerging issues. As well as other strategic regulatory needs. As you can see the core of the Division of Applied Toxicology is to predict and translate for regulatory needs and emerging science. It's a combination of in vivo, in vitro computer modeling. They can develop approaches depending on what the question is.

Here is more specifics about what the Division of Applied Toxicology is skilled to do. You can see some of the work they are currently doing. They have a whole range of organs of the body they can use as models. For acute toxicity, and then in vitro, in silico, they have stem cell models and 3-D models they are using. All of this is to balance what is needed for significant regulatory research outcome. I apologize, as I've look at the slide, I see this text going along the bottom. I'm thinking what is going on? That's why I may look confused. Finally, this is talking about the Division of Applied Toxicology and how they provide toxicologic expertise. To all the different offices within CFSAN. There's a few examples of each. They are working with dietary supplements, we're looking with doing some silver developmental toxicity work. They do significant work for cosmetics and colors. Melamine and we are also expanding some work we are doing for the dietary supplements. You also see there is three circles outside of CFSAN which is important as well. They are involved in NCATS and currently working with the senior advisor for toxicology Dr. Fitzpatrick on really strengthening FDA's role in Tox21 and what do we need to do to make it successful. We are working with the University of Mississippi, we have a couple of toxicologists going down there to go and talk about botanicals and of course working with an CTR. The ideal for OARSA in any of the visions is that we can be flexible and be responsive, and we're wanting collaboration from the groups for the best science.

The series is a partnership between SOT and FDA the Center for Food Safety and Applied Nutrition. It has started as a small idea and it's been fun to watch how it has grown. Among the group. It is the idea to have high quality cutting edge, toxicologic science, for the FDA

employees and the public. This is open to the public. I think the important thing is this is about the science. The latest toxicology update, research, techniques, and approaches. It is not a public forum for discussion of regulatory issues. I think that is important. The way this started was FDA partnered with the SOT to develop these training sessions. With an MOU. So really to talk about shared interest in scientific progress and the disciplines that directly, and indirectly affect both human and animal health and medicine. The memorandum of understanding is really setting up the possibility to do these training events, workshops, other conferences, all for the goal of the newest toxicology. Both for scientists, SOT, and FDA. Again, they are not intended to be a discussion about regulatory issues.

The Society of Toxicology, their mission is to create a healthier and safer world by advancing the science and increasing the impact of toxicology. I think that is very parallel to CFSAN as well. The priorities to strengthen the relevance and impact of it and support toxicologists and really look to the future and expand the outreach and impact globally.

The six previous colloquia hopefully you attended all of these or at least some of them. Very interesting topics from immuno talks to risk assessment to computational in vitro methods. Evaluating human clinical and observational data, I think some of these are important because they give us this regulatory need and the merging of science and regulations. The participation from 2014–2016, it just keeps growing. I think it's just really a testament to the success of this partnership.

The future colloquia in May will be a Safety Assessment Approaches in Young Children. The assessable population and they are currently working on additional colloquia for 2016–2017. The organizing committee is an amazing range of expertise and knowledge. I think you will meet most of them today. Again, thanks to Alan for inviting me to introduce this. The moderators today will be the chair of the organizing committee, Ivan Rusyn from Texas A&M, and Timothy Adams from FDA. The agenda hopefully you got it in your packet. It's full of a lot of information.

A couple of administrative items. One, the restrooms are outside the auditorium halfway down the hallway. You probably passed them when you register. The public cafeteria is outside the FDA so it is public you have to come back through security. There will be cards handed out for those who like to submit questions for the panel discussion at the end of the meeting. Depending on the timing, we may not get all of the questions answered.

I want to introduce Peter Goering who is the president of the Society of Toxicology. He says please do not read my whole bio. Dr. Peter Goering is a research toxicologist and lab leader in the laboratory of toxicology and biocompatibility at FDA Center for Devices and Radiological Health in Silver Spring, Maryland. He earned his PhD from the University of Kansas Medical Center. Did a postdoc at NIEHS focusing on the toxicology of metals and metal defense mechanisms. Here at FDA his research interests include nanotoxicology, evaluating liver and kidney toxic injury, ellucidating new biomarkers of toxicity and understanding mechanisms of metal toxicity. He is co-author of a lot of publications and book chapters. He's a diplomat at the American Board of Toxicology. And a fellow at the Academy of Toxicological Sciences in 2006. I will stop there but I think all of us would agree that FDA is lucky to have him. And as president of the Society of Toxicology.

Welcome from SOT and Introductions, Peter Goering, PhD, SOT President, US FDA, Silver Spring, MD

Thank you, Mary, and welcome to everyone at the colloquia this morning. As Mary elaborated, this is a teleprompter Mary all you had to do is read it. As Mary elaborated, I think the collaboration between FDA and SOT is a very natural one. We have very common goals in creating a healthier world, protecting the health of the public, looking at innovations in technology to advance medical products and safer food supply, as well as innovations in toxicological science. I should've started on behalf of the Society of Toxicology, we are so very pleased to be a collaborator with the Food and Drug Administration on this series of colloquia. I'm so pleased to be a part of helping to organize these sessions.

We have had a number of these colloquia and the success has been so important that we met two weeks ago when we plan to continue with several more through the next year. I wanted to go back to the slide that Mary showed. We have had six of the previous colloquia and had an average of about 60 attendees in this room on-site. Very importantly, we also are providing this training for other scientists globally. We have an average of 250+ that tune into these by WebEx. This is a global enterprise. We have participants from over 22 countries that have participated by WebEx. I just wanted to use this slide also to congratulate our team both that SOT headquarters, and the Food and Drug Administration. We had one of these events scheduled for January 25, that was three days after the blizzard of 2016 occurred. The US government was closed. There was no way that anyone was coming to this room for these sessions. But the FDA, 80 staff, our staff a headquarters working from their homes. Three of them were able to pull off the training, the speakers were remote wherever they were in the country. They were able to pull off the colloquium for that day with about 250 persons participating by WebEx. So congratulate FDA and SOT headquarters staff for that. Look for more of these coming up. At least two more in 2016. Again, thanks to the joint committee between SOT, members and members at CFSAN here for organizing these.

It is my privilege to introduce the coordinators, moderators, for today's session. The first one is Ivan Rusyn. Many of you know him. He has been involved in with the colloquia series right from the start. He is a professor at the Department of Veterinary Integrative Biosciences at the College of Medicine at Texas A&M University. His lab focuses on chemical toxicity, genetic determinants of susceptibility, and computational toxicology. He is on several National Research Council committees, including the Committee on Toxicology, the Committee on Emerging Science for Environmental Health Decisions, and the Committee on Incorporating 21st-Century Science in Risk-Based Evaluations.

Co-moderating with him is Dr. Timothy Adams. He is here this morning. Ivan will be coming later in the morning. I was pleased to meet Dr. Adams earlier. He received his PhD from Catholic University and for two decades he served as a consultant to the flavor and fragrance industry. While he also held the position of professor of chemistry at Montgomery College which I assume is here in Maryland, an outstanding college in Montgomery County. He has served as the scientific director of the Flavor Extract Managers Association and was appointed as scientific director in 1999. He has in 2013, joined the Center for Food Safety and Applied Nutrition where he now serves as a senior scientist working on a variety of graph related issues. Dr. Adams we welcome you here this morning and appreciate your participation.

Speaker Introductions, Timothy Adams, US FDA, College Park, MD

Good morning. I just want to thank both SOT and FDA for providing this opportunity to present some information on the TTC concept which is fairly recent but also on the Cramer decision tree which I was involved in in 1978 so you have an idea as to my age. Today's first speaker is going

to be an individual, today we have an agenda that is the topical issues are the threshold of toxicological concern and the Cramer Decision Tree as I mentioned earlier. Our first speaker is someone who has had an appreciable history in this area. Her name is Grace Patlewicz or Grace Tier. She started her career years ago. She also moved and then went to DuPont in the US. She's a chemist and toxicologist by training. She will be talking today on the issue of past, present, and future for the TTC in the approach and regulatory decision-making. That will be the first one. After that, I will be discussing something in the advancement of food ingredient safety in terms of an expanded decision tree or expanded Cramer Decision Tree. We are working on that here at FDA. The third speaker will be Andrew Worth who will be talking about methods for TTC assessment, much of what he did under the cosmos program. And then finally the Quantitative Prediction of Continuous Toxicity Values using Chemical Structure Information. Jessica Wignall will be talking about that.

The first speaker will be Grace. She's authored over 85 publications and books. She's chaired various groups both here and in the US, and contributed to the development of technical guidance and integrated testing strategy and adverse outcome pathways for the REACH program, which has been in operation for some 15 years. Grace received the undergraduate degree in chemistry from the University of Manchester, her master's degree in toxicology from the University of [inaudible], and the PhD in organic chemistry from the University of [inaudible] in Spain. In December 2014, she took up a position as a research chemist at EPA, where she continues her research primarily focusing on refining read across approaches and using the bioactivity information. Grace, welcome. We are glad to have you.

Threshold of Toxicological Concern Approach in Regulatory Decision-Making: The Past, Present, and Future, Grace Patlewicz, US EPA, Research Triangle Park, NC

Good morning, everyone. It's a privilege to be here and I thanked him and thanked Ivan for inviting me to share about the TTC approach. I hope someone can help direct the slides. I had a nice introduction there. I haven't been at the EPA very long. I have to start off with my usual disclaimer that these are the thoughts and positions of me and me alone and not of the agency which I currently reside with. I had a few disclosures I want to share up front. As Timothy has alluded to I used to work for Unilever and DuPont. And now I am at the EPA and I met my husband at Unilever and I think he still has a few shares there. But I don't think there are many of them. Not enough to do anything. Otherwise I wouldn't be here.

How I wanted to structure my talk was I've been given the privilege to do the scene setting. I get to tell you more about the history and the background and how TTC has evolved. That's how I want to start off. Really showcasing or explaining how the TTC has evolved how it shaped in the thoughts and applications it was intended for peer then I want to shift slightly to start talking about as more questions became raised about how we want to deployed TTC and how it evolved over time. How it's currently use now. Some of the ideas we have in terms of how it could be used in the near-term future and also in the service future. I think Timothy will talk a lot about how some of the tools within TTC are evolving and he will talk more about the Cramer classes. And I think Andrew will be talking more about how it's expanding to other routes of exposure. I think that will be critical. I have a ton of references at the end. Since I've been asked to put some in. There are a few slides there just with citations and web links you can refer to for further information. I have one acknowledgment slide at the end. I always forget to acknowledge everyone that has helped me along the way.

So where does TTC fit in? I have these definitions I've pulled out in a paper from Susan Barlow it's actually a TTC monogram from 2005. It talks about TTC as being this principle about establishing a human exposure threshold value for groups of chemicals for which there would be no appreciable risk for human health. We can do this because it relies on accumulated knowledge and distribution of potencies of relevant classes of chemicals. That's a real mouthful. If we think about it, can we use a wealth of toxicity information that was generated for a vast number of chemicals? Can we exploit that to be able to set a threshold for below which we don't think there will be a concern to human health? That is really all that TTC boils down to. Can we put out what those [inaudible] are and define a line below that I don't think we will have too many issues. Can we use that to make some pragmatic decisions going forward?

So, what that means really is we could allow for an estimate of probability where we don't expect any adverse effects caring for substance of unknown toxicity at a given daily intake. This is really a pragmatic useful tool. If you have a situation where the exposure is really low for humans, for the human person, and you haven't really got any information for a substance you are interested in. This is a good way of being able to say I am not really concerned about this because the exposures are so low, there is no appreciable risk of human health. It's not a get out of jail card you cannot say I will use this to avoid doing any testing where is mandated by some regulatory program. It's not. It's not a way of getting out of doing what you should be doing for a normal risk assessment process.

In certain situations, this type of TTC approach is really sort of pragmatic and those sorts of situations are in its current application. If you have a food flavoring substance. The levels will be low. Is it a pragmatic way of being able to manage the risks associated with those substances without having to do a detailed risk assessment generating new information. If you have food contact materials and we will go into some of these examples and due course. If you have pesticides metabolizing and groundwater TTC is accepted as a pragmatic way of being able to assess those. You can imagine the situation. You have your active ingredient, you uncover you have certain metabolites. Certain ones may be hypothetical. You can't really generate enough to do a test for it. How can you assess and evaluate the risks associated with it without a pragmatic tool like the TTC? If you've got genotoxic impurities in pharmaceuticals. This is a pragmatic way of being able to address this. It's really a pragmatic scientific tool. Building on what we know about the toxicity information of a broad diverse library of chemicals and be able to make some informed pragmatic decisions.

So, there are two types of TTC. One is based on genotoxic information. I call this the general approach. It's a risk-based approach. The other approach is more of a structural based TTC approach that relies on accumulation of toxicity information that was generated in non-carcinogenic studies. If you have your 90 days in your 28 days and your [inaudible] and so on and so forth you can use that to build up your distributions of your NO(A)ELs and to find a distribution of TTC on that basis. It relies on an input of making structures and grouping things accordingly. We will go through both types of approaches.

So, you understand that TTC is a pragmatic tool and I can't reinforce that enough. It really is quite useful. You've got two types of approaches you can go down. One based on carcinogenic information which is more of a risk-based approach into generalized approach versus a structure-based approach where you try to do some pretty grouping of chemicals before defining thresholds from them.

So where does the start from? A long time ago. A guy called Frawley. He looked at a large sample, it may not be large by today's samples, but it was some 220 chemicals where he had

two-year chronic toxicity information on a diverse number of different types of chemicals. Food additives, industrial chemicals, consumer chemicals, pesticides. What he found was that for the vast majority of these, really the NO(A)ELs were higher than 100 mix per [inaudible]. For those substances and I know I have to use my mouse here. But for those substances that were under 10 or under one, he could account for those that we understood something more about them. Those 19 were heavy metal. We have information about them. For the five chemicals we knew something about while they were pesticides with known toxicity and we knew they affected the nervous system and the accumulated in the body. For the most part, if we exclude those we could set a threshold. What he said was maybe we can set a level of 10 [inaudible] per kilogram based on that distribution of those NOELs values he had categorized. This would add a bit of safety assessment to it. He divided by 100 which we were convergent and doing and said we can provide a level of .1 mg per kilogram and below that level we should be good.

Similar analysis by rule. He looked at the FDA Priority Based Assessment of Food Additives. And that had some 159 of compounds of sub chronic and chronic studies. He found that if you think about it intake of 1-10 µg per kilograms. It could pose a particular risk to humans it was quite consistent with the previous analysis. As that foundation built up, to what then became the threshold of regulation for food contact food substances in the FDA. That was passed in 1995. For substances that migrate and had some minimal migration in food that were not food additives per se, you didn't need to do any more than rely on this TTC value or the threshold of toxicological, threshold of regulation as it was known here. The threshold was set, and that meant basically a person over their lifetime could ingest enough of that substance, no more than 1.5 µg per day, over their entire life. Below that threshold you would need to generate any specific testing data it would be more of an abbreviated safety assessment that was being performed.

