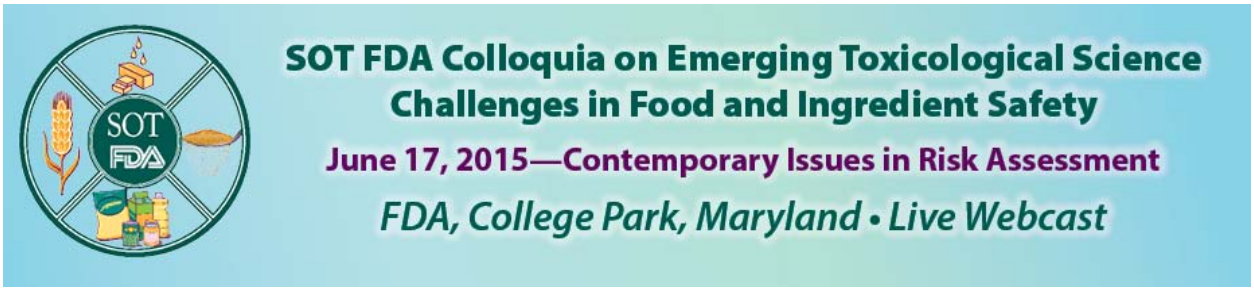

Real Time Captioning

Schedule

8:15 AM–8:25 AM	**US FDA Welcome and Overview** Suzanne Fitzpatrick, Colloquium Chair, US FDA Center for Food Safety and Applied Nutrition, College Park, MD
8:25 AM–8:30 AM	**Welcome from SOT** Ivan Rusyn, Texas A&M University, College Station, TX
8:30 AM–9:15 AM	**Problem Formulation and Scoping for Human Health Assessments** Juleen Lam, University of California San Francisco, San Francisco, CA
9:15 AM–10:00 AM	**Identification and Selection of the Evidence Base for Human Health Assessments** Kathryn Guyton, International Agency for Research on Cancer Monographs Programme, Lyon, France
10:00 AM–10:15 AM	Break
10:15 AM–11:00 AM	**Harmonizing Dose-Response Assessment for Cancer and Non-cancer Endpoints in Human Health Assessments** Weihsueh Chiu, Texas A&M University, College Station, TX
11:00 AM–11:45 AM	**The Use of the Mechanistic Evidence in Human Health Assessments** J. Vincent Cogliano (by webinar), US EPA, Crystal City, VA
11:45 AM–12:30 PM	**Roundtable Discussion** Ivan Rusyn, Moderator All Speakers

---

8:15 am-8:25 am **FDA Welcome and Overview- Suzanne Fitzpatrick, FDA, CFSAN, College Park, MD**

Welcome, everybody to the fourth in our SOT FDA Colloquia on emerging toxicological science challenges. This one is Contemporary Issues in Risk Assessment. I think it's a culmination of all of the

four we're putting together. Different issues in risk assessment. This will cover issues that we're interested in here at FDA. We see them as challenges. We're looking forward to hearing our speakers.

I'm here to give you -- to thank SOT, especially Betty, our guiding light here, and Peter as the president. Alan Rudman chaired our committee and everybody from both industry, academics, we thank you for a very successful series that we hope will continue.

Coffee in the lobby. Bathrooms are out there. No food is allowed in the auditorium. And with that, I'll get off the stage and introduce Peter Goering, SOT President and senior scientist here at FDA.

8:25 am-8:30 am Welcome from SOT–Peter Goering, SOT President, US FDA, Center for Devices and Radiological Health, Silver Spring, MD

Thank you, Suzy. On behalf of of the SOT, I'm pleased to be here and I'm pleased to see the ongoing success of the collaboration between the Society of Toxicology and the Food and Drug Administration.

This colloquia series grew out of a Memorandum of Understanding that was assigned between the two organizations about four years ago. And this is certainly a key marker of success of this MOU. And we hope that it can continue next year. And we have some good information that we might be able to continue this next year. This collaboration is very important. I think the SOT and FDA share common goals in terms of advancing the science to improve public health and advances in regulatory sciences, where both believe in strong efforts training of the workforce and these colloquia will be recorded. And

I'm pleased to say that we have about 400 people participating by WebEx today. We've had four of these in the past. The first colloquium was on the partially hydrogenated oils. We had a significant number of participation. I want to make clear to you that these videos and slides of pest colloquia are available on the SOT website, toxicology.org/FDA. We have the video recordings of the first session and we already have data on the number of views and the presentation downloads from these colloquia. We had a second one on February 23, regarding pharmacokinetic principles to improve safety assessment of foods and cosmetics. And in April we had the third one on immunotoxicology of foods and ingredients. And we had a good participation on-site and by webcast. Today we have about 80 people who have registered on-site and over 400 on the WebEx.

So this is a big success. Risk assessment has come a long way over the past few decades. We've made significant advances in establishing good tools, and good principles, to improve public health through risk assessment. Because of the importance of this tool and the various approaches, I wanted to share a little bit briefly about the foundation for risk assessment and some relationships possibly, the earliest collaboration between SOT and FDA.

Arnold Lehman became the chief of the division of pharmacology at FDA in 1946. And this postwar era was just an explosion of commerce and economies around the world. Foods were being developed, new chemicals were being developed, food packaging materials were being developed. So it was quite a challenge for Dr. Lehman and his team to develop new regulatory approaches. He was also a cofounder of the Society of toxicology. He was asked to be the first president, but because of his employment at FDA, he declined that for a potential conflict of interest. And he agreed to be the honorary president that year. Dr. Lehman pioneered new approaches for safety assessment. If you can believe this, he actually had company representatives bring data to his office. They would discuss this data and decide on safe levels of materials. We've come a long way since then. That sounds like a very easy regulatory

approach if you ask me, but hopefully what we have now is provides good safety information. So Dr. Lehman had a sign in his office. And it said, you, too can can become a toxicologist in -- two easy lessons, each one taking about 10 years.

As I thought about today's colloquia, I think it fits that you too can become a risk assessor in two easy lessons. Each one taking about 10 years. So another pioneer in our history is Dr. John Doull. He has reached a milestone. He is still active in toxicology circles.

At the age of 90 plus. I spoke with him on the phone last week. His professor emeritus at the University of Kansas medical school. He was our society's president in 1986 and 1987. He's the one that told me that he and a colleague would come to Dr. Lehman's office with their pathology slides and they would say I see this through the microscope. I think the safe dose would be 20 milligrams and someone else would say I'm seeing this in my pathology slides. I think we ought to lower it to 10 milligrams. It would say okay, it's 10 milligrams.

John Doull has -- I've heard him say that toxicologists have two roles. Toxicology is what we do, risk assessment is why we do it. So toxicologists have two functions. The first is to elucidate what materials are hazardous. We elucidate the mechanisms. That's only half the job. Many people, scientists, think we can stop there. John Doull posits that we need to carry it a step further and use that data in risk assessment to really achieve advances in public health. I also want to say that the SOT gives a very prestigious award every year called the Arnold Lehman award, and it's given to an individual who's made significant advances in risk assessment. So it's a very good way to honor this pioneer in our field.

So we have today's agenda. We have four expert speakers today. And we have a break at 10:00. And I want to thank Dr. Suzy Fitzpatrick and Dr. Ivan Rusyn for organizing today's session. And I also want to thank our committee of SOT members and also of FDA CFSAN members who have worked hard this past year bringing these four colloquia together. And I want to introduce now

Dr. Ivan Rusyn, he is the moderator of today's colloquia and he will be introducing the speakers and moderating the Q&A at the end of the session.

Colloquium Introduction–Ivan Rusyn, Texas A&M University, College Station, TX

Thank you, Peter. Thank you for coming and thank those of you who are joining us on the webcast. The purpose of today's colloquium is not so much to be a risk assessment 101. As Peter mentioned, these are two easy lessons but they take more than half a day. What we try to accomplish in the colloquium today is to balance continuing education and practical advice as to people start implementing some of the new things in risk assessment.

In my opinion, last year and this year are the years of systematic review and big changes in risk assessment as how it is being implemented. So we have an exciting set of speakers and I would encourage you to be actively asking questions. We'll have five minutes or so after each talk and then we'll have a moderator discussion at the end. Please decide whether it's a clarifying question or a topic you'd like to explore more.

Our first speaker is Dr. Juleen Lam, from San Francisco. She's been one of the pioneers of problem formulation and the NRC 2009, took that too hard. She has all of her degrees of late from Johns Hopkins.

And was at Johns Hopkins and also US ETA before six months ago, crossing the great country and settling in San Francisco.

8:30 AM–9:15 AM Problem Formulation and Scoping for Human Health Assessments– Juleen Lam, University of California San Francisco, San Francisco, CA

Great. Thank you so much, Ivan. Thank you to everyone for attending this and for those of you on the webcast as well. Good morning. Thank you to the SOT organizing committee for inviting us speakers to be here today and speak with you all. I'm just getting myself oriented here. So my talk today focuses on problem formulation. And scoping steps for human health risk assessment in the regulatory and policy world. Identifying tools and recommendations that are currently in development as well as in application to improve these particular aspects of risk assessment.

First, declaring I have nothing to disclose. No financial or otherwise conflict of interest. I'll be talking about four main topics that are relevant more broadly to risk assessment and it's applicable it before human health assessment. Talking about these assessments as a tool for decision and policymaking. And focusing on the initial steps prior to doing that risk assessment of scoping and problem formulation. And how these set the stage for risk assessment.

I'll be focusing on systematic review methods and how these are applicable and what potential they have to integrate methods learned from systematic review into risk assessment in particular into problem formulation and scoping. And how they can help improve these parts of the process. Lastly identifying challenges and opportunities for the future. In particular for food, and environmental chemicals generally, we are aware of a lot of chemicals found commonly in measurable amounts in various food products within the environment, and in particular for myself at UCSF, working on the program on reproductive health and environment, we focus on the reproductive stage of development through pregnant women developing children, and we know they are commonly exposed to many chemicals in the environment both at home, in the workplace and generally in the environment that they are exposed to. We know that many of these products come from food. A lot of these are in measurable amounts and come from a variety of different exposures in the environment, be it something that the animal is exposed to, or something that has bioaccumulated up the food chain or something that is intentionally added or sprayed to fruits, vegetables or other package products in food.

We know that a lot of these are found in measurable concentrations within people. They really are found in everyone. In particular the 2011 study coming from UCSF found measurable levels of over 43 chemicals in pregnant women. In particular for food, we know a lot of these have been found in measurable amounts even though they've been historically banned in the US for over three decades. These PCBs as well as DDT and DDE, although they are no longer intentionally added or in the food chain, they are still present in the environment and people are still being exposed to them. Finding measurable concentrations of these in pregnant women. And we know that we can make a difference. So we know that through policies and actions on the population level, we can actually see this change physically.

As an example when California banned PBDE in 2006 as used for flame retardants in furniture products, in San Francisco we found a measurable decrease in PBDE levels in pregnant women. A decrease of almost 40% from the time period 2008-9 up until 2011-12. We know these types of policies can have a high potential to impact the population's health. This brings us to the question of how we go about

making these decisions. And set the exposure limits to the myriad of chemicals that are out there in the environment.

Risk assessment has been a key tool for informing risk management and decision making. These methods have been refined and improved over the past 40 years ever since federal agencies began doing risk assessments early in the 70s. In 2009 the national academies published science and decisions. This document was pivotal in how we organize, develop, implement and interpret the results of risk assessment. It also addressed methodological challenges within certain parts of risk assessment like how we handle uncertainty and variability. Some of these issues will be discussed later on in some of the later talks.

In particular, science and decisions highlighted this first phase of problem formulation and scoping. Prior to the actual conduction of risk assessment, in how we plan and what issues we'll be addressing in risk assessment. Friends and decisions really reiterated certain criteria they had developed in 1996 document for how they define a successful risk assessment. Emphasizing even over the course of 13 years, these concepts still held true.

My talk focuses around these six criteria and how we incorporate new methods in risk assessment to achieve these core ideals of ensuring we're asking the right question in risk assessment, we both get the science right and get the right science in helping inform our decisions, getting the participation right, as well as getting the right participation, and developing an overall accurate, balanced, and informative synthesis of the evidence. And this is really important because the science is the most important part of this. And we do know that for many of the decisions that have been controversial and challenging, that the actual scientific information that these agencies are using isn't really what's coming under fire. It's the disputes over how that science is interpreted, inferred, what kind of models we're using and other similar policy type issues. Not necessarily the soundness of the underlying data. I'll start by talking about these two concept of getting the science right and the right science. This process is centered around the science base we used to identify, evaluate and assess health risk assessments. So how does one address this problem? How do we look at drawing interpretations from information related to making this decision? And justify it in a way that minimizes potential disputes from stakeholders? This might be one of the most challenging aspects of this whole process. Two recent national academy reports published in 2014 included certain recommendations that addresses this issue.

Incorporating system review approaches into the evaluation of the health literature in this evidence integration process were things that were identified in both of these documents. So systematic review methods have been used and developed and applied for many decades in the clinical health field. So those of you who may be familiar with reviewing evidence for pharmaceuticals or healthcare interventions, this is something that's been around for a long time and has empirical evidence in how they work and can be used to inform these types of decisions.

Systematic review methods are relatively recent for environmental health literature. It's a new topic but it has been developing at a very rapid pace because of the rich body of evidence and literature that's been available from the clinical science field. So there's a lot of challenges that arise when we're comparing the evidence base from clinical science to the environmental health. The flowchart of when the exposure occurs and when we evaluate the literature are the key differences that distinguish these two fields. So in the clinical sciences, typically a new pharmaceutical is developed or new health intervention is developed. There can be an array of in vitro and in vivo toxicity testing, but most of it hinges on the human experimental studies. Randomized controlled trials where people are randomly

exposed or not exposed to a particular pharmaceutical or health intervention. It creates this rich body of evidence for applying systematic review methods and making final decisions. These methods have been around for several decades. Now they directly inform both clinical and health care decisions. As well as billions of dollars in healthcare spending annually.

In contrast we have the environmental field. Where the exposure and the toxicity testing are essentially flipped. Now we have chemicals being introduced or developed and put out there in the marketplace with very limited health effects, evaluation. And it enters the marketplace. The exposure is occurring and most of the evidence base comes from ad hoc postexposure observational studies or randomized trials in animals. So developing systematic review methods for this evidence base is very challenging because the types of evidence that we are getting is very different. We often don't have randomized controlled trials humans. That makes it difficult to apply methods directly from the clinical field. We've been working on this for about six years, taking the methods developed in the clinical field and adapting them for environmental health. Drawing upon things we know are empirically shown to be successful and figuring out how we can adapt these for the environmental health literature. These also have tools to inform policy and health decisions and potential to impact a lot of lives from these decisions and policies being made. One thing I do want to highlight is that in comparison, the exposure is happening at a different time point. We are developing this literature base and making policies prior to the exposure. The exposure is already okay rank in the population. It's essential we are taking these methods and applying them in a timely manner so we can make decisions and make policies.