This value was derived this is like the general TTC I mention so this was derived from looking at carcinogenic data that had been pulled together by Lois Gold. And using a TD 50 dose to derive this risk based approach. There were a number of caveats according to this threshold of regulation. You have to know your substance wasn't already a carcinogen. That would be ruled out already. Your structure would not pose any immediate red flags that would suggest it would be carcinogenic. There weren't any impurities. Certain caveats but at least a very clear threshold that you could apply to food contact materials. There was some follow-up work a few years after that by Mitch Cheeseman. He looked at it to see if we could factor any other information, do we have structural information, do we have acute toxicity information that would permit a higher threshold to be applied rather than the 1.5. He put forward some proposals for those situations. This tiered approach hasn't been yet adopted by the FDA. It starts to bring to bear some of the possibilities that you don't have to be necessarily focused on one level. Maybe there scoped to increase those in certain situations. I think it's been a subject of some debate in the TTC world since then.

We have the general approach and now we want to move into the structural based TTC approach. Around the same time in the late 90s, Ian Munro and colleagues said can we pull together noncancer data from sub chronic studies together to look at devising sort of TTC values on that. Focusing on the oral route only, he pulled together some 613 substances and assigned them into different structural classes using something called the Cramer Decision Tree as I tend to call it. That was something devised back in the late 70s. He could assign things into these three different Cramer classes. These Cramer classes is based on a decision tree of some 33 questions. I haven't got a very good graphic there. This is pulled from the paper and I can only get it in a scratchy PDF these days. You have this interesting decision tree. You go through and it asks you a question on is this something in the body, yes, no, go down to the

next question. It's quite a challenge. I found when I was at Unilever and I talked to someone who wasn't a chemist and he had some struggles in terms of how do you walk through this decision tree to make determinations? It's not necessarily a trivial exercise. I think that's why what Timothy will be talking about and some of the tools that Andrew will be talking about will sort of help you navigate through this working through this Cramer structural class tree.

Nonetheless, Ian could assign things into these three classes. The three classes are you have low toxicity concerns via class I. There's the intermediate, we're not sure, it's not as good as class I, but there are a few flags being drawn. Maybe not something too bad. And class III, more complex structures. Metabolism to reactive products and some indication of potential toxicity. Three is bad, one is good, two is kind of in between.

What Munro did is that he found if he plotted the distributions of the NOELs for these classes he saw a divergence in the distribution plots. There was a separation that commenced with having been them so things that were indeed in class I had a different plot and distribution than those that were in class III. What he did from that was take the 5th percentile NOELs value and adjusted it for human exposure and divided it by 100 as a safety factor. That provides you with what we use quite commonly as these TTC values or the human exposure threshold.

The distribution plots look like this. For the one that is sort of closest to me is class I. The one is furthest away is the more adverse one which is the class III. You can see that separation in profile for those three classes. How that spanned out as you can start to see now sort of what the distribution of chemicals were within those classes. The class II is a little suspect I suppose for want of a better word. There's not that many examples that actually fit into class II. But for class I and class III it's clear we have quite a number of chemicals that fall within them and we can set good thresholds from them. I have typically always sort of thought of the TTC threshold from someone who names them. I'm always thinking of them in terms of what their per person per day limit is. I think common practices more looking at it on the basis of a kilogram body weight. So you can cause some confusion as I found subsequently depending on what someone is referring to. But these are the three classes.

I don't know if you can see that at the back. It will be in the slide pack afterwards. There was a lot of follow-up activity following what Ian did and he worked with another guy, they were part of in ILSI Europe group exploring TTC. What they pull together was how we could integrate the information we know about Cramer classes in a decision tree workflow. How do we use in practice? The first step is to consider is it even appropriate to apply the TTC approach? There are certain situations you really would not. Notwithstanding where if you have to, it's not your get out of jail card, but if there are certain situations where you've got a substance in its highly by cumulative. Like your dioxins. You wouldn't apply the TTC in that instance. There are other lodging aided structures that would lend themselves outside of the scope of TTC. What they did was to find some of these that you knew that if you had a chemical like that don't even bother with the TTC. It's not designed to be used in that way.

The next step was looking at whether there is any structural or potential genotoxicity. This brings us down to the same threshold of regulation the FDA uses in terms of defining the exposure threshold of 1.5 again. If there were no alerts, if we had a TTC of 1.5 we are done. We're good. If there were alerts, you have to have a consideration of whether they were any alerts that would lend themselves into thinking that these are high potency genotoxic carcinogens. Again that rules out using the TTC in the first place. This box that's labeled three, are these high potency carcinogens where we found that when there was an analysis made a

certain genotoxic parsonage and there were some that sort of stood out as these are really bad and we shouldn't consider them in the same way as the remaining structural things that exist.

It looks a little strange perhaps, but I have an annotation there, TTC which stands for the Threshold of Toxicological Concerns. There are exclusions that you take into account. If you haven't got one of these high potency alerts, then we could stop there and say have we got an exposure limit of greater than .15? If we have it, we're done. We can use that level. If that's what your exposure scenario really is, that TTC value will be good enough. If you don't have any alerts, and your exposure level for your particular scenario is greater than 1.5, what do you do next? The next step is sort of looking to see whether your substance may fall into the category of being a neurotoxicant like an organophosphate if you have something like that you need to use a different exposure limit of 18. That is where you fall down that line. I think originally and subsequently, this was glued together with [inaudible] and there is some debate about whether they should remain or whether the [inaudible] should be folded into the Cramer class III. There is something I will draw attention to subsequently.

If your substance is not an organophosphate, we can start to apply the Cramer classes themselves. And depending on what class you fit into that's what defines your threshold. It is quite comprehensive. You have a step? Can I apply the TTC? Do I have any genotoxic alerts to take into account? If they fall into this set of high potency ones, then that drives me down a specific risk assessment, if it doesn't, I can use a .15 threshold. If I don't have alerts I can consider whether my substance has got some neurotoxicity potential. If it doesn't, I go down the Cramer class III. It sounds complicated from the flowchart, but when you get to use in practice it's pretty straightforward. Again, Andrew has a tool that can help you work through this in a more systematic way.

I mentioned some of these exclusions before you start the TTC. And I mentioned dioxins and compounds there are some other exclusions to take into account. If something is by cumulative you don't want to use the TTC, we haven't evaluated it for proteins or steroids or metals and organometallics. Don't bother. If you have mixtures of substances where you don't know what the chemical structure is it makes it difficult to try and use the TTC. You don't know what the structures are to be concerned about. And also radioactive substances are excluding. That being said, there are many instances where a TTC approach could be applied.

As the discussions evolved in terms of how we can expand the scope of TTC, that is where some of the questions start to arise as well as refinements that need to be done even with the tree itself, what the representation of the underlying data set was, and Andrew will talk more about that. But also some questions about the workflow itself. Given the number of chemicals that fall into class II, do we need a class II? I think that is still up for debate. Carbamate and organophosphates are grouped together under a TTC threshold of 18. Maybe the carbonates are sufficiently protected under the class III. The TTC approach overall is based on an oral route of exposure. As the discussion went on in terms of how we can use TTC, questions around what is the utility, can we apply? Can we come up with the TTC approach that is relevant to other routes of exposure? For instance, the inhalation routes of exposure. There have been interesting work by Sylvia [inaudible] back in 2009, 2010 looking at building up an inhalation database, repeated exposure inhalation to say can you use TTC values generated from those specific data sets. Andrew will talk about the cosmetics project. There the interest was from the cosmetics industry looking at can we have a TTC value that’s relevant for a dermal route of exposure? There are discussions on how you can extrapolate from the oral route to the dermal one. Is that a strategy or do you need to derive a specific dermal database in order to derive your TTC value from that? That is something he will allude to.

There has been other work in terms of looking at TTC for other specific endpoints. The [inaudible] back in 2010 looked at can we have a TTC specific to prenatal development toxicity? Is the existing TTC threshold sufficiently protective using the more general data set that is being compiled? I do want to touch on the skin sensitization because it's not an area really covered by the existing TTC. I have a few slides that talk about some of the work that Bob Safford had done. Building on and using the same principles of TTC but could be defined thresholds for skin sensitizers.

One thing I said all along is TTC is based on a lifetime exposure. Specifically, in the pharmaceutical world. Maybe there are instances where a shorter exposure period can be considered. If TTC is assumed for lifetime exposure there are situations where you could have a higher TTC value for a shorter duration. Maybe less than a year. I forgot to mention in that workflow when I had there was a proposal saying if you have something that has a genotoxic alert, maybe for a duration of year or less, maybe you can get away with a higher threshold if you had some additional information like a negative test, maybe there's a scope to have a higher TTC value. I think the same conversation was going on in the pharmaceutical industry. If we were looking at genotoxic impurities where the risk benefits for a particular pharmaceutical can be more clearly articulated, could you have a higher TTC value because the scenario could be you are not taking that drug for a lifetime. Perhaps it’s something like an antibiotic you are taking for a shorter duration.

There have been proposals and things that were accepted in terms of having a staged TTC value. I want to show a couple of slides for that. Within my own industry in the past, we also looked at could we use TTC values for trying to set some kind of occupational threshold limit where we are dealing with workers are maybe something is for a shorter duration where we manufacture something for two years, can we use TTC in that scenario rather than thinking over lifetime exposure for 70 years or so? Could we tolerate a risk of [inaudible] instead of the usual one in 1 million?

There was a guy called Mueller. I forget what company he worked for. He put forward a proposal of this staged TTC of if we look at a year duration, or less than one month, can we have different TTC values? I've pulled this table out that there is a reference in the back that alludes to this. What is been agreed, so under the ICH M-7 there is agreement to use a staged TTC depending on the different scenarios. The value is different than what was proposed originally in Mueller. And under the European Medicinal Agency there is also a staged TTC. There is prior art in terms of being able to use these values for shorter durations.

I mentioned Bob Safford that had done some work on TTC and skin sensitization. I'm just going to rush through in terms of what he had done. There was a series of papers that came out. I think it opens the possibility of applying the same principles but still using TTC elsewhere. He wanted to investigate could he start the TTC approach but in the interest of developing new acronyms he used dermal sensitization threshold or DST as the threshold in this case. But could he define a threshold below which there would be no appreciate bold risk of sensitization. He followed the same principles for the threshold of regulation. He used what was known or used to be the new substances database ELINCs at the Joint Research Centre’s European Chemicals Bureau and we pulled out all the information we could on classification and labeling looking at what had been classified as the sensitized or not and could we use that to define what the probability was of something being a sensitizer. Then he looked at the distribution of sensitizing potencies. Using what we had called the master table that was published in 2005 for a set of 211 chemicals where we had local lymph node data. And we could define from that potency

threshold that's defined in the lymph node we could define a human exposure limit. And then convert that to an acceptable exposure limit using a different assessment factor depending on what the product type was. What he was interested in was can we define this and adjust it depending if we are interested in the skin cream or hand lotion or shampoo? And so forth. He wasn't interested at the time to define a specific threshold, but could you define different thresholds depending on the use type?

Subsequently he did more refinement on a larger data set and started to define things based on structural classes. But rather than use the Cramer structural classes we use some what we defined chemistry domains. We could been things according to what we expected the reaction chemistry and how that is important for sensitization to be. And then look at defining these thresholds from them. He defines one really looking at one particular threshold depending on whether something would be a sensitizer.

In subsequent work he did more work in 2015 looking at can we define thresholds for substances we know are sensitizing. Think of it this way, the first scenario was looking at can we define thresholds for specific product types, if we take into account different assessment factors. The updated it to look at can we to find a threshold like a threshold of regulation for skin sensitizing information and then can we refine that depending on what we have whether something is likely to be a sensitizer or not. It's like have you got an alert for sensitization, if you have then you've got a different threshold and if you haven't, you have a higher threshold you can use. That was that type of approach.

These are the reaction domains. We bend them in different mechanistic classes, whether something is [inaudible] accepting reacting or whether it shift space and so on and so forth. It's akin to something like the Cramer classes but grouping things that are more appropriate for the endpoint of interest.

One other step that was taken and this was just last year, and this was akin to what you have in the ILSI structure based TTC approach. A similar analysis was done for certain sensitizers if they are really reactive and really sensitizing. And a TTC approach would never be applied and exercise was be able to identify those particular ones. And there were a couple of samples alluded to from that set.

One last thing. One thing we presented at SOT this year was looking at if we have the scenario where we are trying to prioritize many chemicals, thousands of chemicals, is there a role that TTC can play to help prioritize those types of chemicals? What we had done was look at the predicted exposure values published by John Wambaugh and can we use that to look and compare against the TTC value to say are we comfortable that these are the chemicals we should focus more attention on? Or perhaps tear down what we do with the remaining compounds. This is a graphic from his paper from 2014. We have a ranking and prioritization of some almost 8,000 chemicals with respect to predicted exposures. And where we had data and it overlaid. We re-plotted this and the other graphic accumulative frequency by the exposure production. If you look here, these are the chemicals, this line represents the Cramer class III threshold. We haven't done any other clever refinement or processing it through the workflow. If we just say here's our TTC from Cramer Class III, where should be focusing our attention? It's really a false start or proof of principle about where we could go forward from this. It's really some early work that we have done just exploring the idea and the utility of the TTC.

What we found was less than 5% of that is the 95th percentile was above the Cramer Class III, we haven't screened them so we have any alerts, we haven't screened them whether they are

organophosphates which is quite likely, but at least it gives you an idea where should we focus the attention. I know this graphic didn't come out too well, it will be a lot clearer in the slides. I have included a web link for the actual poster you can read for yourself. We kind of try to embed the same, in some fancy cartoons, the workflow that was pulled together to try to prioritize what could your steps be in terms of looking at the TTC and how that could then inform you to focus more attention on that 5% in terms of looking at what bioactivity data we could generate and how we could refine the risk assessment going forward.

Take home. Because I'm out of time. I hope you can see the TTC is a pragmatic means of prioritizing the testing of chemicals with the assessment of chemicals where you have situation of very low exposure, and when you really don't know anything about the chemical to begin with. It doesn't overrule traditional risk assessment. It's been established I think for oral exposure and it's something that's continually being evolved and challenged but other routes of exposures and applications beyond food contacts and additives and genotoxic impurities and pharmaceuticals. On that note, I want to finish with my, I said I had a few reference slides. My acknowledgments and really, Bob Safford was the one that got me interested in this which is then when I went to the JRC I begged Andrew to help develop the tool which is what he will talk about. And some internal discussions based in DuPont where we sort of resurrected the idea of TTC and identifying ways in which we could apply it for business specific questions. And on that note I thank you for your attention.

Adams: We have time for one question if anyone has a question related specifically to the slide presentation.

Advancements in Food Ingredient Safety Prioritization: An Expanded Cramer Decision Tree Schema, Timothy Adams, US FDA, College Park, MD

My discussion today will be on something that really started back in 1974. It was odd, but I worked with Greg Cramer, and [inaudible] the authors of the decision tree article. I thought I would take a step back and first of all I have no conflicts, that I know of, there may be some, but I have no conflicts. Second, I take the perspective that Dick Hall used to push forward. Here we are, we have a substance water hemlock, I don't know if anyone knows it, but it's one of the most toxic plants in the United States and in Europe. It's all in North America. It has a substance in it called [inaudible] if you rub it between your fingers and put in your mouth is lethal. Within two seconds exposure, five seconds exposure is long enough. This has a NOEL that's been estimated and somewhat verified that it's somewhere around 1 µg per person per day. That represents a quadrillion molecules. The other thing he always compared it to was water. He said in terms of water, the safe level is about 4 L. If you go beyond that and you start to have electrolyte solution. There you talk about 100 billion quadrillion molecules. He said as you can see there's a wide variation in terms of what man can endure, but they always have some threshold. The idea of threshold has been bound up with me ever since 1974.

Today I want to present to you a short history of the decision tree and how it evolved into TTC, some application of it which Grace mentioned mostly, some current and past issues of the TTC and the decision tree in terms of the formation of it, the structure base, questions about its origin and its underpinning, and then a proposal for a new expanded decision tree, one that we have been working on and off in the industry for some time but now at FDA for the last three years. I also want to talk about how we developed questions for the decision tree. Something that was never explained in the original article. We're going back to the fundamental basis upon which

the decision tree was originally developed. And then talk about how we might update it. Which is a little different from the current approach taken by most.

Just stepping back to 1958. In 1958 when the amendment to the Food, Drug, and Cosmetic Act came out, about a year later the flavor industry in order to comply with the grass regulation, sent a list of about 1100 substances to FDA and identified them as being in use in the United States. FDA's initial response was we will take a sampling and do the toxicity study, acute, sub chronic, and chronic, on a representative sample of the most highly used flavors. Because this would be part of what was the new grass regulation. Guys like Hagan and [inaudible] and Long, during the 60s, and early 70s, actually did comprehensive studies of flavor ingredients. The government's resources were quite limited. And so they have another approach. They approached FEMA with the idea of collecting all the scientific information related to flavors and flavor safety in a contract which they signed with FEMA, the Flavor and Extract Manufacturers Association in 1974. And in 74, the industry and university people from Catholic University where I was, and University of Maryland work Greg Cramer was and Catholic U where Dick Ford was they started to research the literature and put together a group of what was 65 SLR's containing information on 1100 materials. Those 65, each one of those 65 SLR's was really organized according to chemical structure, the available metabolism in pharmacokinetics and the available toxicity data at the time. And this was then submitted to FDA and as a sidebar to all of this, the intelligence gain from covering 1100 substances and treating the scientific data for those, there were thoughts of how can we put this into a more rational approach and that's how the decision tree was born. We saw that when you treat it [inaudible] that indeed the ketones there were some underlying basic toxicity and metabolism that could link them all together. This is what turned out to be questions within the decision tree.

Now of course Grace went over the Cramer classes and as you can see them there you have the three NOELs with the TTC values that she listed. And of course, the underlying feature was that if your intake was above the threshold you either required more data or asked additional questions or if it were below the threshold for that class it was considered safe. This was the whole flavoring substances and began in 1996 and it continues on today. And there been more than 2400 substances subjected to safety evaluations using this methodology.

Of course, at some of the, Grace had a similar slide, but this is just the different types of applications of TTC and the decision tree. I need not cover that she did a good job with it.

What I've always thought was important was the analysis over time of the Cramer decision tree which was published in 1978. Initially Phillips in 87 published an article in *Food and Chemical Toxicology* that said there were certain issues here and we need to include more functional groups, there were certain questions that were concerning questions that aren't structure related, they are related to source. This is the common component of food and they certainly pointed this out as being a weakness of the decision tree. Later publications by Andy Worth who's here in the audience and others point to some of the other information that was missing that would help to understand the Cramer Decision Tree better and document the underlying or underpinning data that allowed the questions to be developed. Some of the basic critiques in the analysis is what you need to eliminate non-structure-based questions which I will point out later. You need to specify the metabolism and toxicity underpinning each one of the questions in the decision tree. You need to update the question to reflect current state of science. Because it is now 40 years old and you need to document mode of action and species difference which might be improved or might be evolved. These were comments that certainly were made and in 1999, 2000 the flavor industry thought about developing a new decision tree. In fact, [inaudible] and

Dick Hall said we need to put an effort into actually expanding the decision tree and bringing it up today.

So, my objective at FDA is to continue doing a scientific update on the Cramer Decision Tree. To increase the chemical space that the decision tree covers, we'd like to apply it not only to flavors and cosmetics but make it applicable to food contact material, the space and pesticide residues etc. We would like a broader application. We like the expanded decision tree database, one that integrates structure in a more comprehensive way. That means we have a group of [inaudible] or [inaudible] that lead to a reactive intermediate that lead to a specific neurotoxic response. We want to tie that together in a question. So we want to track the molecule through the body, and point out the toxic endpoint, the intermediate that caused it or resulted in it and that was part of the objectives of the decision tree. Also, to identify the lowest NOEL just as Ian did in his work with the original TTC, and to pick it for the most sensitive species of the longest duration study and make sure that species was relevant to humans. Now part of that was to institute a scrubbing method. That is let's scrub the data to eliminate things like alpha-2 globulin that isn't relevant to man, let's really discuss whether or not for stomach effects are relevant to humans when they are seen in animals. This idea scrubbing is something that we want to take seriously. It's not just a casual scrubbing of this or that. It's a scrubbing of the data to make sure we identify truly relevant data for humans.

Part of the criteria we developed, and this has been in use here for the last three years by individuals working on it, would have a duration factor, in some respects the same as Ian used, we are using four for studies of 28-90 days, we are using three for studies of 91-98 days, and one for studies longer than 98 days. These are correction factors. We have a 90 day study. Whatever NOEL we developed for it we divide by three. You can argue there should be other factors. Some use six or 28-98 day studies. It turns out that when you compare the 90 day and 28 day study in the NTP program you will see a factor of three is conservative for 90 days. A factor of four or five is conservative for 28 day study. We have relatively few 28 day studies compared to 90 day, so in the end, you will see it doesn't make a great deal of difference. If it's a repro study we will use it only maternal or paternal affection study we will use duration factors but if it's a true development reproductive effect we will not alter it, we will use a factor of one. Relevant species we will pick the most relevant species based on relevant factors that is in alpha [inaudible] such as a mode of action in if it's known in the bladder if the female or males are more sensitive. We can actually go to the data and to identify a mode of action. If you use single dose NOELs if the related substances in a particular chemical group have similar NOELs. We will pick a single dose NOEL if needed. We will replace NOELs from the Munro database. If there is a more rigorous study that has a lower NOEL that's been performed since 1996.

Also, an objective is to expand the chemical space by increasing the number of NOELs to approximately 2000. That's just a target. And to make structural variation more important than the number of NOELs. That is, we are not interested in collecting 300 NOELs on every organophosphate, we're interested in a variety of structures that are organophosphorus compounds that provide a basis upon which to derive questions for the decision tree. Therefore, another aspect of that is if you pick 300 organophosphates you are overrating a particular class. That really is not what we desire. We also would like to validate the NOEL data by consulting authoritative bodies who've already published on those NOELs. Therefore, we will go to [inaudible] to FDA, and we will look at the NOEL for that study for that particular compound to see if it agrees with what we have derived. I would say maybe 80% of all the study NOELs have been at least agreed with in an authoritative body. Another aspect is that we will use [inaudible] per kilogram per day really because some of our compounds differ from other compounds by

order of two orders of magnitude. That is 100 times difference. This would make in effect, if we use milligrams per kilogram per day. These are all part of our criteria that we use.

Here's an example of how the thinking works. We tried to create a biochemical map of a substance. And then look how changes or functional groups perturb the data we derived. Here are three substances. Furan, which gives you a reactive intermediate which reacts to the protein the target organ is liver and from what we derived from NTP studies and other literature the data the NOEL is somewhere between [inaudible] per day. That's fairly toxic. However, if you put a metabolic handle on furan, like Furfuryl alcohol, you get as a metabolite and if the glycine conjugate and it's readily excreted in the urine. The target is also the liver. When you get the high enough level you will see very similar types of response to the organ as with furan. The problem is it that a dose level which is much higher. Therefore you're looking at competing pathways. You have what is called the ring opening of [inaudible] on one side which is intoxicating the other hand [inaudible] is detoxicating. And that competition changes. Eventually when you saturate the level of acid, you bring it up to a high enough concentration and you exhaust the glycine pathway and then ring opening becomes important. However, when you move it to put a [inaudible] group onto it and change it now you're looking at a molecule in which sulfur dominates. They react with the blood and you get hemolysis and bilirubinemia and also, it's more toxic. The questions you develop in your decision tree must differentiate between these three molecules and that would be the basis for development. We are looking at very small changes but that's enough to differentiate.

Here's the Cramer decision tree versus right now the preliminary expanded decision tree. The number of questions in the Cramer Decision Tree are 33 with the structural decisions of about 50 because some of the questions in the Cramer Decision Tree are multi-part questions. And for the expanded decision tree we have 47 questions with about 127 structure decision points. We've at least double the number of decision points in terms of chemistry to better define, better specify in terms of chemical structure in the relationship to toxicity. Of course, Cramer had 247 in his original database. We're now well above 1800. These are single NOELs for individual chemical structures of a wide scope or wide range of structures. And then in terms of the number of classes the original one was three. The TTC's were of course three because of the three classes that were defined, and now you're looking at six classes and six TTC's. We are increasing that simply because we have more information and we have more knowledge of specific differences between molecules and therefore we are better able to classify in terms of toxic potential.

Here is the old Cramer Decision Tree.

That network you, that didn't work. Can you get me back? Okay, back one. Back one.

The original Cramer tree as you can see step one, that's class I, that is the question that many objected to because it is a question on is a constituent of the body. And of course, we know there are many substances that are constituents of the body that are toxic at low levels. You have formaldehyde in your body, etc. With these compounds certainly would the put in class I and that was a common criticism is that they should be there.

Another one if you look in the Cramer tree just to pick out a few points, you see question 22 down in the lower right-hand corner. That particular question is, is it a common component of food. And you see how many questions come down through the tree and wind up at 22. So you are facing the final decision on whether it is in class II or class III on whether it is a component of food, and that really is not a structure-based question.

And so, some of the comments that were made really deserve, they really have merit and deserve to be addressed. And so, this is why we have these trees by the way, the line it is divided up. It is not important that we go through each step. But what you see is really [inaudible] being dealt with differently in the tree.

Here is what the new tree looks like. Far greater complexity. One is now a question about food nutrients. That is, does the compound for carbon dioxide [inaudible] and pneumonia, etc. That is, is it a compound that you eat and it enters the fatty acid the pathway, does it enter the [inaudible] cycle, [inaudible] which makes a lot more sense. But we now have questions like two over to where we are doing with phosphors compounds and we differentiated between those organophosphorus compounds that have different degrees of toxicity. Depending on certain electronic factors, [inaudible] factors, certain factors in the molecules that we are differentiating, and we actually have different classes for organophosphorus compounds. So, we're dealing with the difference between an organophosphorus compound that is used as a pesticide, versus an organophosphorus compound that is used in food contact material as an antioxidant. Which provides a better correlation of the real world. And you'll see our tree has far more complexity than the original tree, but we have enough data to support this.

If you look at the bottom part of our tree, you see that in blue, the question 28 to the question 47. 47 is at the end of the tree. You always default there. That question is really have to do with is the compound absorbed or not absorbed. If it is not absorbed then excreted from the body easily, that goes into group 1. If it is absorbed and we haven't had any other place to put it, it winds up in the default category which is class 4. So, we are able to deal with all chemical substances that have a defined structure within this tree. So, we are not just passing off organophosphorus compounds or [inaudible] into a particular subclass and applying a factor, we're looking at it in terms of real data with real underlying metabolism and toxicity.

If we look at the classes themselves, you see class I as I said is a class that is simply non-toxic. Essentially food. Class II is also fairly benign, that's fairly low toxicity because these compounds participate mainly in phase 1 and phase 2 metabolism of detoxification. The readily excreted from the body and only at very high levels do you begin to see differences between the man and animal and the flip over to intoxication pathway. So, I'm talking about menthol, like menthol in mint, cinnamon aldehyde, etc., and that's a class that really has considered more or less pretty safe to consume.

We have class III that's intermediate toxicity. Some of these show actual species differences between man and animal. And I will talk about that in a minute. But things like polysulfides, furfural, but certain dose levels, you get to a dose level where you start to differentiate, and you will find out what the rat and the mouse do with it is different than what man does with it. At least from the information we have. So that's why I have animals doesn't equal man.

If we look at classes IV, V, and VI. IV is what used to be the old class III, Cramer class III, and that has structural features of the parent or metabolites that suggest toxicity. Materials like acrolein, which is a well-known substrate, and pyridine, compounds like this.

We also dealt with high molecular weight molecules in the next class, class V, and now we get toxicity at very low dose levels. Down to 10 to the -3 to 10 to the -5 mmole/kg. To give you a perspective, now we're talking about somewhere between, where something has a no effect level around one or 2 mg per KG which is considered pretty toxic. [inaudible] which is a pesticide and fumonisin B1 would fit into class V.

And then class VI. Typically, these are very potent organophosphates substances that would certainly be considered for chemical warfare, naturally occurring sodium and channel blockers and toxicants that are really potent over short intervals. But we are trying to cover the full chemical space from compounds of molecular weights 3000 down to compounds of molecular of 30. We consider about 15,000 studies at this point and we've derived about I'd say 1750 NOELs but we're still looking but we're in a state of a state of shall I say 95%.

Now this gives you an idea of the preliminary TTC. You notice the bottom line is the Cramer class. Cramer class I, II, and III. And you see how it is divided in this presentation. I've divided them up so class I under the expanded decision tree and class I under the Cramer decision tree is a significant difference between the log for the TTC value. That is class I under the expanded decision tree and class II under expanded decision tree are really contained a lot of materials that would be class I in the old Cramer decision tree. And you see that the TTC's are significantly higher. That is to be expected given the fact you're asking better questions and you are pushing molecules across a greater number of groups. And that makes a lot of sense.

So, class III is equivalent to our old Cramer class II which was sort of the one that has no substance in it. Now that particular class has about 250 substances in it. Distribution of substances across all six classes are about 250 for number one, for class I, class II about 300, class III about 250, class IV about 700, class V about 150, and class VI is fairly small, as you might expect there's only 50 or 60 substances in it. So, we're looking at this to sort of a better set of question, more structurally based, being better able to define the toxic potential across a broader space of chemicals.

This is pretty much as you can see when you get to class IV, it is a little bit or it's in the same range as the old class III. And class V and VI of course are much lower and they should be, now we're talking about really toxic materials, both natural and not.

So, where are we here? If we go through this, there's some basic steps that we take. We just don't take the data and let it fall out. You need to organize the data in some way so that you can develop your question. And so you look first at the range of structural variation within a chemical group. You look at the RAM and you look at Durand with five carbons. You look at it with two carboxylic acids, one carboxylic acid. To see whether there's consistency among target organs, where there's consistency in terms of change when you add metabolic [inaudible] which are intoxicating versus detoxicating.

So that's the first step. And then you try to identify metabolic pathways that are present which are the detoxification pathways and the competing intoxication. And ones what would be how do they compare to those in humans.

The other is the effect of additional functional groups and structural changes and what happens when your skeleton changes, you go from five carbons to 15 carbons. So, we're looking at that. And then we're developing the structure-based questions that account for these metabolic options and toxicity, and must be consistent with the NOEL changes that we're seeing among a group of materials. No one NOEL is ever that important. It is the group of NOELs for this particular group that has merit. It's because there's too many variables. The only independent variable you have when you are collecting this type of data is the structure of the substance. That is absolutely known. So that becomes independent variable. A toxicity study always involves one thing, it is dependent because it is dependent upon too many variables. You can never repeat the NOEL for a study repeating the same species, the same animal [inaudible].