Now, how can these approaches from systematic review be used for risk assessment? I'm going to back up a little bit first and talk about the planning and scoping and probably -- problem formulation. Starting with planning and scoping, what is the step? It's a discussion between the decision-makers, risk managers who are making the policy and decisions at the end, with relevant stakeholders. The risk assessors have a -- a supporting role where they contribute the information they know but it's this conversation between those who are impacted and affected versus those who are making the decisions. We have a determination of the hazards, potential mitigation options and a decision on the scope of the risk assessment and what question is going to be answered and addressed in this assessment. The table is adapted from the science decisions report, where these are the key elements that are discussed at the planning and scoping phase. You can see the balance between things that impact the stakeholder, looking at sources of exposure, potential exposure pathways, what the hazards might be, and then balancing with that, with what the risk managers can contribute, looking at here are the sources of exposure, what are the mitigation options we have for mitigating those exposures? What are the options for mitigating the exposure pathways? These are the things that are discussed during this planning and scoping. This conversation between those affected and those making decisions with risk assessors being there to absorb this information and figure out how it's going to be incorporated into the risk assessment process.

That's where this next step, the part two of phase one comes into play. Problem formulation is a discussion between those making the decision and the risk assessors as well as any technical stakeholders who can contribute to this process. What you're doing is taking the discussions that were coming from the planning and scoping and actually putting that into a detailed technical plan for how we're going to do the risk assessment. This is all part of the planning process and all of this is linked to the policy that is ultimately a potential option. In problem formulation, it identifies the sources, environmental stressors, exposed population, trying to figure out how we're going to tailor the risk assessment to address these issues. I want to highlight that planning and scoping is something that happens first where you are addressing these issues with those impacted stakeholders but problem

formulation is something that occurs in parallel and along with the planning and scoping. It's the feedback loop of going back and forth between planning and scoping and problem formulation and continually changing the technical plan so that your addressing all the concerns coming up during planning and scoping. This is planning. We're trying to figure out how we want to do this risk assessment. What are the technical aspects of the risk assessment? But ultimately, after -- when you are planning the risk assessment, it comes down to asking the right questions. How are we going to address a question that's relevant to stakeholders and explain it in a way so that it makes sense to everyone? So everyone knows what your addressing and have you make sure it's the right question to be asking? In systematic review this is one of the tools that they use in order to formulate the study question. It's called the population exposures compared to group and outcomes. It sounds like a very simple concept but it's very challenging to do before hand where you're looking at potential exposures or health outcomes that you're interested in and trying to hone in on who you're interested in addressing the study question four, what are the potential exposures that might be of concern, what are the compared to groups you're going to using your evidence base, and what are the health outcomes of concern? So this is an example of a study question formulated as a PECO statement.

One of the first case studies that we did applying these systematic review methods in environmental health. Our study question was a pretty simple one. We wanted to start out with a proof of concept state -- case study. Our study question was looking at fetal development told exposure to PFOA. And starting out with the basics about the PECO statement, the population being assessed in this part of the case study were humans studied at a reproductive or development of time period before and/or during pregnancy and fetal development, looking at exposure and classify and how we're going to looking at -- to be looking at PFOA during the time before pregnancy and/or during pregnancy for females, looking at the comparator group of comparing humans exposed to lower levels of PFOA than the more highly exposed humans, making sure you're looking at this comparison across a range of exposures, and our outcome was one relatively -- looking at effects on fetal growth, birth weight or other measures of length. I'm not going to go into the approach for this. Again, this might be something that could take not 10 years but a significant amount of time. I want to show it highlight our findings from this case study.

Our ultimate conclusion was that PFOA is known to be toxic to human reproduction and developed based on sufficient evidence of decreased fetal growth and humans. The PECO statement -- we did evaluate the evidence for humans but we also looked at nonhuman mammalian species. So taking us through that systematic review process we had defined steps for how we were looking at both the human and nonhuman evidence base and coming to these final conclusions. But ultimately we come up with the bottom line statement based on the evidence through systematic and transparent review to make a conclusion about our study questions. These findings were cited in a recent regulatory rulemaking proposal by the European chemicals agency that is proposing to restrict exposure to PFOA.

As you can see from this PECO statement, it's a proof in concept case study. It was relatively simple. I forgot about this slide. This shows you how we typically do things in environmental health. A lot of the times we'll do what may or may not be a systematic review where review authors go out into the literature, identify a relevant study base, talk about the findings from this and that study, put it in a table and are asked to make a final bottom-line recommendation. Most often, you're going to find a statement that looks something like this. Inconsistent associations are several birth outcomes including all of these outcomes related to fetal growth. Among the five population studies that measured in this paper looking at both PFOS and PFOA. The command in birth weight was associated with PFOA exposure. It could be similar, it could be stronger, it could be weaker. Not helpful in terms of making a decision. One of the key issues with these narratives, expert-based reviews is there isn't a systematic

process where you can see how they evaluate each one of these studies, how they come to final recommendations on the overall body of evidence. It's this one by one study and then one study seizing positive finding, another one finds a negative study. In systematic review procedures and in the navigation guide approach, there are set criteria for how we're evaluating the body of literature as a whole and what constitutes evidence that would either cause us to upgrade or downgrade our findings on the strength of the evidence for answering our final study questions.

So this is a comparison on how systematic reviews can help increase the transparency and reproducibility of making these final decisions on the evidence base, comparing to these expert-based reviews that have typically been done in the environmental health literature and may or may not be very useful for making decisions. So now coming back to the simple study question that we started out with, you can make something -- change the study question to make the question more complicated, to assess either a broader set of exposures or a broader set of health outcomes. So this is a draft protocol put out by the national toxicology program where they also have a very similar process for systematic review approaches and evaluating the environmental health literature. And they are asking a slightly different study question. Looking at PFOA or PFOS collectively and looking at a broad health outcome. So the population is similar. The exposure is similar, but they broadened it out to PFOS and more specific about how these exposures will be measured. And this informs how their inclusion-exclusion criteria for the body of literature will be drawn upon. But their health outcomes are what are more complicated. They have these primary outcomes. These are the only ones I'm showing but they also have a long list of secondary health outcomes. This can be challenging because you are asked to outline your collection of health outcomes at the very beginning before you have technically seen the body of literature. This is really important to be able to talk with stakeholders and risk assessors and figure out, what are the health outcomes we're interested in? Also talking with risk managers to see what are the health outcomes to make decisions on and can inform policies? And asking the right question can also be challenging in different ways. Not only just what you're addressing in the general question but who are you asking the question for?

This is where the stakeholders come and play. Switching gears and showing you a different case study looking at blood levels in children. And looking at setting -- the EPA and set of studies they use for reevaluating their and a seat in 2008. Looking at setting blood levels and health outcome they were looking at was a change in intelligence measured by IQ or MDI score. This is the set of literature that the EPA used for making their decisions. I think they reviewed 6000 studies and came down with these four most relevant studies for answering the questions. What these estimates are showing are an estimated change per 1 microgram deciliter, across children. This is generally across a collection of children. So using this set of studies, they actually came up with their final recommendations and it was .15, I believe, micrograms per meter cubed in air. You will notice I also showed categorized by high STS children versus low SES children. This is the pattern where high SES children are showing less of a decrease in intelligence compared with -- to the mean where as low SES children are showing this greater decrease. I'm so sorry. So socioeconomic status. It's an indication of the economic status of the child. A high STS child would be somebody who has a little bit more money, in a better neighborhood versus low SES children. Thank you for asking that question. So children who are -- have high STS, doing better economically in terms of family, have less of a decrease in intelligence compared to the average child. The average child is what the EPA is using for making this decision. If you are looking at lower SES children who might be exposed to more environmental stressors, might have more economic challenges in their life, actually have a greater decrease in intelligence compared to the mean. When you are stratifying by economic status, you see this difference in children but if you're setting this regulatory decision or this decision based on the mean of children, you might be missing the sensitive population.

So this lower SES group has greater susceptibility to lead effects on IQ. This is an important thing to note when you are making regulatory decisions based on the population. Setting regulatory exposures standards at the mean doesn't necessarily protect all of the population. That is the ultimate end goal.