Guarantee. So, I think that this type of approach does have merit. And this is our approach to question making. [inaudible]

I am over a little. May I go through this part?

Let me just show you how we think about this. And some of my colleagues are in the audience. But this is our thinking on this. There's a simple substance, a group of phenols. We have a group of 50 phenols. They are compounds that could either detox, they can undergo phase 2 [inaudible]. This is also hydroxylation product. The hydroquinone can undergo phase II conjugation which are detoxification pathways. Or it can be oxidized to a quinone, and a quinone is more reactive, and of course will react with [inaudible] and activated with stress. It certainly is a target for this sort of thing. So, we're looking at here competition between detox and intox.

We go to the next slide. The analysis shows you materials that form sulfates or glucuronide. They are easily conjugated and excreted. On the other hand if you look at compounds that are a little more difficult to conjugate. These have large groups, [inaudible] groups, it becomes difficult to conjugate them because they are blocked. And so, these compounds tend to go towards the quinone. That is, they will head or be metabolized more to the quinone because they can't be conjugated so here if you look at the NOELs, there are two orders of magnitude less than molecules before him. That I considered before.

Now here's how species is involved in it. These are species data. If you look at the species differences, here is one of the ones that blocked, that’s the 2,6-di-t-Butylphenol. That has good data. Some of these molecules do not have this type of data but this is good data. If you notice the rat, the rat’s not a good conjugator. The rat tends to form hydroxylation and eventually, it goes towards the quinone. On the other hand, humans, humans are very good conjugators but not that good hydroxylators. So what does this tell you about sensitivity? The rat’s more sensitive to intoxication than the humans. So, the human’s a less sensitive species. If anything, the rat needs the safety factor, we don't.

If you look at the comparison of the two, there's two groups of compounds, you see that there is [inaudible] data which agrees with metabolic data which allows us then to come up with sort of a recommendation and this is very preliminary. But here are the, in the column after the name, that’s the Cramer decision classes. They are in three, two and one for the very hindered compounds. But all the rest of them were in class I in the Cramer decision tree. In the expanded decision tree, those three blocked materials, the blocked materials I’ll call them, they are in class III whereas all the unblocked ones which conjugated very easy, they’re in class II. So you see the differences in NOELs but [inaudible] magnitude less than the NOELs for the [inaudible].

But then, if you apply a species-adjusted factor, that is, instead of using 100 you use 10, then you see that there is an increase in the overall NOEL. The lower NOEL for the substances. That would put them in the correspondingly different classes. And I think this is something that goes against the grain of most of us, we would rather include safety factors to protect us rather than remove these artificial factors with a factor of 10 by actually having good data to justify. So, I am sort of engaged in this. With some interest. And we're going through the various groups and we plan on publishing this sometime around the end of the year.

So, my conclusions are number one, new science allows for greater specificity and questions relating structure to toxicity. ADME data underpinned EDT questions that link structure to endpoint toxicity. Inclusion of more elements and moieties increases the applicability to the EDT

to greater scope of chemicals present in food. And EDT better delineates TTC classes compared to Munro classes and provides a better understanding of structure activity relationships.

Thank you and those are some of the references. And I like to think Szabina Stice, Antonia Mattia, Sylvester Mosley, and Renata Kolanos all of them of which have worked on this project. And a big thanks to the Flavor and Extract Manufacturers Association for their cooperation with FDA on this particular project. Thank you.

It is time for a break.

Audience Question: [inaudible]

Adams: We will start at 10:30. Sorry.

Rusyn: And I think it was very educational because we do strive to make sure that these are continuing education like activities. And our next speaker comes to us from the European Commission Joint Research Centre. Dr. Andrew Worth is one of the household names in the structure activity relationship modeling. He is a very eloquent speaker, he got his degree in linguistics from Oxford University, so we will have a chance to witness that. But Andrew has really been one of the leaders in not just developing quantitative structure activity relationship models but actually implementing them into regulatory decision-making. So, I am looking forward to your talk, Andrew. Thank you.

In Silico Methods for TTC Assessment, Andrew Worth, European Commission, Joint Research Centre, Ispra, Italy

And thank you, Ivan, for that kind introduction. It's really nice to be here again, please bear with me. I am losing my voice. So I will try to do my best to get through this.

So, I would like to change gear now talk a little bit about *in silico* approaches to TTC assessment. [inaudible] A brief reminder about the decision tree and how that is basically the basis for this approach. I really want to explore two themes. One is how computational tools can be used for the application of TTC approach. And the second is how computational approaches can be used to integrate new science. Then I want to talk a bit about where the field is heading in terms of the issues of route to route extrapolation and oral to dermal extrapolation for cosmetic substances.

So, as you heard, the TTC approach essentially [inaudible] approach based on the assumption that there is an exposure level below which adverse effects are not expected to occur. With these effects we have thresholds that relate to the three Cramer classes.

These are embedded in the Kroes decision tree along with a load of additional questions that are also structure based. For example [inaudible] neurotoxicants and so forth. So, this is another representation of the decision tree that Grace showed you earlier. And as you see already there where there is potential for the structure-based method because a lot of questions are effectively relating to structural features.

So, the first question that we really need to address in this approach is the likelihood of genotoxicity and carcinogenicity. And it’s advisable to use two kinds of approaches, really. And

there are very many models that are publicly available. Some of them are based structural alerts. Some of them are based on [inaudible] but there are essentially two kind of approaches. One is kind of knowledge-based approach with human knowledge represented in the form of structural alerts associated with genotoxicity. And the other statistical-based approach derived from machine learning.

I'm showing some examples here of some of the models that are publicly available. The first model is the CAESAR Mutagenicity model [inaudible] now available as part of [inaudible] software. And these are very easy software tools to use. It's advisable to use the two kinds of approaches. The structural alerts-based approach the tendency to overpredict genotoxicity whereas the statistical models have a tendency to correct to those overpredictions.

Another resource which is really worth looking at is the Danish (Q)SAR database. And this is basically a database of QSAR predictions but it's very large and it's searchable in multiple ways. So, there are (Q)SAR predictions for over 600,000 chemicals based on 200 (Q)SARS including some commercial tools, fairly accessible from the Danish websites as you can see the bottom of this page.

So, a large part of the findings in the decision trees relying on the Cramer decision tree. And like you heard before this is the set of questions for which there are yes and no answers. And they basically take into consideration structural features associated with toxicity along with a limited number of considerations related to metabolism and elimination and natural occurrences of substances in the body or traditional foods.

So, they have a great idea that they should implement the Cramer decision tree into a software [inaudible] and we call the software form Toxtree and this has been very useful. From the development in the Toxtree in the additional row bases have been added to the software. So, it is much more broadly applicable tool. You can see the software basically the top left this summary of the compound properties. Bottom left you can see there's a representation of the molecular structure. The prediction is given in the top right. But what is also important is the reasoning behind the predictions. So, the actual series of yes/no answers to the questions ultimately leads to that prediction. And this is important because the experts can evaluate whether or not they agree with this particular prediction based on that line of reasoning. You can download the software from source Forge. And you can also use it online. And it's worth noting that there are a number of [inaudible] within the software [inaudible] in terms of applying the [inaudible] approach. The most commonly used one is the Cramer scheme. There is another decision tree called [inaudible]. It is not very widely known but it has six additional questions on top of those in the Cramer scheme. The basically correct some of the false positives and false negatives and it also has an extended database with natural components of the body.

In terms of genotoxicity predictions, there are [inaudible] nucleus and [inaudible] decision tree is also implemented. So, you have [inaudible] exposure estimates, you can insert that along with the chemical structures. It will [inaudible] prediction and it gives you an indication of what that is safety are not.

Another tool that's freely available is the OECD QSAR toolbox. This is a very complex tool that allows you to do grouping and read across and it has within the workflow a number of profilers which are essentially structure-based prediction tools. You also have the Cramer decision tree and the Cramer with extensions. These have been coded in a totally independent way compared with the software and that leads to some discrepancies in the results [inaudible] with pieces of software. And I will expand upon that in just a moment.

Here's an example of flavoring. There some opinions that provide some estimates [inaudible] want to do a TTC assessment on the substance. The first thing you do is you would apply a rule-based or a QSAR for genotoxicity. [inaudible] you will identify immediately the structural alert [inaudible] in the saturated part of the substance.

Now you know that the TTC for genotoxic carcinogens is 1.5 µg per person per day. So there may be some concern. Then you look at the example and there are 11 [inaudible] in the Danish database. You find that most of them, 10 in fact, are negative, and there is one positive prediction. So, you might think, okay, I don't actually think this is a genotoxic carcinogen. I think we have TTC value. But you may also wish to look experimental data or generate some experimental data and in type you can do that. You will see the opinion experimental *in vitro* mutagenicity and chromosome aberration data are negative and also experimental data for *in vivo* micronucleus test is also negative.

So, in a sense we have concern for [inaudible]. Now we apply the Cramer classification. And see the toolbox. And the TTC value for Cramer class I is 1800 micrograms per person per day. To the estimate for exposure is less than the value when they see as being consumed under safe conditions.

Now I would just go through this process in a bit more detail if you are using the Toxtree software. So, the first thing you need to do in Toxtree is select the [inaudible] different decision trees which one you want to use. Have Cramer extension or the [inaudible]. And that's quite straightforward.

You can also view the decision tree, so you can see the full decision tree. You can click on the different notes in the decision tree and get a summary of the question that's being asked of the node. You can get examples of substances for which the answer to the question is yes or the answer to the question is no. They are very useful and [inaudible] in the [inaudible] tree.

Next and most important thing is finding the right chemical structure. You need to enter this into Toxtree in order to get the prediction. The way I always do it is go to the quality assurance source chemical structures such as ChemSpider. If you put in the substance you can type in the name, you get official representation of the structure and you get a number of identifiers. For example [inaudible] is a one-dimensional representation of the chemical structure. Or the energy or entity are also one-dimensional representations of the chemical structure, so you can take the smiles and you can insert that directly into Toxtree to make a prediction.

So you basically copy paste that into the top line, that's the smile and you press the estimate button. During the Cramer decision tree in the get the prediction that's low. Cramer class I represents ingredients and you also have to change the reasoning. You have an explanation that goes through [inaudible] with dialysis or you could see particular [inaudible] basically the one no, two no, three no, and so forth. And it's worth recording that in case there are differences of opinion later on with the prediction and expert judgment.

And finally you can save and export the results depending on whether you simply want to view the results in a PDF file or whether you want to import them into some other computational tool with the computational processor.

So that was quite a simple example, one substance. Another kind of scenario that might arise is evaluating in the metabolites in the [inaudible] active ingredient. For which experimental data is

not required. And in fact, European Food Safety Authority recently published a guidance document on how you might approach this using the TTC concept, using QSARs, using grouping and read across. [inaudible] and your comments 2 May from the website.

And what's nice about this as it provides the work flow on how you integrate TTC. And within the appendices of the guidance document you see a few words of examples, so this is one example of [inaudible] for example, which has a whole load of metabolized substance not just in ground water, but in plants, in cattle, and so forth.

So, you have before quite a few evaluations of the Cramer scheme and [inaudible] the Cramer scheme. And essentially these evaluations in the focus on the ability to predict NOEL values, and whether there's overprediction or underprediction. And how the tools implement the original thinking behind the Cramer scheme. But the [inaudible] of all of these studies is that there is a need to update the Cramer classification scheme. Some authors advocate very minor modifications, simply clarifying the questions, others a total overall, and others perhaps as we just said in the previous presentation something in between.

So, here's one example of how Toxtree and the toolbox may come up with different predictions for a given substance. So you have this bicarbonate. If you process this through Toxtree, the prediction is that it is Class II, intermediate. But if you go through the QSAR toolbox, it's Class III. So this is where the chain of reasoning comes in handy because you can see where the two tools depart in their reasoning. And essentially, it's at this question, question 17, is the substance readily hydrolyzed? Toxtree considers this substance to be hydrolyzed, the toolbox does not, and then the consequences from there essentially result in different predictions.

So that was an easy one. But in other cases, the rules are basically not easy to understand. And there's a whole bunch of questions within the Cramer tree that really, it's hard to understand what the original thinking was. And in fact, the same authors that provided the example I just mentioned have also identified thirteen different questions that could actully be cleared up. Not actually changed but better interpreted and better described. And I have a lot of respect for the people who apply Cramer on a manual basis because I have read it many times and I still fail to understand [inaudible] and this is how it tends to feel.

Actually evaluated over a thousand fragrance ingredients. Comparing expert judgment and to independent experts with the predictions generated by Toxtree and the toolbox. And this is very useful because they did this overall, but they also did this on the basis of individual chemical classes. And [inaudible] the guidance really on how you can apply the QSAR tools. So you can see that at one end of the spectrum making predictions [inaudible] you're likely to get discrepancies between the software tools and between the software tools and expert judgment. But the other extreme is if you have [inaudible] for example, it's likely that basically the prediction will [inaudible] with expert judgment and you don't need to think about it too much in detail.

Now this is all about using existing tools for the concepts. However how you might use chemo approaches to develop the approach and the work I'm going to draw upon was all carried out in context of the COSMOS project. This was in 2011 to 2015. So, the project is over, but we've got some of the tools that's available from the project. So, there were really four main parts to this project. First was developing the new toxicological database. I will come back to that in a moment. The second part which is pertinent to this presentation is developing TTC approach further. And there was the whole module parroting of the model development, molecular

modeling for different kinds using information from the database. And finally, the mathematical model, *in vitro* to *in vivo* extrapolation with the system of these models.

So, the COSMOS database is very valuable resource that's come out of this project, but we currently have version 1. Version 2 will be released on the [inaudible] of April. Now we were in a very good place the beginning this project because we had some donations from FDA. For example, the whole architecture of the database was donated by the FDA and based on a similar system they use in-house within FDA CFSAN. And a large part of the [inaudible] data was also donated. During the project we essentially integrated additional data from the scientific literature, from additional regulation software that will be [inaudible] database. And we have done that for around 230 substances.

What you also find in version 2 is skin permeability data for 470 chemicals. And what's also notable in version 2 is as multiple export functions so can download the data from the database. For example a TTC data set that I have been talking about in a moment. And this is a publicly available resource. There's a very short and useful webinar on how to use this. If you do one thing after this presentation I would urge you to look at this database.

Within the database you also have what's called cosmetics inventory. And this is important because we needed to evaluate the methods and models against the universe of cosmetics. This is essentially the intersection of the European cosmetics [inaudible] with the US list [inaudible]. And you can see here some of the characteristics in terms of these functions and the structural features that are available within the inventory. There is not experimental data for all the substances. But you do have the names, the structures, and other identifiers, so a very useful resource for doing chemical space analysis.

Now one of the main challenges in the COSMOS project was really to evaluate the applicability of TTC approach to cosmetics. And part of the problem now is the scenario the chemical space within the TTC data set and Munro. So, we built our own TTC data set to really extend out the cosmetic relevant space. We also applied rigorous data selection duration and evaluation procedure. Basically, only picking certain study types and species and certain durations of study, taking essentially [inaudible] undergone some degree of quality assurance already. Certain criteria for the selection of NOAEL and this was supplemented with some expert evaluation of 25% of the study results. So if, for example, we saw that there were large differences in the NOAEL values from different sources of data, then that would undergo additional expert review. And this was carried out by [inaudible] working group. And also, all substances that were in the lowest tenth percentile with the whole distribution were further evaluated. So, we think this is a very high-quality database and suitable for the assessments of the TTC approach. For cosmetics.

So here are some statistics that summarize the data set. That's been generated. You will see that there's a larger prevalence of the Cramer class I substances in the COSMOS TTC as compared with Munro [inaudible]. What's also interesting is the percentiles, the points of departure for the TTC values. You can see that in the COSMOS dataset these sorts of departure are actually slightly higher to Cramer classes I and III than the original Munro dataset, whereas Cramer class II is somewhat lower. So, this is basically consistent with the observation that we have more innocuous chemicals within the data set that shifting the thresholds to higher values. It's also notable that Cramer classes, the TTC values for Cramer classes II and class III are more or less indistinguishable because it's a smaller number of chemicals within Cramer class II and the shifting of the TTC values. So really this is consistent with an opinion by the European Commission Scientific Committee on Consumer Safety in 2012 that was essentially

saying that we should disregard Cramer class II, at least the cosmetic relevant applications, and treat all Cramer class II substances as Cramer class III for the purposes of safety assessments.

Another way of comparing TTC data sets is in terms of the structural features that they contain. So here you can see a comparison. If the blue bar and the orange bar are the same, the same length and equal size, then you have the same sort of prevalence of substances within the data set. So, you can see from this that substances that are more relevant to cosmetics are more enriched in the COSMOS data set. So [inaudible] and hair dyes. And this is a major concern when we presented an earlier version of this work [inaudible] way back in 2011. Basically, the cosmetic state because [inaudible] so that's a shortcoming that I think we have now addressed satisfactory.

The other thing we want to do with the data set is look at new ways of building models. New ways of the building the TTC decision tree and we've been using what's called chemotypes. You can think of these are really enhanced structural alerts. They're enhanced because in addition to the general fragments information, they carry information on the properties of individual atoms and bonds and fragments within the overall fragments. But this really allows for more nuance, it allows for more subtle discrimination between different kinds of substances.

And again, there's a publicly available tool that will allow you to analyze a given data set in terms of predefined chemotypes. So, this is called the Chemotyper and it contains a library of public chemotypes called ToxPrints. And this software tool was funded by the FDA [inaudible] and molecular networks.

So, one question we started to ask ourselves, and this is rather preliminary, is can we associate chemotypes with potency classes? And could this form the basis of a new TTC decision tree or could it form the basis of QSARs for predicting NOEL values? So here what we've basically done is divide a whole distribution of NOEL values into subsets. So, each subset really relates to the NOELs for all chemicals containing a particular chemotype and then you compare the average NOEL value for that chemotype class with the overall distribution of NOEL values. And what you notice from this is you can indeed start to associate certain chemotypes with lower potency or higher potency substances. So for example [inaudible] are amongst the more potent substances within the overall TTC data set, whereas alcohols are amongst the least potent ones.

So, one issue is really the chemical space. The other is related to the route of exposure. And of course cosmetics must be applied to the skin, so what we do about that? One approach would be to simply redo the [inaudible] using the database. As far as I'm aware, nobody's actually done that, partly because of the size of these databases and concerns over the quality of the dermal toxicity data. But [inaudible] preliminary attempt has been carried out to compare the dermal TTC values of Cramer classes I and III with the oral equivalents. And what you can see is that the dermal TTC values are higher than the equivalent oral ones. So that suggested if you simply use the oral TTC values for a dermal exposure scenario, would you be [inaudible] anyway.

What they also did in this paper is develop a decision tree. So rather than using the applied dose for the skin for a given exposure scenario, you actually evaluate the bioavailable dose following skin penetration. So this implies the need to use [inaudible] data and that's why we have a substantial amount of that in the COSMOS database. If that data is not available, you can in fact apply a QSAR and the QSAR will predict the permeability coefficients which you can use to predict the Jmax and predict the systemic dose that is crossing the skin following the

particular exposure scenario. So, this is a decision tree comparing the bioavailable dose following skin penetration with the oral TTC value. And that would be protected because as we just saw those values are actually lower than the dermal equivalents.

Another way of dealing with the question is developed the concept for an internal TTC value. And vacancy publications on this in effect cosmetic [inaudible] is starting a new project to look at this in more depth so essentially you can bypass the problems of external exposure by different routes by deriving an internal threshold of toxicological concern.

There are two ways of doing this in effect. One is external [inaudible] and convert them to some forward symmetry [inaudible] viability QSAR and modeling for example convert the whole range of external lower values for example internal logically effective basis. Another possibility is start with the distribution of toxicity data to generate the cell line with organs with a screening for example. Chemical database [inaudible] in the [inaudible] you get the point of departure and you can look back on that to get an external equivalents toxic dose TDK modeling. And it's going beyond chemistry-based modeling, it’s biologically based modeling in order to drive the internal TTC. But this is something that is starting to be explored in the research community.

There's various software tools available to apply the TTC approach. With copy additional predictions you have to apply expert judgment in order to come to a conclusion. The chain of reasoning provided by this software tool can be very useful in this regard. Computational methods can also be used to develop the TTC approach and I gave the example of chemotype classes can be related to potency classes. And finally dealing with the issues of oral to dermal extrapolation, how chemistry-based modeling is being supplemented with biological modeling. And with that I would like to acknowledge the many colleagues and collaborators within the COSMOS project that provided the work that I’ve shown you here today. Thank you for your attention.

Audience Question: So, what has been your experience with actually using these in practical settings. We have a lot of tools to choose from. They give you sometimes opposing or different predictions. What are you hearing from the community that is actually using these tools? Are they using them widely? Are regulators paying attention and accepting the data or are we in the middle still?

Worth: The technical developments have created a good problem to have in the sense that we have multiple tools. But that also leads to some possible contradictions. We did a survey back in 2011 with [inaudible] that was referred to with the panel of myself and we really looked at who was using the tools and which decision trees they were using. And they are broadly being used. And then particular for those applications that we heard about not intentionally with this type materials [inaudible] and so forth. So, people are or were at least crying out for better guidance on how to use software tools and some indication of when we need to overrule predictions from the software tools. And we’ve seen some of the later papers by [inaudible] they’ve really done that in a nice way, so they’ve really cleared up some of the confusion that really started from the initial implantation of these decision trees into software form.

Audience Question: Thank you for the nice presentation. I was intrigued by your analysis for the dermal TTC. And of course the question has been coming at us all the time. What about dermal exposure? You have these oral intake exposure TTC based, can you apply those to dermal? I think I heard you say I just want affirmation is that if you're screening and the use the oral TTCs, and you screen, and you find it's not a priority below the value of exposures that are okay if you would. For dermal, the dermal exposures, then you are essentially done. You can be

done. In terms of priority study. So, you don't really necessarily need to go to the next step of the dermal specific TTC. Unless you don't pass that first screen. Is that right?

Worth: Indeed, and that's the proposal in the paper by [inaudible] the whole decision tree, the different steps are really going for more refined assessment. The picture I showed actually goes beyond the picture the paper by including [inaudible] modeling as a final step. But that's essentially the philosophy behind it.

Audience Question: First screen [inaudible].

Rusyn: Well our last speaker is Jessica Wignall. Jessica is with ICF International. And works at a number of projects with US EPA and American Chemistry Council Long-Range Initiative program. Jessica got her bachelor's at the University of Virginia and her master of science in public health at the University of North Carolina Chapel Hill, actually my laboratory a couple years back. So, Jessica has been working on some of the projects that are trying to be a little more quantitative in terms of predicting the point of departure and larger variety of points of departures. So, Jessica.

Quantitative Prediction of Continuous Toxicity Values using Chemical Structure Information, Jessica Wignall, ICF International, Arlington, VA

I did want to start by thanking the organizers for inviting me. Thank you. To talk today you guys have been hearing about TTC all morning. I want to shift a little bit of talk about other types of predictions that we might want to make for other decision concepts. And specifically using chemical structure information.

So, none of the views presented, they're all mine, they're not from any funding source. Like Ivan mentioned, I'm currently at ICF International. This work was conducted underneath the direction of Dr. Ivan Rusyn. It was also conducted in collaboration with the steering committee of people listed on the slide here. And so there was input from EPA, [inaudible] and then California EPA and academics such as [inaudible] San Francisco and David Reif at NC State. So I just wanted to start by acknowledging the collaborators upfront that I could acknowledge they largely influenced and have a lot of input into this work.

So quick little context overview is that while TTC does classify chemical structures into classes, there are some decision context in which we might want to have a quantitative value. For making our decisions. The quantitative values such as [inaudible] itself might cause harm to humans. And while it is hard to generate those [inaudible] with ample testing data, never mind actually going to something like using QSAR data we still want to try, we're still going to see how well it goes.

These alternative methods not based on animal data do have uncertainty built into them but we can quantify that uncertainty and then we can use that uncertainty in our decision-making context moving forward. And there is hope at the end of the tunnel, I will come back to us at the end but, in general our ability to make these models and make these predictions is based on having good underlying *in vivo* dose response data. And we are in the process of extracting and getting more and better *in vivo* does response data.

So, the outline of my talk is the general workflow when you're talking about building these kinds of models that we been hearing. You start by collecting the data sets so that's the most important component of the project. What data sets are you talking about, what data sets form the basis for the model? And then once you have those data set built for QSAR models and there's different decisions you can make when doing that, and ultimately you want to apply those QSAR models in different decision-making contexts. So, in the talk, I will specifically talk about data sets that might be used in decision-making, including points of departure like NOEL we've been talking about all morning. I'm also going to spend a few minutes talking about benchmark dose values and ways to standardize the calculation of benchmark dose values. They're a type of point of departure. Along with other toxicity values.

In the context of building the QSAR models themselves we're going to talk about considering the chemical space and really similar to some of the things that we have been hearing. Using concepts that are considered for the chemical space of the data set. I will talk about how to evaluate the performance of the models that we are building. It's important to benchmark them and have an understanding of how well they are performing with the uncertainty and you can quantify that uncertainty.

Then the application context would be addressing that no data no hazard challenge so if we're talking about chemicals without any data without any testing data or animals or any other sources, what are our solutions? And then come back at the end to talk about [inaudible] which is a user interface that we've built that allows you to access some of these QSAR models.

So, this is little bit of risk assessment 101 so bear with me in the foundation for this if you understand and you can zone out for a minute. But point of departure cannot be using decision-making. We have been talking all morning about NOAEL, so that would be coming from a dose-response data set, what dose is shown to not cause an adverse effect. That's defined as NOAEL.

However some date sets is that you may not have a NOAEL in which case decision-makers tend to rely on LOAELs or the lowest, the dose at which the lowest observed adverse effect is seen. There some limitations about using LOAEL and NOAEL but they don't take into account the shape of the entire dose-response curve. And that's where benchmark dose modeling comes in and so for benchmark dose modeling, like I said it's basically an approach that fits the curve to the entire response data set and takes into account all that information.

So, based on your specific biological response of interest, on the Y axis, what does is estimated to cause a response? And I say estimated specifically because we're talking about modeling and is a certainty around it. And also, what's actually used as a point of departure the decision-making is called the be BMDL or the BMD lower [inaudible] the lower 95% confidence limit in the BMD. The BMD being the central estimate. These are different types of departure used decision-making, BMD is a little bit newer obviously than NOAEL and LOAEL and can be applied in certain context. So, for example for talking about noncancer cancer assessments, noncancer test [inaudible]. Typically applied manufactures to the point of departure and derive a reference dose to reference concentration. Cancer assessments will extrapolate using the slope from the point of departure in order to estimate the risk.

The BMPs are ways, points of departure that can serve to take advantage of the data that is inherent in the whole dose-response curve. So, they're data-driven points of departure. So, there are limitations to that. They are time intensive. They take time to calculate and walk through. They can be complex. They require expert judgment. For example, if you want to, you

want expert input on the biological response that you care about for context, you want expert judgment on the statistical models that you apply to data set. And you also want the expert judgment interpreting the results of the benchmark dose modeling. So historically has been a very time-consuming process. And are some data sets that are not [inaudible] for dose modeling so for example you are not the affected any dose, you could still identify a NOAEL, but you would be hard pressed to actually model a response and get a benchmark dose. So, while benchmark dose is data-driven, there are still contexts in which it's important to also use [inaudible].

The limitations to benchmark doses, we try to approach some of those imitations and identify strategies to work around them. And that's through an effort that we did a couple years ago. Where we standardize calculation of benchmark doses and BMDL for a large number of chemicals. And this is in the context of essentially collecting data sets that we think would be of interest to decision-makers when we talk about modeling. We went through and collected 880 dose-response data sets for over 350 chemicals that had toxicity values and using the toxicity values to refer to reference doses and factors and points of departure. And all of those types of values I'm calling toxicity values.

So, we collect all the data into the database and we apply the standardized benchmark model approaches. So, we follow the EPA guidance and we used EPA benchmark dose software to do this. So essentially, we automatically ran all of our dose response data sets through the benchmark dose software tool and came up with estimates of the benchmark dose and the BMDL. So even in the context of limitations on the slide we were able to show that about 75% of those data sets can be modeled using this approach.

So, we were still able to batch process hundreds of chemicals essentially. And end up with a data set of benchmark doses and BMDLs for over 300 chemicals even after this automatic processing.

-And you're probably wondering [inaudible] when emphasize how much [inaudible] this is important, it still is obviously, but what we wanted to do is compare the batch calculator benchmark dose against experts derived benchmark doses. So there are about 40 chemicals that, 40 chemicals for which assessments include benchmark doses and were able to compare directly our calculated benchmark doses against those assessment benchmark doses. And you can see here that the correlation is very high. Extremely high. [inaudible] between [inaudible]. They were also converted to the human value [inaudible]. So even the batch processing not necessarily relying on chemical specific expert judgment was able to generate value that correlate very well with [inaudible].

One component of this process that I did want to take a moment to talk about that I think might be of interest to different decision-making context is this type of approach can also be used to evaluate other types of data. So, for example if you were looking at a cross [inaudible] for a given chemical. The specific figure is showing [inaudible]. That's the assessment based on the NOAEL that included consideration of critical effects of body weight changes [inaudible]. And so, we said what if you were to calculate the benchmark dose for each specific data set for each of those critical effects? What the distribution potential points of departure look like? So, on the right in the yellow might be a little hard to see but those are calculated BMDL for potential points of departure. This is a standardized way if you were to add in more studies you could look across studies in a standardized way, it will let you identify outliers, potentially the idea being that you are applying a standardized approach to different data sets in order to see what the data looks like.

And so coming out of the effort we essentially are sharing the, accepted that BMDs and BMDLs are useful points of departure and they can be calculated in a standardized way. So, while historically they've required a lot of expert judgment and statistical knowledge and experience to generate them there are approaches that are more standardized in a straightforward way. Making assumptions but then the batch calculated BMDs can actually be used for many purposes. I already mentioned evaluating for example the weight of evidence of data around the chemical. It has been applied different approaches there similar especially to the BMD with high-throughput active data and calculating point of departure off the concentration response curve. I have seen it applied with [inaudible] data as well, so at what point is the activity happening? And ultimately, these kinds of batch calculated BMDs can serve as data sets for QSAR modeling. Which is little here. And for those data sets the BMD's that are calculated a standard way and are used with other toxicity value.