Protecting the population through these decisions and policies. Now, the graphic I showed here is from a Gilbert study in 2006 where you might be asking yourself, a five units the decrement in intelligence, make that doesn't have such a huge impact on the individual. That's what this is showing where you might have a distribution of people, intellectually impaired people with lower IQ levels, the mean here, at 100, and intellectually gifted people at the higher end. This distribution of intelligence across the population. If you're talking about a five level decrement in intelligence on the individual scale, if you look at that on the impaired segment of the population, the five level decrement in intelligence will have an overall impact on the population and will actually move this population level down. That actually impacts the numbers of intellectually gifted people at the higher end. So the small individual difference in intelligence on the individual scale can have a great impact on the population. This is why protect against adverse health effects really does need to consider the most vulnerable or the most susceptible in this population.

So the next topic I'm going to address is asking the right people. So stakeholder involvement is something that was strongly suggested in science and decisions. It's something we really lack. In particular for cumulative risk assessment, affected communities haven't been involved in this process. But it is important to ask them to get involved up front early, ensure balanced participation from multiple stakeholders who might be affected by this decision and be inclusive of a variety of different sectors in your decision. This is important because it increases the transparency of the process so that stakeholders understand how the final decisions are being made. They have a say in the decision, they can contribute into this planning and scoping stage of the decision where they can identify the types of exposures and health outcomes they as a community or they as any other sector of stakeholder might be most interested in, and can inform the process of how we actually do problem formulation for risk assessment. It also potentially can lead to more effective and efficient risk-based decision-making. If you are involving stakeholders upfront and they have a say at a variety of steps in the process, they can follow the steps for how you're making the decision and also have input into the process and potentially highlight what is most important to them.

So in particular for the community, this is who you want to address. To ensure you're asking the right people and addressing the right questions that they are most concerned about. So you want to make sure that they have a place at this table in making decisions. So this is what is referred to as the six chairs model. Seven chairs around the table including the decision-maker and six other potential stakeholders that might be very important to bring to this table. This was published in 1993 by Dr. Thomas Burke who is my advisor at Johns Hopkins. I think it's still relevant today. Asking yourself who needs to be at the table? Who needs to have a say in these decisions we're making? I'm showing an eight chair table here because the original scan from the book is not a very high quality so I was trying to re-create it. Is difficult to find a diagram of a table with seven chairs. It's relevant because when you're making a decision, you never know who else you might be missing. It's important to have a chair and a question Mark where you're looking at the decision you're making and looking at the people around the table and asking are we missing any one? Is there anyone who might be involved who needs to be at this table and having a say in what we're doing? One of the key things that was identified in science and decisions is a lot of the stakeholders who might be most impacted by this decision don't necessarily have the scientific and financial resources to provide timely comments and be involved in this process. A lot

of times these are the people who you really want to be a this table and helping you make these decisions.

The NAS review of the EPA Iris process, published recently, recommended expanding opportunities for technical assistance programs like those of the Superfund program of EPA and extending these to under resourced stakeholders to balance public input. This is actually something that EPA IRIS has been doing in their by meeting -- bimonthly meetings. The NAS agreed to arrange for scientific experts through a contract with EPA to invite them to attend and participate. This is continuing for the APA why monthly meetings. The one that's occurring today, which Dr. Cogliano is involved with, where the stakeholders are being -- experts are being identified and invited to attend these meetings. And provide scientific input. That's a really important step forward in addressing the need to bring the relevant people to the table to have these discussions. But you can imagine as your extending these resources, you're inviting people to contribute and opening this conversation, these things can drag on for a very long time.

So science and decisions was also very adamant in saying that you need to have time limits. You need to set stopping dates for the rules and the decisions and you need to ultimately come to a final decision. This is really important thinking back to the flowchart of where the exposures are occurring in chemicals and the environment. They are already occurring. So we need to be making decisions in order to protect the people's health from the chemicals that we're most concerned about. It's this balancing act of insuring your addressing the right question, but ensuring that you're able to make a decision as an agency. That can have a huge impact on the populations -- population's health. One of the last things is the protocol. The protocol is an essential component of the systematic review. In part, it's a very detailed plan of steps to be followed. So it involves the pico statement. McPeak a statement is a part of the protocol where you are stating the study question that you're going to be addressing in your assessment.

From this PECO statement, once you identify the study question you're interested in, this will outline the inclusion and exclusion criteria for what body of literature you're going to be searching for in the environmental health literature. And within the protocol you have sections that identify the methods you'll be using to search, locate and identify the relevant evidence that will create this evidence base. What your plans are for extracting the data and analyzing it and the key here is that you're developing the plans before you actually see the data. This is an attempt to minimize the bias you might have when you go in and look at the body of literature, and then decide this is how we'll evaluate the set of literature. Forcing you to make these upfront decisions. It also includes details on how you will be evaluating the risk of bias which is internal validity of each study, what the quality and strength of your body of evidence is and how you plan on integrating evidence across different bodies of literature, looking at human literature, animal literature and potentially mechanistic literature.

This is important not only to organize yourself and to incorporate all of the feedback from the planning and scoping and formulation as well as minimizing the bias as you're going through and evaluating this literature base. But also being upfront and letting people know this is how we're going to be doing this evaluation with this assessment. Increasing transparency and also letting people know this is what we're going to be doing. I think that increases the conversation between relevant stakeholders so people can look at a different steps of your process and know exactly how you propose to handle each step.And that's what this slide says. Minimizing bias and how you're evaluating studies, increasing transparency and reproducibility of the process and again, this is Katie. Allowing for stakeholder input in the onset. Allowing them to see exactly how your proposing to do this assessment and allowing them to have input on various steps of the process by looking at how you propose to do it, what they think about that and

providing feedback for anything you might need to change or that you have missed, that you haven't included into your plan.

One of the things we've been exploring at UCSF which is relatively new is looking at registering our protocols. This is something that's commonly done for systematic reviews in the health and social care fields. Something that's gaining attention in the environmental health fields and forth epidemiology studies in particular, it's relatively controversial but I wanted to discuss the tool itself as a forum for pre-registering your protocol before you actually implement your systematic review. So this is a website database that's run by University of York center for review and dissemination. It's an international database and it does currently focus on prospectively registering systematic reviews for the health and social care field. There are a variety of informational health systematic reviews in this database including a couple of the ones that are currently ongoing at UCSF.

So what Prospero does is it creates a permanent online record. You register the protocol -- at the very latest before you are extracting out the data from the studies. Anywhere in between, developing your protocol and your plan up to screening the literature, prior to when you are ready to start a data extraction. You can register your protocol. This creates the permanent online record. Obviously for anyone going through systematic review or health assessments, things tend to change along the way. So there might be a body of literature that you had not anticipated. There might be statistical measures you hadn't incorporated into your data analysis plan. These are things you can change in your protocol but through PROSPERO, it creates this documentation every time you go change the protocol. And you are asked to provide justification for why you made that change.

This is a key step towards increasing transparency of this process. By putting it online upfront and allowing for changes in the process along the way. Along with justification for why those changes are being made, so that it can be understood by all who are tracking this process. I'll show up really quickly a screenshot of one of the case studies that we are working with. Systematic review looking at development of exposures to PBDE's inhuman neurodevelopment. This has been registered in protocol. The web link is right here below. There's an individual unique identifier, CRD number so you can search for that. And there's also a screenshot on the sidebar of when you go onto this website. It allows for revision notes. So we uploaded this document in -- on April 20 see. We haven't made any revisions but we anticipate having some revisions. Each of these will be documented along the way in this process. UCSF has been developing this navigation guide systematic review process.

The National Toxicology Program and the office of health assessment and translation has been working in parallel on their own systematic review. I showed their PECO statement statement. They've also been working on developing this handbook. This handbook is something that outlines their approach to conducting systematic review.

So it's like a meeting to plan a meeting. But it's very important because it outlines how you are expected to be conducting these systematic reviews. Outlining the key terms and defining them so they can be used consistently across all case studies. Outlining how you plan to do problem formulation, scoping and subsequent steps of the assessment, and the idea behind this handbook is to explain this is how we're doing systematic review and ensure consistency across different systematic reviews, increasing transparency and setting expectations for stakeholders who are following this process. Outlining exactly how you're going to be doing this process will display what to be expected when you're doing a systematic review so that people can understand what to expect and how you plan to do these

reviews. There are still obviously challenges and many opportunities for systematic review and problem formulation and scoping in general.

So again consensus among the involved parties is a key thing that I alluded to a little bit earlier when you are involving stakeholders and trying to get balance and adequate participation from a variety of stakeholders. There's going to be a lot of different opinions for how you make these decisions and what you ultimately end up doing in terms of risk management. So this requires a balance among competing values looking at different stakeholders, allowing for their input, but balancing this and understand you may not always come to a consensus, but trying to incorporate the input from stakeholders as best as possible and remembering that ultimately a decision does need to be made. Ensuring that you can hear everyone's voice but in a timely manner. Looking at the assessment in general, when you're doing problem formulation and planning and scoping, you want a broad enough scope so that you're capturing potential effects and the affected populations of interest. Again, there's the balancing act with time and resource midstream. So these study questions if they are too broad can require a huge amount of time and manpower. You can be identifying 6000, 7000, 60,000 relevant studies to your study questions and ultimately, for a systematic review you need to be evaluated each one of those studies for relevance, inclusion, exclusion and then evaluating internal validity, combining the evidence to integrate between human and animal and come to a final conclusion.

So with a broad scope, you might be facing too large of a question. With a narrow scope, you might be missing populations that we care about the most, that might be the most impacted. So it's this balancing act and with systematic review, I think we're outlining different steps in the process and hopefully through this transparent documentation of how we're handling the evidence, this will allow for identification of different ways that we could balance between the broad scope and balancing time and resource constraints. In particular I want to highlight that focusing on sensitive and vulnerable populations is something I think we should think a little bit more about because oftentimes these are the ones that are most highly impacted by these decisions. This can help narrow that scope of finding the health effects and people who are most -- have the most potential to be impacted and focusing our decisions on this population base. Oftentimes they have both higher chemical burdens as well as increased susceptibility to toxic effects because they are being exposed to a myriad of different chemicals in different my mental stressors.

In conclusion, systematic review processes are important. They can increase transparency and consistency across different health assessments but also serve as a tool to increase communication with stakeholders. Potentially improve decision and policymaking and we are only six years in two systematic review processes for in rental health. But we've demonstrated already through various case studies, through UCSF as well as OHAT that these processes can be applied to environmental questions. Ultimately remembering this -- these overarching considerations for what constitutes a successful risk assessment, we need to ask the right question, ask and involve the right people and remember the ultimate goal is protect from existing exposures already gone -- already ongoing in the environment.

To acknowledge the SOT FDA food safety colloquia organizing committee. In particular Ivan Rusyn, Betty and Suzanne, which have been extremely helpful in organizing our presentations and getting us all here in this room as well as two of my UCSF colleagues in particular, Tracy Woodruff and Patrice button who have been working on the systematic review for many years. I put three pages of references for all of the references which I will not spend any time on but I will provide my contact information here in the event of any follow-up questions you may have. Thank you very much and there's no clock in here. I have no idea what time it is. Five minutes for clarifying questions.

[Applause]

[Indiscernible -- low volume]

Dr. Lam
I was looking at your face in the audience. You look familiar. Hi.

Unidentified
[Indiscernible -- low volume] -- and I do a lot of premarket reviews and we have a very limited time, 120 days to make a decision. This type of methodology is often not practical for the types of premarket reviews. Sometimes we do review chemicals that have an extensive database. How would you recommend condensing the type of approach that you do or that you've outlined to make it a little more practicable for risk assessors or reviewers like me to try to do a more comprehensive review in a shorter amount of time?

Dr. Lam
Can I ask you a clarifying question? What are some of the aspects of those types of reviews that you think would not be amenable to systematic review? Just the broad literature base or the individual assessments that must be made on each study?

Unidentified
Sometimes we might have a common chemical that has many, many studies associated with it in the database. We have to make a decision. Sometimes we have our own internal data and we have to incorporate the public literature into our review. So we do all that on a very short timescale. When you are outlining involving stakeholders and getting feedback, we don't have the time. As an extensive post-market setting, we do have the time. In a premarket setting, we have much more limited mandated timescale.

Dr. Lam
Absolutely. Great question because I tried to highlight that several times. It's a very challenging thing to not only identify the right stakeholders but get them involved in the process. I agree with you. 120 days is a very short turnaround. It may not be feasible to involve stakeholders from all different sectors, but I imagine there are parts of the systematic review process in terms of how you evaluate the literature base. That can still be done in a relatively short amount of time. Through that transparent documentation of what decisions are being made, what evidence base is being used, that can be something that is provided to stakeholders. So if somebody may not be able to get involved in that 120 day period, you still have this document that outlines exactly how that decision was made. In that way, systematic review can help increase that transparency. I imagine within that short timeframe, you can probably identify stakeholders who are of most importance to that decision. So if you can outreach to stakeholders who might have a say, that he know are interested in this decision, 120 days is not a lot of time, but you can make that effort to get them involved. Having those stopping rules and the time point where we have to make a decision at that time works in your favor because you need to make a decision at that time point. I don't think that excludes the possibility of involving stakeholders. It just makes it more challenging.

Unidentified
I agree. I wasn't saying we're not involved.
Dr. Lam
Absolutely.

Unidentified
The timeline is a challenge. To expand the scope sometimes.

Dr. Lam
In some ways, that does work in your favor. You know you have to make a decision at that time point.

Paul Hanlon, Avid Nutrition
Paul Hanlon, avid nutrition. I think it does a really good job of highlighting one of the issues that we face all the time in having a lot of discussion around problems, defining the problem is great. So building a little bit on the question before, getting those stakeholders involved, so getting back to what Ivan said at the beginning, risk assessment is two lessons of 10 years. When you have stakeholders who are not as used to risk assessment, how do you make sure you keep the problem focused on as you talked about, risk assessment, risk management and not going down and defining the problem as being either only exposure or only hazard? If you are making risk management decisions, how do you make sure that you get all the stakeholders on board that this is a risk decision, not necessarily a hazard to decision?

Dr. Lam
Yes. That is a really challenging question. And again, it's not something that I have a short answer for. It's something we need to work towards. Understanding that this communication between risk managers and stakeholders with risk assessors having that technical assistance of providing that focus for that conversation is key. Having the risk assessor and risk managers trained on that very topic is something we need to do. Risk assessors generally are very poor communicators. And I think increasing that part of training is key because then it will help to formulate that conversation in a constructive manner. I think that's something we need to focus a little bit more on.

Paul Hanlon
Thanks.

Mary Jo, Office of Food Additive Safety
Mary Jo in the Office of Food Additive Safety. I want to make a clarifying remark about commerce makes the laws and they oftentimes put restrictions on what we can do in the premarket arena in terms of not only the time limits but who can share information. And so we have to work in certain constraints. I think I agree there are ways we can maximize, but public involvement and stakeholder involvement, oftentimes we don't -- we can't really have a whole lot of room. So we're trying -- when you talk about balancing, there are restrictions on premarket reviews that don't apply to post-market. So just to be clear about that. Because I think when you look at the clinical versus environmental, food is in the middle depending on whether we are pretty or post-market.

Dr. Lam
Thank you.
-- pre-or post-market.

Dr. Rusyn

Thank you for the presentation and great questions. That was a great start. Our next speaker is Dr. Kathryn Guyton from Lyon, France. The senior toxicologist with the international agency for research on Cancer which is part of the World Health Organization and before taking on that position at the international arena, she was with the US EPA for about 10 years. And she is very well-qualified with degrees from Hopkins. If you want to shorten your time to risk assessor, go get a degree at Hopkins. You can see the trend here. Kate?