So, for this, results that I will show the presentation of the numbers I'm talking about here. So, we been hearing other efforts to collect this type of data. This is a specific data set where we are interested in building models not only for points of departure. But because the departure of interest decision-makers that might want to apply their own factors are their own margin of exposure. But also, we were interested to see if we could build models for the final assessment values that already take into account some uncertainty and extrapolation.

As you can see here [inaudible] the sources from diverse sources such as [inaudible] program. Office [inaudible] etc. And so, when I talk about the toxicity values these are the numbers of chemicals of talking about and these are the list of chemicals. The toxicity values.

So now we collect the data set moving into the second phase of what were talking about actually building the actual QSAR models and again the point being that we want to predict values of interest to decision-makers. TTC to this point was deciding of chemicals were in certain classes in order to decide how to move forward. The chemical poses a concern are not. Here where actually calculating [inaudible] calculating actual doses for the actual values, toxicity values that are shown on a previous slide.

And we talked about this in general. Touch on this in an earlier talk but I just want to emphasize, so in QSAR modeling chemical structures can be represented using numbers. So that would mean you could break down structure into the different components and you can generate different data sets that describe those structures. So, does in include [inaudible], what's the molecular weight? Also, a different, you can imagine the different ways to describe the chemical structure using numbers. And there are data sets and descriptor sets that describe chemicals structures [inaudible] of numbers.

And ultimately way can do is use statistical models to describe the relationship between the relationship between chemical descriptor values of the outcome of interest. With the statistical models can be used to make predictions of new chemicals for which you don't have that outcome information. So, a given chemical description matrix based on instructor and then you based on the numbers are able to estimate the [inaudible] they build models [inaudible].

And then I just want to note that essentially as were discussing that your ability to have model predict chemical structure depends on the input chemical structure. So, your model is trained with chemistry space and so it's important to recognize that you make predictions about new chemicals that you want to check and see where the chemicals fall in terms of the chemical space of the training.

If you're wondering about [inaudible] this is how it compares. The blue dots reflect the toxicity value chemicals that I was referencing a few slides ago, the green dots reflect 32,000 chemicals that EPA has used recently in a collaborative activity prediction project. So the large data set you have chemical space here. TTA can be used to describe the chemical space of, or can be used to compress the multidimensional chemist descriptor space into two dimensions. Explore the variance of your data set in two dimensions even though [inaudible] multidimensional. And we're showing that there is a lot of overlap between our data set and the cost related data set a 32,000 chemicals that might cover the chemistry space you might possibly be interested in. So, we are showing some good overlap and some good, broad coverage. With the toxicity data value set.

And because I don't know if you notice but this included inhalation reference concentration and inhalation unit risk and so there are some chemicals that might be outside the oral exposure space. And allowing us to expand a little.

So, when you're moving forward to building QSAR models, in general, when you are evaluating the performance the model there are some considerations to keep in mind. For example, any performance metrics should be calculated based on the external data sets. So, what was not used in building the model. So this is something to take a look at when you're looking at a QSAR model or any kind of model and it's giving you a value of what's predictive and we're doing this good of a job you want to make sure that values are calculated on external data sets. But no matter what, no matter what kind of model you want to build we are limited to how good the underlying data is. So, there's experimental variability [inaudible] is the only independent variable is the chemical structure information. There's still variability in the experimental data itself. So, we can't ever be better than that experimental variability.

And that in general model performance is improved by using larger data sets with large numbers of chemical. The more data you have the more robust your model is. And then if your data sets are closely related so if your outcome is narrowly defined mechanistic outcome, related to chemical structure you will build a more predictive model than a model for something that maybe more ambiguous. And so, these considerations do have implications for predicting *in vivo* outcomes for the environmental chemicals. Were experimental data is very limited and variable across species in that [inaudible].

So, this is just showing some examples, QSARs we've mentioned this, these are [inaudible] model earlier performance because predicting the binary outcome the performance would be measured in terms of false-positive, false-negative, true positive, true negative, so that's the top left graph there. If you are modeling sort of a continuous outcomes such as the top right which is the inhibition of the thyroid hormone receptor bindings. Then you can plot for example experimental values against the predicted values and come up with correlation between those values. Obviously, the correlation being perfect, and we were perfectly able to predict the experimental values. In the bottom graph which is hard to see but just to point out that similar to the continuous outcome of the findings if you're predicting outcome like LOAEL, and this is coming from TOPKAT which is a proprietary model, the performance of model is using values for specific structural classes ranging from about .8 to actually to almost 1. These are some ways you can report the performance of your model.

So, in the context of fact that we have this in a QSAR and we been spending all morning to talk about the different types of [inaudible] why is mine special? I don't know if it is. Here we have some objectives and you may want to meet those objectives and that's the context here that's

important, to predict continuous outcomes of interest to decision-makers. Especially points of departure. So again, if you are decision-maker and you did not want to, or you wanted to apply your own uncertainty or your own uncertainty factors to your data you want to start with the point of departure. If you wanted to extrapolate a slope for noncancer [inaudible] which is not typically done, then you might want to start with a point of departure.

Want to facilitate transparency communication by using publicly available descriptors, algorithms and external validation. So, this is something there are the proprietary [inaudible] out there, proprietary software is out there, so something we wanted to do was really be as transparent as possible. So, we specifically used two publicly available published descriptor types of TDK data. Especially between those two descriptor types, each descriptor type with the predicted outcome. And we'll average those predictions to develop the consensus model so that we are taking into account different types of structural information.

[inaudible] with algorithms which is more structural specific [inaudible] that you apply in QSAR that's based on the concept of a decision tree. So we have been talking all morning about decision trees. Random forest is a [inaudible] is going to randomly sample from your descriptor set so you are not pre-defining what that looks like. It will build a tree that it thinks is very predictive.

And then for the cross validation which is a way to ensure that all of the performance metrics are reported based on the data set which is external to the model at the time.

And then the third objective is providing the online portal. And so ChemBench allows you to build QSAR models using your own data sets and your own choices. For the specifics and descriptors. But like I mentioned earlier we built a front-end user interface to access the data which is available at ToxValue.org for the models we talked about today.

So now that I have built these models I'm going to show you some data on how well they are performing. So, the model performance did vary according to the toxicity and the number of compounds that we have. That what we did show is all models did perform significantly better than [inaudible]. So, and then I will talk more about the Q squared is a metric similar to an adjusted [inaudible] one is perfect. And [inaudible] no predictive ability in our model. As you can see in the NOAEL models we're hitting a Q squared of about .5. The distribution is the bottom left spans an order of eight magnitude and you can see for a given prediction we are able, for a given chemical we are able to calculate the error between our prediction of the chemical and the observed value. So that's where we are getting the uncertainty of able to quantify the uncertainty. So, I have my 487 chemicals. For each chemical I have a measure how well I predicted that chemical and how well, and how poorly I predicted that chemical.

And so, if I average all that information across my chemicals, what I end up with is our average model error. And so that average model error can apply new predictions. And provide uncertainty range I'm talking about. So, for example the bottom left are three models. Again, the model Q squared is reported there. And the value to the right is the consensus model prediction of the absolute error, sorry the absolute error of the consensus model prediction in the log scale. In the average 90 [inaudible].

So, for an unknown chemical for which you're calculating descriptors for and predicting the outcome you can actually apply uncertainty to that prediction and have an idea to my prediction of the value nut then there's uncertainty around it. Based on the model screening we think it's, the uncertainty ranges from plus or minus .67 or .8 [inaudible].

So, given some of these models if you are in the QSAR you might be expecting to see .7 or .8, that's when you're talking about various well-defined outcomes. So in our case here we will talk about these systemic toxicity outcomes even models may be considered low predictivity do provide information that's useful to us.

So, for example the graph we are seeing is that the black line which represents our consensus model is significantly better than the dotted gray line and the gray line says what if I would just to predict for any new chemical that it was the average NOAEL. So basically, picking the midpoint of our observed value range and saying what if any new chemical is predicted to be that average NOAEL. And you have the data with primary chemicals inform future chemicals. So these QSAR models are significantly improving on that. We are still adding information even in the context of the fact that we're talking about [inaudible] of about [inaudible].

As I have a talking benchmark doses I do want to show you some benchmark dose results. They were not performing as well as some of the other models but were limited in the chemical space and the number of chemicals, so we anticipate that additional chemicals will allow us to build more predictable models. And still show that especially for BMD they statistically improved over that model predicting average. And it makes sense that they might be more noisy and account for uncertainty. And there will be more variation across chemicals [inaudible].

So, putting the theme a little more context. This goes back to can't be the expectation that we can be predicting better than you would see in experimental variability. So, the first in the graph on the right the first line is showing based on EPA reference database source, how well is subchronic rat data predicting chronic rat data? There's three and a half orders of magnitude of uncertainty around those predictions. And so that's already telling you, okay I think that subchronic tests are predictive of chronic tests, there's still uncertainty around it, then how do our models fit into that kind of context?

In the next line is rat chronic versus mouse chronic. How predictive is rat chronic of mouse chronic going across species? And you can see that models that were built on the averages if you predict the average BMD or average NOEL, you're still talking about two orders of magnitude of uncertainty and then on average the consensus models have about one and a half orders of magnitude of uncertainty. So, you can see that in the context of these baseline expectations again we are still improving and adding knowledge to those, to the data.

So, we can't talk about QSAR without talking about mechanistic interpretation. The descriptors sets that were used in this model, like I mentioned, were [inaudible]. So specifically, for [inaudible] models we looked at the descriptor sets here where we said which descriptor were best at reducing the error of a model. And these are a little cryptic. There's a link at the end in the references what each of those means. But the bottom left, it's really the middle plot with triangle. [Inaudible] three for example is a descriptor that takes into account molecular branching of the compound. And takes into account the atomic weight and you can see the chemical structures with similar activities are grouped according to those descriptor types. So, they would be good at differentiating top versus bottom predictions. The same with that right graph which is using [inaudible]. [Inaudible] information on the mechanism underlying why the model is predicting certain chemicals in certain ways.

So, summary but not done yet. The standardized BMDs can be calculated in an efficient manner which allows you to move quickly and minimize the need or the time it takes for expert judgment in certain decision contexts. These QSAR models can be built to predict these quantitative

values of interest to decision-makers. And these results can be presented in the context of your baseline expectations. So, what reasonably can we expect when we're modeling experimental data? What kind of variability are we talking about there?

And I mentioned this briefly, but we do have assumptions inherent in aggregating those kind of data together when we're talking about systemic toxicity. In general, the more homogenous your data set and the more homogenous your outcome of interest, the better, but we needed to balance that against the need for robust training sets and the ability to gather a bunch of data together. And the reason we had to make those big assumptions that we were aggregating those data was because we had limited *in vivo* data for model building. I did mention earlier there's efforts underway at EPA right now to extract additional quantitative data from the ToxRefDB animal studies and we're going through and extracting the quantitative data for all treatment-related effects that were seen in those studies. So even if there wasn't the critical effect we will potentially end up with data sets that describe the same effect across compounds even if it wasn't the critical effect. So, we potentially build more, better models and apply those batch calculated BMD approach to calculate BMDs and the create a new data set for models.

And so, coming back to the application, we really want to put this in context that we're not talking about data-rich chemicals. We're really talking about chemicals with very little data. For example, if you wanted two screens that generate screen value or if you had to make risk management decisions about emerging contaminants, prioritize chemicals, or evaluate alternatives. And the input would be, if you have chemical structure information you can generate the inputs. The reason that [inaudible] are similar reasons to why TTC is limited in chemical structure space. Things like organometallics and inorganics are difficult to calculate these chemical descriptors for in a standardized way. But for those we can calculate those descriptors for, we can generate these quantitative outputs. Again, one more time, there's uncertainty around them, but you can know what that uncertainty is and then use that for your decision-making context.

So real quick specific applications. If for example there was [inaudible] had site-specific sampling showing chemical for which there were no data available, no toxicity values, the decision-maker might need to make a quick decision. They may or may not have access to experts and they may need to make a decision very quickly. So, these type of approach in these models could potentially serve as a resource to the stakeholders who need to quickly predict hazardous compounds that lack data.

And I mentioned ToxValue.org. It's an online portal, it's in development, it's very much in development. It's on there right now if you were to go back to your computer for those of you in the room or those of you in the webinar, if you were to type it in right now, you could get to it and play with it but it is in development still. Step one you enter the compound information either with smiles or some other identifying information and you can upload your own list of chemicals using SPE. You would confirm that you uploaded what you thought you uploaded or that you picked the chemical you think you picked, and then you could pick the models that you want to run. As for example if you were interested in both oral inhalation or cancer and noncancer you could select and run your models. Like I mentioned they are hosted on the ChemBench, this is just a front end access to access the [inaudible].

And then retrieve the predictions. And so ultimately be able to download those and keep them. And again, there would be an uncertainty applied to each of those predictions. So I will say again, just one more time, that it's under development so if you're interested in learning more

about it just email that email address and we can keep you up-to-date as to further development.

And then there are some references. Thank you.

Rusyn: We will have a chance to discuss this in the roundtable. But if there is a question or two from the web or the audience, please ask now. Let me start. Jessica, so you did mention that in discussions with different decision-makers, there seems to be a preference for what type of number they need. So some people they need the actual final toxicity values it already takes into consideration uncertainty and all the uncertainty factors that have been applied and others are saying, just give me the NOAEL and I know what to do with it, you know, I'll divide it by 100 or a thousand and I have a number in a margin of safety in mind. So, in your opinion as you have been in kind of consulting and dealing with the regulators and the regulated, do you see a strong preference for a certain type of number or does it always depend on the decision context and the preferences will change dramatically?

Wignall: I think the idea that when I heard the preferences were data-driven. And the fact that you have to apply uncertainty factors that are default values that may or may not be data-driven. So not necessarily data-driven defaults. So, there are some reservations with applying those values. Whereas if you're talking about like a point of departure like the benchmark dose which is derived from the entire dose response curve, it's making use of the data that you have. And then you can decide case-by-case what margin of margin of safety might apply or what uncertainty factors might apply so you are able to, essentially the better and more data you can use to apply your decisions and the more you can step away from using defaults, the better.

Rusyn: Well thank you all. So we would like to ask all the speakers and Dr. Jacobs to join us at the table. And we will get going with the roundtable discussion.

[Note: temporary drop of audio and video recording]

Roundtable Discussion
Ivan Rusyn, Moderator
Kristi Jacobs, US FDA, College Park, MD
All speakers

So, if I can ask the speakers to come down. We don't have any questions.

Well thank you all for the presentations. And we do have an opportunity to give a couple of people to make introductory comments. We actually do have online Dr. Cramer. We have been using his name I think throughout the morning. So he was supposed to listen to this morning's presentation. And he wanted to say a few remarks. So are we ready to put him on?

Okay so we will then notify you when he is actually dialed in and we will let him speak for a couple minutes.

We also have panel that is from the agency -- Dr. Kristi Jacobs. She has been with the agency for a number of years and she has been one of the leaders actually pushing TTC approach not only within the agency but also in collaboration with the [inaudible] and [inaudible]. So, Kristi that she would like to give you an opportunity to say a few words if you would like.