9:15 AM–10:00 AM **Identification and Selection of the Evidence Base for Human Health Assessments–Kathryn Guyton, International Agency for Research on Cancer Monographs Programme, Lyon, France**

Thank you for that kind introduction. Very pleased to be here. I want to thank the organizers for all of their hard work in putting this together and thank all of you for taking the time to be here. So I'm going to focus on how do we identify and select the evidence base for human health assessment? As Ivan mentioned, I'm working at the World Health Organization, at the international agency for research on Cancer. I'm going to give you my perspective from there but I will also kind of point to what are some of the common themes that may also apply in your own work? And hopefully we can go into those more in the discussion.

Let me begin by saying I have no conflict of interest to declare. In my talk what I'd like to do is pick up on some of the themes that Juleen elegantly presented to you and talk to you about the protocol that we use at the monographs program. And I'm really going to go into some of the experience we've had recently using some online tools and different techniques to try to address some of the challenges that we face and I think everyone faces in doing these kinds of assessments. And then wrap up to say where we may go in the future as a community to address some of these issues. So just to begin, the IARC monograph program is a program of hazard identification. We produce these books, which have been -- have an orange cover and they identify causes of human cancer. They can be chemical, complex mixtures, occupational exposure, et cetera. They are not necessarily risk assessments.

Hazard identification is a key part of the risk assessment process but risk assessment involves a lot of hire math. And we basically don't do that. We do meta analysis and tests. I'm going to leave the hard problems to Weihsueh Chiu. And he will get into more of those types of things.Juleen mentioned there's been a lot of recommendations on the systematic review topic. I'm showing you hear some publications in the United States national research Council, as well in the international arena. There's been some landmark publications about how this applies in the international scene with respect to World Health Organization, recommendations, key concepts like conflict of interest management, how are you identifying your evidence? And so forth. You might look at all of this and say first of all, two of these books have to do with formaldehyde. So this doesn't apply to me. I think that's actually perhaps not to. Some of these is basic guidance. The risk assessment science has embraced a lot of these recommendations. You also might say I'm a very busy person and I don't have time to read all these books at this US NRC putting out. Let me break it out for you. What is in his books?

It comes down to three things. Evidence, evidence, evidence. How did you get your evidence? How did you identify it? What's in and what's out? How can you communicate it? How are you displaying it? Also what is your protocol for evaluation and percentages? Of these are the themes that Juleen highlighted? I'm going to talk to you more about these in detail. So the IARC monograph program has a preamble we

publish with every book. It really came from the very first one that was published in 1971. Had the first version. The version I'm showing you here is on our website. It's a public document. It was developed at the time when Vince Cogliano was the head of the program. So all credit to Vince. It covers really two aspects. First is general principles and procedures. This covers a lot of the issues like who is doing these evaluations? For the monographs, our evaluations are done by working groups of experts. These experts come from around the world. They've published in this area and we screen them for conflict of interest. We disclose the pertinent interests in selecting them. But we have other types of participants. We invite observers who -- representatives from different national government -- the Secretariat which is me. To help all of this. We do have a timeline. We publish -- announced a topic that's going to be evaluated. Within a year, that evaluation is finished. So we understand some of the time pressures that you are under. A year may sound like a long time, but it goes by really, really fast when there's a lot to do. We have a lot of our nuts and bolts work -- working procedures here. Additionally we layout, how -- what evidence is going to be reviewed? How is this going to be done? How is all of this going to be incentivized to regular -- together?

Now another giant book I have to read, but I want to show you for the people in the room, this is it. 23 pages. You can read this in the time that I'm talking. I'm not going to tell you which is the better experience. You're going to have to figure that one out for yourself. There is a break later. If you're on the webinar, you can go right to this link and you can click through each one of these things and read. It's not a lot of text but it does give a lot of clarity. It's a public document. Everybody understands the rules and this is really, really helpful in avoiding all kinds of problems.

The other thing to emphasize, this is the same protocol, no matter which agent you're evaluating. Chemicals, occupation, something really complex? That gives you a consistency over time, a program that's been around only 40 years. Compared to the greater history here at the FDA. But this is where you can get consistency and where you can get people to understand what do these evaluations mean? This is compared to all the others. So I am talking about our preamble, but think about which of these may be helpful in your own work. So basically, the scientific review and evaluation can be broken down to three things.

There's three types of evidence that are evaluated. Cancer in humans, cancer in animals and the most important and the most fun, the mechanism. These are systematic reviews of the individual lines of evidence and considered together in an overall evaluation. I'm going to describe this overall evaluation procedure very, very briefly at the end. So try to stay awake until that happens. Most of it, I'm going to talk to you about how do we identify what evidence we're going to include? So according to our preamble, the literature is collected by IARC, us. And we provide this to our working groups. But they are expected to supplement this. These are the world's experts. We're counting on them to do searches and find out what happened. So this is a process that's been working pretty well. It can be a very extensive literature, hundreds or tens of thousands of studies. In the old days these were mailed around the world in giant boxes and they were received like Christmas presents. I don't really know. But we moved on to more electronic distribution and you can imagine, this evolved from a laundry list alphabetical to using an indexed reference database, whether you use N note, you can document this in your mind in CBI searches. So that has evolved. I think they're still have been some challenges in this is what I'm going to focus on now. Really the working group has wondered, how did you guys come up with this list? When did they do these searches? What were the terms? Particularly with the mechanisms, there's a lot of these things. Did you look for all of them or which ones did you and didn't you? What about let's say you're doing a topic that's already been evaluated in the past. What about all

those old studies? How do we bring those in? Where they captured? -- were they captured? These are basic questions that apply to any systematic review process.

So I'm going to talk about some of our recent experience and try to tackle some of these. I'm going to talk about a recent evaluation that we did in March of 2015. Specifically on the group of organophosphate pesticides we looked at. I'll show show you some of what we did with respect to that. We have our three types of studies and so let's look at cancer in humans and cancer in animals first. Since I work in mechanisms, I think these other topics are really easy. So we don't have to spend a whole lot of time. So we worked with Andy Schapiro, who is now at the DHS NTP program. He developed this online program called Project.org. You can go to it right now. And log in there and see what it does. So this program allows you to capture -- you can type in a description. This is my cancer in human studies. You can put in what database it is, what was your search text? This is something that I can say we made up but we really didn't. We looked at what a lot of other people were doing. We sent it by some of our expert working groups, you can type that in and all of this stuff is documented. When did you create it? When did you update it? What did you find?

We last updated this a day before the meeting. So this gives you the reassurance. I would say the number one challenge for experts is they didn't find something. Something has been missed. So this is one way you can really address, we looked for this recently. We're up to date. Let's look at cancer in animals. Basic search text. You are seeing that we're bringing back a few more references. You thought, this is cancer in animals. These are [Indiscernible] there cannot be 31 animal bioassays. These are going to bring up studies on our favorite mechanisms, but this is a way you can manage this, document this and this was really helpful to us and to our working groups. One of the things that the system allows you to do is apply these little tags to track the disposition of each study. So they function as your exclusion criteria and you can make up the tags. However you decide to do it. You can see for section two which is our human data, we have a certain set of exclusion criteria and then inclusion criteria. They are either case-control or cohort studies. If you found 700 case-control studies control studies, you might want to divide that pile by occupation, by cancer type, by so forth. You can do all of that. This system will allow you that flexibly.