Jacobs: Okay well hello. I am [inaudible] Jacobs. I'm a supervisory debt toxicologist in the Office of Food Additive Safety. And one of the reasons I think that they added me to the panel today was much of the work done recently with great effort by the BHL and [inaudible] to look in the, 2014 they held a workshop looking for a state of science for TTC. Looking at the Cramer decision tree. And if you noticed one of the things that came out in the presentation the original decision tree, and we has been established in a number of publications using a direction fax and the impact of how the dose is administered with specific endpoints. So, WHO and [inaudible] look at the current state with the [inaudible] to come up with a harmonized decision to and an evaluation state of science look at the classification scene was set for the purpose of prioritization for assessing chemicals with low exposures. That report for the workshop was recently published earlier this month. Along with all of the public comments and the response to these comments. That final report and comments can all be found on the [inaudible] website if you Google search for WHO is PDC 2016 you will find it in the first [inaudible] pockets I thank the organizers for allowing me to participate.

Rusyn: Well thank you again we are welcoming questions both from the audience and also from the line. So I will get a computer in a minute. And we will be able to monitor Westerns to identify the web participants. I would like to thank them for participating and not just the lectures but also the roundtable.

Let me start by perhaps trying to build a discussion here in some sense of order. And I think you don't start with actually some data to this you decide which is the right way to do this. And whether it is quantitative or qualitative decision. At the end. I have heard from speakers today that both Ranches presentation, we built our own data sets. And then we do something with it dishy that the mechanism overlap in the chemicals that go into the different competition of tools whether they are qualitative or quantitative. And there is promise a better data. Which we have still yet not explored. But my question to the panel is, and maybe Grace you can start because you are now with EPA, Nick I sit on the throne of information that I think we can mind for at least the next five or 10 years. And maybe open top mining whether she rather than difficult underground mining.

So, do you believe that actually we need still to work on better data sets? So, do you think it's enough now? You know a couple hundred here there. Dermal inhalation. With areas or do we still have some ways to go?

Tier: I don't think I can answer that from an EPA perspective. It would be more from my own perspective. And I think for those that have known me for many years my mentor has always been I haven't got enough data. [Inaudible] and eight even had a [inaudible] spot and of weird area. But I think there's been tremendous progress over the recent years pulling together and compiling some of this data from different sources. I think the COSMOS project was very notable in terms of the way this data has been compiled more cosmetic information. I think this opportunity in terms of explaining them and evaluating the data that is being collected under each. We have the Munro data obviously and with Tim has been presenting this morning. I think that for me, maybe we are almost there in terms of having enough data, but we have many opportunities and many challenges in terms of being able to evaluate the quality of the data and the sea to understand [inaudible] are not. And the uncertainty. But also, the challenge in terms of understanding the chemical landscape and where the overlaps are. So maybe the short answer to question is, I think we have made tremendous progress. Maybe we don't need to do much more election. It's more trying to understand what we have and where the opportunities lie.

Rusyn: Tim and others please weigh in as well.

Adams: Yes. I think, I agree with Grace with the point she made from our experience, going through even the Munro data sets and the data to drive NOAEL. We found probably 20% 25% of the studies could be replaced by more recent studies that were more robust than the study that was cited. So it is a question of consistency among the industry and the idea of standard assays with regular, rigorous guidelines really have shown a lot to improve the data. And I think somehow a retrospective review of all the data would be able to clean up a lot of what we have as confusion at this point in regard to endpoints and NOELs and etc.

Worth: Andrew. Yes I agree with that. I think it's also important to note that a few years ago, not so long ago the access to the Munro data set in the PDF, versions as long as chemical structures that are downloadable from the three database or the SOS data site. And represent the more, same data in a more transparent way in the reasoning to make it lead to different NOEL and it's all captured in the three database or the date is not a date in a sense but the transparency and the whole process is much more clear.

[Speaker not identified] Yes, I would agree. I think that's important point that it does exist about, insist her to find a pocket sometimes hard to communicated end-user. So the more that the data that people are using in the matrix are available in the market others can review it. And just a follow-up on Grace's point about taking a minute to understand uncertainties and the reduced ability and an information inherent in the data we already have think that we do analyses that's what we should inform the next steps. So for example we identify studies and to be replaced. Let's go do that. And build a better data sets. So, I don't know that yet but with the generate more data but I do think we need to collected to make it usable for the think this.

Grace, thank you for mentioning actually the REACH data because there's a lot of interest and hope that holy grail intellectually deliver but there's also a concern with the fact that these are trust me, this is the best, submitter, visa data I have. And the quality of that information is really unchecked.

So, while they are on surface a lot of data points and studies, the fidelity or certainty of those being true is not -- is nothing very [inaudible].

So, the efforts perhaps EPA and maybe collaboration with DFC and other European colleagues to actually go and extract the meaningful data set from those studies. Because on the one hand it looks fast. On the other hand of his a lot of concern about the quality underlying quality of information there and you start using them for decisions and modeling, you may run quickly into lack of trust because underlaying information was not [inaudible]. So, are there any thoughts Andrew with the JRC again you are not speaking on behalf of the European Commission and I understand that what's your opinion on that?

Worth: This is a huge resource. But as you mentioned its largely unchecked by [inaudible] basically you have some type of [inaudible] there and that's the registering opinion of data.

Other parts reach dissemination sites in the quality data and [inaudible] that are linked to the authorization for some understate of authorities are actively engaged in reviewing the data. And more generally I think in a sense for Jessica's presentation at capturing uncertainty the more data you have, the better pitcher you can get. So, have distribution of all the NOELs.

And speaking of the data, it's not just numbers. These are certain types of numbers pocket some of these numbers that are reported I think for REACH or NOELs it's not the underlying information because on most other studies were probably multidose response. And so again trying to think about this response modeling of today rather than the understanding of what the point of departure 2034 years ago when the studies were done.

[Audio and video recording resumes]

Rusyn: The point of departure do we actually need because the line of discussion this morning has been on NOAEL. And I think it pains a lot of regulators to still be clinging to the NOAEL as the point of departure du jour. So, get some opinions as to whether NOAEL is something we should carry on in the future or should we be thinking more about benchmark dose and some other types of more contemporary ways of describing the point of departure that we really care about? So, Kristi, do you have an opinion again not from the Agency view but you've been in this business, so does NOAEL serves your needs?

Jacobs: Well before I touch on the topic of the NOAEL or BMD or a BMDL as a point of departure used I a regulatory risk assessment, I think the issue of study qualities and well-established inclusion criteria and exclusion criteria is something that has a more fundamental point when we start to build the database and derive the values that we need to consider. I think Andy mentioned it a bit in talking about the COSMOS project but as one of the two seers of the database, it was very informative to sit through with a group of very experienced toxicologists and look at the Munro references and the data that was used to drive those NOELs and some of the newer data that has been published and setting up a inclusion criteria in terms of the number of doses that need to be tested. The study duration. The species selection. These decisions make a fundamental impact on what the final NOEL choice would be. And going and establishing the criteria early on in the process is a critical, is a critical point.

Part of that COSMOS project was looking at each of those individual NOELs when there was a difference in the NOEL reported of greater than tenfold. No matter where it fell within the class, to say, if there's two studies out there and a NOEL differs by tenfold, toxicologists need to take a look at that and they need to make a reasonable decision about which one of those NOELs should be entered in the database. And this was a long painful process that took multiple years to go through and look at each one of those. But it definitely impacted the quality of the data set. And I think that's one of the more critical things. I think choosing a NOEL or a BMD or a BMDL, these are, there may be a preference to use one over the other but in terms of what number you apply the safety factor to, I think making sure that the study that you are choosing is more important that the quality of the study is the best quality that you pick whether you derive a BMD or you choose the NOEL for the study is not as important as the individual quality of the study itself.

Rusyn: But ultimately the decision tree or some sort of a model is going to produce a number so what you talked about is what goes into that building of some sort of classification or quantitative or qualitative model. Once it's built it only produces one type of a number. So it's less important to only think about what goes in, but also what comes out. So if the only thing that comes out is a NOAEL is that satisfactory for decision-makers or do they, would there be better decisions if what comes out is a quantitative point of departure? And again, maybe it's impractical at current stage to think about producing a benchmark dose or thinking of classification trees that can actually be more precise about the point of departure. Rather than a NOAEL. So has there been any discussion on thinking around that?

Jacobs: I think WHO report considered the question on whether or not ht enewer databases should consider BMDL and thought that there might be some merit to that but thought that the NOAEL was a sufficiently protected value to use and that that has been demonstrated to be conservative enough to be protective. And so, the agency, we'll use the data available to make the best decision. We haven't made a determination one way or the other.

Patlewicz: I think it is important not to necessarily sway us down one solution fits all cases. My own opinion is to be as flexible as we can. Maybe a NOEL is good in one scenario, maybe a BMD would be better in another depending on what sort of question are you are trying to address. I think we have a screening level risk assessment project that we're doing in house and we're looking at all sorts of toxicity values and we will evaluate them accordingly, depending on what the question is and what sort of modeling we can do based on that type of data.

Adams: We always need to have a second level of oversight with regard to the NOEL you are actually looking at. Or whether you use BMDL. And that has to do with the number of substances that have data related to that material. So if you have enough data across enough materials, you can see trends and see whether the NOEL is going to work or if you have to choose a BMDL approach. It is looking at a larger data set. Then stepping back and deciding what approach you need.

Rusyn: We do have a few questions from the web. And we will get to them later, but I would like to take questions from the audience.

Audience Question: When you can get a BMD, you can get a LOEL. Not necessarily a LOEL, but a NOEL. The BMD comes with a measure of uncertainty. And one should be able to look at the difference in a set of chemicals between the LOEL and the BMD and ask how often is that difference within the degree of uncertainty of your measurement? And it really does not matter whether you use the LOEL or the BMD, you will get about the same value especially when you are looking at large numbers of chemicals.

Wignall: We did a paper, I cannot remember the exact ratio but compared BMDs against the LOAELs and I think what we found was that traditionally NOAELs are thought to be about the 10% change and BMDL was about that, so actually they were lining up closer with the NOAELs than the LOAELs. But I want to add, from the perspective of QSAR, and I would also be curious for Andrew's opinion, too, I've learned over time that it can be slightly problematic to think about modeling a lack of activity. And so that NOAEL is a no activity level. So the idea that we're building these structural models to predict a lack of activity, whether or not that should be taken into consideration.

Worth: I think the systematic way to do it would be to use the benchmark data as the sort of starting point to building the model. If you're jumping straight into the safety assessment, I think what we're hearing is that multiple options apply.

Rusyn: And a related question to this, everyone agrees that local models are more predictive than global models. This is true when finding a relationship between structure or metabolism and the potential outcome. And again we went through this this morning. Do we need dermal TTC versus inhalation TTC versus oral TTC, and then are there certain types of classes and the Cramer classification tree? And this will lead you down a certain path should you determine that you have a certain type of hazard or an alert. If you have a small set of chemicals that are structurally similar and you have your new chemical that's very structurally similar, it [inaudible] very accurate prediction whether it's true or not. One of the ways to localize the models is

through taking into consideration the mode of action or the adverse outcome pathway. We had a couple of questions from the web on that. So can I get maybe a little more discussion because, you know, folks did mention [inaudible] if it's that alert, do that. And that information currently is based on the respective mode of action or what we know or don't know about a particular chemical. So if we think of going further and developing and refining some of these classification trees and models, what do panelists think of not just, we all say that mode of action is important, and more of that type of information will be useful, but how we can actually start using that information. So, if I get some points on that, I would really appreciate that discussion.

Adams: From my experience, taking a mode of action approach is available provided you have good enough metabolic data. Then you can work with something. Andrew had a slide, that talked about the confusion around [inaudible] carbons. And that particular question has to do with [inaudible] toxicity [inaudible] it could be refined and has been refined in our tree to be a [inaudible] less than 10 carbons [inaudible] with only four carbons on one side of the [inaudible]. In that process you take about 25 or 30 [inaudible] and reduce the level of concern for 20 of them while you have five that meet the criteria. Because it all surrounds one particular mode of action and that is [inaudible] that second carbon [inaudible]. And you will take that information and the next question you ask is does it apply to any other, are there any other arrangements like that? And then you can find [inaudible] have exactly the same range. And have the exactly the same neurological effect. Once you understand the mode of action you can ask questions that are very specific which give you true outcomes. Not outcomes that are pulling [inaudible].

Rusyn: These are great examples, but I think they are few and far between if you look at the larger number of chemicals. So, would you agree to that?

Adams: Actually, it is less and less. One of the more popular journals, which is the Journal of research in chemical toxicology. This is part of a movement in chemistry to have integration with toxicology that did not previously exist. And *CRT* as a journal, a lot of the work that's being done in the mechanistic work on various toxic endpoints, really has expanded over the past 20 years. It is part of the obligation of the discipline to better integrate into chemistry that is going on right now.

Rusyn: Grace, do you have a view? How is that relevant or not relevant to the TTC like effort?

Patlewicz: I think that is a difficult one. I think prior to joining the EPA, I was very much, and I'm still, a strong advocate of local models. Having some mechanistic transparency. A lot of the work that I've done in the skin sensitization area is very much underpinned in that type of way. Being now at the EPA has kind of opened my eyes to the issue of how many chemicals we might want to evaluate, being more open to the idea of where to global models fit within that sort of scenario? And the work at the OECD, too, in terms of the [inaudible], I've been really interested in that, but at the moment, it feels like we haven't really got a critical mass of [inaudible]. I'm constantly challenged by, I have a chemical, I may have lots of chemicals I need to evaluate at the same time. I want to make use of that [inaudible] information but I do not know how to navigate through that. For me at least, having a tiered approach in terms of using global modeling approach. We have been doing a lot of work recently in read across in this sort of scenario to be able to build on some of the examples that Jessica has alluded to today in terms of being able to hone in and focus in where I can make use of that mode of action information so I am not so overwhelmed.

Rusyn: Do you think this is another layer on top of the tree? Because you have to decide what local model to use. Do you think there should be a second step actually predicting what type of hazard you may actually have and then go into that local model or do you bypass it altogether? Because somehow I think it has to be a tiered approach. But I think the field is struggling to understand where do you stick mechanistic information. Do you stick it on top or do you come back to it later? So we can feel better about the decision we made.

Patlewicz: I do not know. It seems like you are trying to pull me down one way or the other. And it does depend. If you have a large library that you need to sort of prioritize in some way, I think the tiered approach makes sense. If you are looking at one chemical at a time, then yes you can go down in a more focused way. And it will lead you down a different sort of scope. I do not think there is a right answer. I do not think there is one solution. I am interested in how you can use some of that type of mechanistic information, like from using the [inaudible] information for instance with chemical descriptor information. Can we merge them? Can we blend them in some way that we can make more meaningful predictions? And that might [inaudible] where you might generating or using different types of information to make a decision.

And the other thing I found that was fascinating from Tim's talk, I have been thinking about say TTC in a very compartmentalized manner of, well, we have Cramer and we assign things in classes and we use that to decide what the threshold is. This reinforced for me, how these different disciplines are blended together. Your talk today made me think, that's kind of what we do in read across from an expert point of view and we should be more clever in complementing the different activities that are going on.

Rusyn: I apologize if it seems like I am picking on you. Andrew, will you comment on your experience? How to blend data with purely structural approach? Do you have an opinion? Do you see a practical way rather than that the data is out there, we should use it? So, how should we use it? Can we answer that question, or maybe in five years?