For the volume 112 evaluation, we actually task to our working group to do most of this work. We tried to learn from what they did for volume 112 and applied it ourselves in the next round for volume 113. For their subsequent review. This is just an evolving process, gaining familiarity and ask me in about a year whether we are still using this or whether we've evolved to something else. But we really did find it to be useful. One of the things this allows you to do, Andy is very good at creating these fun visualizations which show the size of these nodes, how many papers are in each bubble, what was excluded and included, and this is a screenshot but if you are on the website, you can click around and make these come and go. So this shows you for the cancer in humans and cancer in animals, what was searched, what was found and what you ended up with. You might be wondering, there's actually a whole lot of total references. Like 1200. There's very few in these other two categories. So that brings us to the next topic, the mechanism. It's always a big one. It has some special challenges. I'm going to share with you some of the insights that we've had for the volume 100, there was a review of all the known human carcinogens.

This was an opportunity to really look broadly across the mechanistic data and to come up with some advice for the future. We've also had some recent advisory groups way in on this as well. So the mechanism group according to our preamble is to consider representative studies to give a concise description of the relevant data and issues. However, there's an increasing volume and complexity of

this literature. As I said, what is your main fear? Something was missed, something really important. So one of the recommendations was maybe we could use what were identified as the 10 key characteristics of carcinogens, you might be wondering what those are. Hold that thought for Vince. He's going to explain it to you. For now you need to remember this is how many there are. 10. That makes it really easy. That's not an infamous number. It's not join on us. So could we make some progress and systematically looking for these 10 things and would that give us the appropriate coverage that we really would want?

You might think doing this systematic identification mechanistic study is absolutely impossible but I want to point out that Mary Cushman, sitting right here, actually did a project on this when she was at US EPA. And what she did was take a topic and develop a question. She then went to the literature, opened up some recent reviews and started to say what are the mechanistic events? What the totality of things I need to search for? And she identified those. She showed them here. These are the cute little bubbles. These are the different topics. Then she did specific searches on these and went through this whole process on the left-hand side. These eight steps. She published this and I think it was really, really terrific work but also quite groundbreaking to get into this area of how can we take systematic review into this area of enormous complexity and also volume of studies? So basically, for our volume 112 assessment on parathion and the other agents, we took a similar approach. We took our 10 key characteristics, which Vince is going to tell you all about. You can probably imagine what they are. They are going to include genotoxic to you. So you can make a search for things that are related to genotoxic tea. These are our search text here. And then you can document all of this. You can also as I say, if you are not using HAWC, you can capture this in your mind and CBI searches. This brings back quite a few more references than maybe you have ever wanted to look at. On the other hand you've got them in there. And it's about 300 on this one topic. And we found that these key characteristics -- the terms are very much overlapping.

So having a system where everything is integrated into one bowl and once you've dealt with this study, no matter where it came from, it was also very, very helpful to us and our experts. And you might think you are taking a tremendously long time to go through 300 studies. I want to point out that Jason Fritz, he may be on the line now, he can do approximately 90 of these studies an hour. It's quite amazing to watch. Once you get the facility, it's really a process. Is it in and is it out? If it's in, which of these 10 do I put it in? It's a relatively straightforward process. It can be done. It does take time and attention, but it's not impossible. As I mentioned, now we have the issue of what about the things we identified through hand surgeon? This is always going to happen. Somebody says, you may be searching with the best term ever but you didn't find my paper. What about all those articles that we've got in the archives? Parathion was evaluated in the past. In 1983. Trust me we actually have paper copies of all the stuff. And we got out these old boxes and here are all these studies. Most of them concerned mechanisms. Nevertheless we wanted them in front of our working group. So I just typed in all the little PubMed IDs and brought them now into the database. So they are now all in here. And you can see just on our statistics page, I can tell you we had 59 of these that came from direct import, not things we found from our searches. But they are still included. If you go into some of these NRC guidance books, you'll find this is the kind of information they are recommending that you document. Because it's not like these papers fell out of the air. The actually came from somewhere. Whether it came from one of our working group members, from the archives, we have all of that written down and I can trace all that back if asked. So the mechanism as I mentioned is always a little bit messy. But remember if you have something complex, the best idea is to break it down into small pieces. This is what this shows. If you keep clicking on these little bubbles, you can get into our what we excluded, only for three reasons.

It wasn't on-topic. There wasn't any toxicological information for parathion. There's a lot of studies on approximate -- those were not pertinent to our evaluation. So we put them in the own pile. The included studies, we divided them by the 10 key characteristics and other topics we need to cover according to our instructions to authors. All of this is described in our instructions to authors which is available on our website. So it is a public document we did put out so people understood that we were using some of these tools and how we were approaching some of this. So then from here, you can actually on the upper right, you can download references from any of the little bubbles, show the abstracts, you can edit the tags and do all kinds of nifty stuff once you have this here. This is very, very helpful to have a dynamic system like this when we actually have our working group together with us to facilitate discussion. So when Ivan Rusyn says, who searched for this? We assign this to so-and-so and now we can make sure we've covered that topic which as I said is our biggest concern.

So there's other sources of publicly available data. Here at FDA, we certainly do too. For cancer in humans most of the information is in the published literature. We do have a public call for data that closes a month before the meeting. Trust me, people will be sending you their recently minted articles. We have to have a stamp that it was actually accepted. If it has been, we will put it right in front of the working group, no problem. For the cancer bioassays, these are done by a large study, not usually done in the academic setting. They are done by large groups around the world. So we are in contact with the US NTP. You can go to their website and see what they have done. And that's where you're going to get most of that information. For the mechanisms, it's a special case. Most -- a lot of this information is in the published literature. This is frequently voluminous but there's also a lot of data that's available in the public domain from the different governments including the US government, which I think is actually a leader in this area.

For the volume 112, we did include such data including information that may be included in the bioassay reports. This might come right down from York NTP study report. It might come down from another US database like the Tox Ref database. We've also been encouraged to explore using this in the progress in that regard as well. So in the last few minutes, I'm just going to go very quickly to what do you do once you collect this information? I'm going to show you some of the ideas that we're pursuing and go through this a little bit quickly because I know people have different systems but I want you to be aware of how the information I've just showed you can be really, really helpful for your next steps. So I have all my information. Now what do I do? Your grade school teacher probably told you before you start writing, you make an outline. This principle holds true today as it ever did. And we use -- because we have multiple authors contributing from around the world, we have a -- a web-based system to do this. You can use many other tools to accomplish the same thing. You can go old school and you can have a structured set of folders on your computer. If you have multiple authors, you can put this on a shared drive. There's many, many tools out there, open monograph is an open-source one if you're interested to take a look at that. A lot of journals use the same kind of thing. We have our own homegrown system that we use that we house on our own servers and so forth. But because of this background work I've shown you, this outline is now a little bit more supercharged. I can show you the amount of literature that's on each of these topics. I can show you what expertise we need to cover this literature.

Again, our assessments are done by the working groups so I can be sure to get one of those experts on our committee. But if you're doing the work in-house when you go to. Review, when you are engaging with stakeholders, you may want to reach out to somebody who maybe has experience in receptor-mediated effects or genotoxic you or whatever the topic may be to make sure you have that coverage. So the step two, you can actually do this in HAWC. Managing all of your included studies. And Schapiro,

when he finishes -- when he finished his Masters, we were able to entice him to spend the summer with us to try to -- he really expanded HAWC to deal with a lot of the issues I've just shown you and then we said, we are interested in really standardizing our table format. We're working -- a lot of people send tables in Word which is a wonderful wonderful tool. Or Microsoft excel. What you get is like when your kid brings home artwork and you're like, look at this thing. I think it's a horse. It can be in any format, it can look like anything and it can be very, very difficult to try to have some consistency. So we've tried to standardize these table format and to have a database architecture. About the first afternoon of Andy's visit, we realized what he built which goes into dose response was really much too heavy for us. We wanted something really, really lightweight and he said great, there's this programming language that was invented three weeks ago. And I'll use that and make you a little tool. Because we would not want to be out of date, absolutely not. He did that and built this thing called the table builder. This is an open-source tool. If you'd like to use it, we can get you the code and you can go with it. But you can build a lot of the stuff with other existing tools. Like I said take a look at HAWC. It has a lot of these functionalities. One of the things you need to do is include all of your studies. Once we had our studies, we could actually export them via excel and upload them somewhere else.