Worth: I think we are on the road. We have seen quite a few papers that have explored this idea. And also, within the COSMOS project, we were exploring the use of [inaudible] to identify modes of action [inaudible]. And you can rationalize that into the initiating event and so forth. This is one example of you can bolt onto. But the difficulty is, this is evolving and it does give us an opportunity to make the decision tree more refined. It has not taken on the integrated knowledge. So it is about the philosophy of the design and approach. And should you sacrifice some of these [inaudible] just to keep it simple?

Rusyn: Not to belabor the point of how to use *in vitro* data because we did have a couple of colloquia that addressed these and especially making small steps first before you learn how to walk and run. Has there been any desire or at least consideration for the bioactivity that I think will surpass the availability of the *in vivo* toxicity data? And are we creating TTCs to predict [inaudible] *in vitro* toxicity classes and then with the [inaudible] models [inaudible] how to relate that concentration to *in vivo* dose, or are we not there yet? And again, this is not a loaded question. I'm just trying to get for the audience an understanding of where we are or are not right now so folks out in the audience can actually start thinking about these new approaches in the future.

Jacobs: I do not know if we are there yet. And that would be my simplest answer. It would not be a TTC topic if we didn't say that [inaudible] how to layer the formulation of this question. Looking at whether or not you can look at structural-based global model to determine whether or not you need to start to overlay more weight of evidence. And if there is any data out there how

you would consider it. Whether or not there are more local models, how you would consider it. I think starting with exposure for me makes a lot of sense. Exposure and also why are you asking the question. If you were to consider the reason why you are asking a question and the potential exposure that you're dealing with, it may lead you down a different path. Depending on your need. The orders of magnitude below whatever the TTC value might be for a given exposure, you could say that is enough. I do not need to look at another model. You have to consider the mechanism of the action. The more that exposure creeps up, the more information you are going to seek. So that as a decision-maker you can say, I have confidence that there is a reasonable certainty of no harm for this particular chemical at this particular exposure.

Rusyn: For food use compounds it is easier to estimate potential exposure and that may or may not be true for environmental or occupational exposures as well. So, I think we need to be careful and not project that exposure as something, as a riddle that it has been solved. It does depend on how local you are in the chemical space. So, Jessica?

Wignall: I would like to talk about consumer products. I am not sure if this has been considered in different committees, but that idea of combining considerations of product type and exposure duration. In the past, there have been considerations of dividing classes. Or having questions related to product type and how products might be used. There have been questions about exposure duration and how often you might use a product. When you're talking about consumer products, which I would argue, even maybe more than food, is an exposure route of concern that would be relevant to are a lot of us in the room. That if you combine lotions that might stay on your skin all day versus something you wash off and considering the exposure data especially coming out of EPA. Being able to nail down how different chemicals are used and in what products and what that use scenario might be.

Rusyn: And the other question, and again we are trying to solve what data, what point of departure, and what other layers of information you're putting on top of that. But I think the elephant in the room is really your confidence or certainty [inaudible]. Are you taking that into consideration? And when you start with NOAEL and you have to think of species extrapolation. And a TTC class is a class, it's just that. There is a range in which you assign the chemical. In the experience of the speakers this morning, what can you say about uncertainty? Is it already enough incorporated because each class is a range of potential exposures? Is that enough of the uncertainty cushion there already or is that one of the barriers for regulatory acceptance of TTC?

Adams: One of the advantages of looking at a large database, you can have different layers of confidence. There may be a class level, but if you look at the particular chemical group that constitutes most of that level, they can be subdivided into chemical groups which can if they are large enough they can develop their own TTC. And so it becomes possible in a group of [inaudible] to define a TTC for 25 of them which have a full range of structural variation that you deal with. And it becomes a question of how much data you have and how many layers you develop.

Rusyn: I think what you are describing is that confidence in actually deriving the classification schema. And then the regulators and the regulator dealing with that one particular chemical. And do you then take the uncertainty of your classification scheme, and slap that on every compound that comes through? Because the regulators worry about the chemical that they need to look at and not necessarily how you have done the classification. So what can be done to increase confidence or reduce uncertainty? Does anyone have any thoughts?

Patlewicz: I do not know how to answer that. But I think it's a critical answer. Making uncertainty more explicit and finding ways that we can quantify it is really critical. And I welcome the talk that Jessica gave. That despite some of the inherent data variability that we have, we still can do something with that data, but we need to get a better handle on quantifying uncertainty and identifying all those sources of uncertainty.

Rusyn: Kristi, do you have something to add? Looking at the report that's coming out at [inaudible] is there a part of it that deals with uncertainty and variability?

Jacobs: I think in general the literature on TTC has demonstrated that it's reasonably conservative with prioritization and screening purposes. I think they also acknowledge that, I think Grace used the not get out of jail free card. It is not meant to supersede a regulatory requirement for data and that's important to acknowledge as well. But I think, a couple of times today we have seen the distribution curve of NOEL for Cramer classes I, II, and III. That is a very broad distribution. And the TTC value is derived from though lowest 5th percentile with a hundredfold safety factor on top of that. It would be nice to see a large study, I would read a paper that were to come out and to say, I am going to compare every reference dose and every safety value and compare that to what TTC tier it would be and figure out what is the ratio of those numbers. That would be an interesting paper. I certainly would read it and I haven't seen it. But I think anecdotally the experience that we have is that that TTC value is generally protective enough. That the [inaudible] and reference doses that are being calculated for individual chemicals don't tend to fall lower than those.

Wignall: I have a question, and I think Andrew you alluded to the fact that OECD QSAR toolbox and Toxtree for a given chemical generated different classes for that same chemical based on the answer to one question, and there's expert judgment needed. But is there a way to quantify when expert judgment would lead you to one answer versus another? Or is there a way to identify the decision points at which two different people would make a decision and that's where your uncertainty might come from?

Worth: This is what the authors of those papers did, in fact. I think it was somewhere between 75 and 85% concordance between expert judgments in the predictions from either the toolbox or Toxtree. And for some chemical classes it's higher, and for some it's lower. And another thing, it may not be the source of uncertainty. There are different ways of deriving dietary exposure estimates. I mentioned having the [inaudible] survey [inaudible] other methods and that's where you almost need to consult an exposure scientist, so you do not have to worry whether it's Cramer class II or Cramer class III. That is dominating the overall assessments of margin of safety.

Rusyn: I think this borders on my next question, which is clarity and complexity and confusion that some of these predictions generate. And this all, I think, comes down to how we communicate this information to the regulators and also the regulated. I think many of the decisions are made at small and medium enterprise level, where there is not a lot of expert judgment. I think it is good that we should think of this. But people on the ground really are the ones that are using this. I want to start a discussion or get your opinion, what is most important in terms of communicating the outcome of the classification? And having a confidence to a certain degree decision? What regulators are most interested in about? Are they interested in how open and transparent the tool is? Or do they really need to understand what data went into building that particular classification schema? Or is it just the ease of use of the particular tool and they'll use whatever is easiest and doesn't involve a lot of expert judgment and they would have to grapple with differences of opinion that when they use more than one tool? So, the

communication part and the transparency and [inaudible]. So, can we have a discussion on that? And Andrew, maybe I'll as you first because you have to go and communicate with decision-makers. Because your day job is that you are an expert. [Inaudible] a lot of people who are using this information are not experts and to them, it's a [inaudible]. What has been your experience? And maybe this applies to TTC specifically about the types of questions and what [inaudible] regulators are asking you the most about?

Worth: I think that is a difficult question but we do need transparency. For example, transparency over the algorithm, transparency over the underlying data, and so forth. And if you provide all of that, you can provide a package of information that's very substantial indeed. But I think we are missing something. The people's confidence does depend on the kind of approach. If it's QSAR and it's [inaudible] there's a tendency to disbelieve the output. If you are getting the same prediction from grouping and read across approach, the inherent sense is that it makes sense. This is semi-automated read across which is a kind of hybrid between local QSAR and read across. It is very interesting, to see which approach is accepted. I don't think it's really ultimately about, too much, about the volume of information you provide and whether it is commercially and publicly available. But it is more about the sort of believability. And maybe the best models are not believable, but they are true. Because the human mind can only grasp so much. It's a conundrum. I don't have a better answer than that, I'm afraid.

Rusyn: Well, it's easy to give these [inaudible] answers, but at the end of the day it comes down to a particular person, staring at a particular chemical. And maybe the cartoon that you put out is really reflective of what goes on in that room, but somehow the decision is made. Kristi, again, maybe not speaking for the Agency, but in your particular experience, what are the aspects of people coming out of that decision room, and feeling semi-good about the decision they have made?

Jacobs: For one thing, experience that the agency has had so far, is the adoption and the use of ICH M7 guidance that looks specifically at using [inaudible] approaches for impurities and active pharmaceutical ingredients. In that guidance, and the approach that they have taken there, is that they would like to see both approaches used. They would like one predictive model that's kind of the black box sort of a model approach that you mentioned, and then the other would be an expert rule-based decision-making approach to the *in silico* QSAR assessment. And that guidance, if you use it as a measure of what people may feel comfortable with, if you use that guidance as an example, they need to see both approaches taken. And that it is the contribution of the purely model-driven black box type chemotype, or, I'll use the term chemotype here just to use the term chemotype, that approach taken where models are derived completely *in silico* versus an expert rule based, the two together provide the necessary information for the regulators to be able to rely on that decision.

Rusyn: But that would imply that you need to apply some sort of weight of evidence, right? So you still would trust the expert or the black box more. Are there any preferences? And again, there are no fast rules for that, but have you seen some preferences? Is it truly psychological, that I have known this person for 20 years and I just trust them more than some OECD toolbox thing that I do not know the people that build it? So, I think it is, as Andrew said, very person to person and personalizeable decisions, but can we move away from that? Or do you see any effort to move away from that and standardize some of these decisions?

Jacobs: As a living breathing person, I certainly don't want to remove the person from the decision-making process. But just to clarify, and I am not a psychologist, but there are some of these even standardized approaches that are more based on expert judgment and expert rule

making. If we think about the expanded decision tree that Tim talked about earlier today, the decision on which classification, which class a chemical gets [inaudible] is based on the experience of what we know about the metabolism and the experience about what we know about how a specific chemical exerts its toxicological effect. As the expert, the rules can be derived. So those are expert derived rules systems. You can then overlay a chemical program which can then do that process for you, as has been done with Toxtree and the OECD toolbox and others. And that's what I'm considering as an expert rule-based approach. Not necessarily that an individual sits in a room and makes a decision, but they've used that sort of *in silico* approach as opposed to something that might be more like a more traditional black box QSAR approach which is going to make the decision just based on the chemical connectivity and the different chemoinformatics type information and deriving models based off of that.

Rusyn: I agree with you that as far as job security of regulators, we're nowhere near computers replacing regulators. I did see couple of, at recent SOT meeting there was a poster on using supercomputing such as Watson to make decisions. And I think it's an interesting approach to take, but I do not think toxicologist will have to worry too much about being replaced by computers in the next 50 years. Speaking of 50 years, so Tim, because you have been doing this for a while. So reflecting back on your experience. Do you see a measurable change in attitudes that people have towards experts versus machines versus some sort of hard and fast approach, just push a button and get an answer, and go home at 5:00. So, has that been the experience that you had in the last decade or so?

Adams: You should make that forty decades, I mean four decades. For about 20 years of my life, I worked in the industry with an expert panel. We evaluated flavors on a regular basis. And their attitude was that, and this is even for experts, they would not worry too much about a very low exposure substance and they could, they'd ask the chemist, is there any good metabolic pathway that you can envisage here that could take care of this, oh yeah, don't worry about it, this gets metabolized [inaudible] and it gets excreted. But if it came to a compound that caused a tumor in an animal that was somewhat controversial, or no one had looked at the slides, let's say a [inaudible], they then the chemist would step back, and the pathologist would come forward and would say now hold on a second. He said this is getting closer to me and my animal and I need to see the animal. So, the more important the question, the closer you want somebody to the animal species, and the way they want more information in terms of expert judgment, they'll always migrate there, even for experts. And the people who then hear the results of that, they have more confidence. They know that an animal is connected. And they would rather, they think of their pet at home, but that's how they connect to the decision that's made.

Rusyn: Does anyone else want to weigh in? And I'm not implying that people had decades of experience but still, at least some experience. And do you see changes in attitudes? Andrew, you have been dealing with regulators for quite a while. And OECD QSAR toolbox is a relatively recent development and I think if you talked to regulators, they really [inaudible] that particular effort, but I'm not sure how much it's actually being used because it is a black box versus an expert judgment. So, do you see a change in attitude on the regulators side?

Worth: From the colleagues I interact with at [inaudible], I think we are seeing that, but what we're also seeing is greater attempts to codify expert knowledge and to distinguish between predictions and conclusions and to have sort of workflows that really are a combination of the two, but at least you can trace what is a prediction, what is a weight of evidence based on multiple predictions, and where does the human factor come in? And maybe we haven't quite done that well enough, but I think we have seen a lot of progress in that respect. And I think in

general, people are sort of accepting these new methodologies but in that context, you know, these decision support tools rather than decision-making tools.

Rusyn: If you think I ask difficult questions wait until you hear the one that came from the web. And I wanted to use this one to kind of wrap the discussion. I would like to get an opinion from everyone, and maybe a short one. The question is, what's the most important thing we should keep in mind when considering the use of the TCC? How do we improve regulatory acceptance of the TTC?

Adams: From my point of view, TTC is a prioritization tool that, it leads one to classification for further judgment and further evaluation. You need additional information for something that follows a higher level of toxic concern. And the first, as Kristi noted, and others, first bit of information is what level of exposure you have. In terms of my response, at the [inaudible] level you're at, the better you have an exposure that you can explain, the better communication you're going to have going forward.

Patlewicz: I'm going to cop out of this one and say I think I would agree with Tim.

Worth: I think this question has come up about not only about TTC but about QSAR and other methods. And I think my answer would be that given the caveats around the approach, it's better than nothing. If you don't have toxicity data, and you don't have the legal mandates, the required toxicity data, then at least you have the basis for making some decision.

Wignall: And I think I would echo what Andrew's saying in saying that it's the use context and that depending on your decision context and what kinds of data you have available, TTC allows you to organize and process that data, then you should apply those kinds of tools. And if you have other data to inform, then also apply those. And so what kind of uncertainty you can handle as you make your decisions and then finding the tools that allow you to make decisions within the realm of that uncertainty.

Jacobs: I think the same thing that would probably apply for any scientific field, advancement in that field, and that would be more data sharing and quality information being available. The more high-quality data that can be pulled into modern databases, the more confidence in general that people are going to have in what we do with that information.

Rusyn: Thank you so much and it has been a very interesting day. I'd like to thank the speakers again and I think they deserve a round of applause. I'd like to thank the audience both here in the room and on the web for sticking around and listening in. I'd like to thank our hosts and sponsors, the Food and Drug Administration, for supporting this activity. I'd like to thank folks at the headquarters of the Society of Toxicology for making all the arrangements and being a great help in organizing these. And then finally, I'd like to invite all of you in the room and on the web to review previous offerings through this FDA SOT Colloquia because all of these sessions are recorded, the slides are posted, and you have a chance to share them with your colleagues and this is at no cost to you. And again, thanks to sponsorship by Food and Drug Administration. So, thank you again for participating and we will see you in May at our next colloquium. So, thanks again.