You this is something you can do. We can manually add in all those government reports and other sources of information. You can see this is really cute when you can click on the little blue things and it takes you right to of mad. You have a close relationship with the data to where it's going. So I'm going to flashing you some of the tables that we created for each of the data streams. It's really about using predefined fields for us. Putting drop downs where you can -- ability to toggle between numbers and plots. When you're thinking about doing a meta-analysis or if you don't do a formal meta-analysis, looking at these risk numbers is really, really, really really helpful. Strengths and limitations. Anything you're going to do in systematic review, you need a place to capture that. And there you go. The animal bioassay data is quite standard. So it's really, really amenable to standardizing these things. You know? What do you need to capture?

All this information I've listed here, if you have a favorite statistical test which you like you -- which we do, we just coded those and they automatically done as soon as the data is entered in there. So reduces the debate of whether that was a significant result or not. We can show you the P value and debate from there. Strengths and limitations, you can export this to different formats. You can use your own database. This is just an example of what we're doing. For genotoxic the, a lot of these studies are quite standardized in terms of what they do. And basically we have a system of drop-down menus depending on what you are choosing.

If it's a study on bacteria, there's only so many tests that can be done. So this reduces the amount of variability. And then you can sort your results. If you want to look at the human results are only my score only chronic studies, whatever you want to do is much, much more flexible and you can capture dose and duration which we do at a high level way. If you're going onto dose response, you probably want to focus on that aspect. So in the last couple minutes I'm going to just briefly show you, why have we been talking about this and where is all this going? The IARC monograph preamble specifies the decision criteria. I'm just showing you what is the criteria for judging animal studies? If you don't have this written down in your protocol, I would strongly recommend it. We use standard terms like sufficient, Limited and adequate. A definition of what those are. 99% of cases are going to be in these categories. If you want to go with the type of evidence and put it in sufficient and it doesn't match the criteria, you better have a strong rationale.

Otherwise we expect it to be, because there was multiple -- positive results or a single unusual result, this creates the consistency I was talking about across the assessment. When you are making comparisons, a lot of people understand if you say there was sufficient evidence, this is what means. As promised, I'm just going to briefly show you how do you integrate and how do you describe how to do this? For us, this is showing for inadequate evidence or higher, evidence in humans on the left, evidence in experiment animals on the top. Now we have our groups. This is how we describe our evidence group. One carcinogenic to humans. Usually sufficient evidence in humans.There's other levels of evidence and one of the main points I want emphasize to you is even if you don't -- have not established a causal relationship, it is possible to make a conclusion such as probably carcinogenic. These are the typical situations where this occurs. But remember that I told you the mechanistic people who have waited through tons and tons of data, they are the smartest people in the world, at least I believe so. How do they play in this game? Essentially, they can apply a get out of jail free card or go directly to jail card or all other kinds of cards they want by moving these overall evaluations up and down. This can occur, you can move something all the way into group one based on a finding of a really strong causal relationship in exposed humans but you can also move things up and down based on evidence that's -- from the mechanistic data set.

This is to keep in mind, described in our preamble, but before this current version we're using from 2006 there were mechanistic upgrades going back into the '90s. Just as an example. Just to summarize what I've talked to you about today, systematic review is fundamental and cancer hazard identification -- we have a preamble. This can be done in a short document. Who, what, how, and where. Evaluations are conducted. We found online tools can aid in identifying and managing voluminous and complex scientific literature and a lining tabular presentations with where you're actually actually going in terms of strength of evidence and overall evaluations. We found this to be really, really helpful and expect to move further in the direction in the future.

So I want to thank all of you for your attention. I want to thank the organizers and also I've been I want to thank Andy Schapiro who I showed you some of the tools he has developed. I want to thank -- Andy was photo shopped into the picture of thevolume 112 working group. As he says, larger than life. It's not so hard to be tall in France but when you have Ivan Rusyn in the picture, it is difficult. Through Photoshop, this can be accomplished. I want to thank the wonderful staff of my colleagues in Lyon, France. So thank you.

[Applause]

Dr. Rusyn

Hopefully while we're getting questions from online, Kate, before we start, we have a few minutes. I wanted to ask a question about where you started putting mechanistic evidence. We will have Dr. Cogliano later today explaining some of the actual case examples, but what do you think are the best steps to avoid perception of bias? You know with mechanistic data you can move things around into hire or lower hazard categories. How do you think some of these systematic tools and defining the protocol and the question and having transparency about what evidence do you have? How does that help to dissuade that potential perception of a bias?

Dr. Guyton

I think that's really a great question. Our process is one of. Review and consensus. Having first of all to avoid bias, I think addressing issues like conflict of interest, we have a very strict policy. I encourage others to follow that. We publicly disclose any kind of funding and associations. I think this is one area

that you can do in terms of financial bias. In terms of other types of bias, it's important to have a balance of views in any group you're getting to evaluate. Any body of evidence. I think that's certainly true. I think having a protocol and also using these 10 key characteristics, they -- this is a really helpful step because you are actually searching for all kinds of things. I'm going to quote Martin Smith who said this really helps you get away from the problem of it's driven by whoever is in the room. Now you can search for all kinds of different evidence and you have to really look through that systematically. Versus you bring in your Jean talks guy and he sees the world through that lens. It's a great lens but maybe you're missing some other aspects of the database. So this is really helped and we've seen how it's changed some of the dynamics of the discussion by having all of that evidence on the table and people not worrying, if it's not this, what might it be? What's the overall strength of evidence? So I think those are some of the directions we see that going.

Alison J Kretser, ILSI North America
Alison J Kretser from ILSI North America. My question is on the databases that were searched from your slide deck. I really enjoyed your presentation. PubMed, did you also search beyond other different databases where journals are not extracted into PubMed?

Dr. Guyton
That is a really great question. We typically do look into other databases as well. So we had -- when I was showing where we actually upload all the references into, that's a place where we can capture them know matter where they came from. HAWC, I believe at the moment, is a strictly -- Andy is going to expand it. That's a really good question. You can certainly do that. And we certainly do. It gets bigger and bigger.

Dr. Rusyn
Just to add to what Kate was saying, you can add anything you found otherwise or someone volunteers something is confidential [Indiscernible] something else the -- all sorts of things in there.
To follow up on my question, really, if you only looked at PubMed, are you open to criticism that you didn't look at all the literature that isn't in PubMed? So that's, how do you address that? Because it is a balance between resources. And so forth. And so I was curious how you looked at that or the selection of databases for journals. The database that has other journals that are not in PubMed. So was there a process?

Dr. Guyton
Absolutely. We basically have a documented workflow we follow in terms of all the databases that we would normally search. And people bring stuff up all the time. As I say, you can do your systematic searches. You're always going to find evidence from other places. For us, the idea is to try to capture all that together in one place and make sure we have it in front of our working group and make sure they've considered and discussed that. So what Ivan is saying, it is absolutely true and you can enter anything you want into HAWC. You can type in any kind of reference you want. It's just that automatic pulling in with the PubMed ID is going to grab it. So that's a very easy system to work with PubMed. Doesn't mean you can't capture all your other literature there. We actually chose not to do that in HAWC. We did that in our other system so we used HAWC for PubMed which is the primary source of our literature. I'll say it that way.

Dr. Rusyn
A couple questions online from Angela Hofstra. How do you ensure that negative studies are captured as well?

Dr. Guyton
So we wouldn't include or exclude if you saw the criteria I showed. This is not based on the study results. It's what the study examines. So in most cases it would be -- things would only be excluded if they weren't about the chemical -- it was a chemical that we were interested in. Or it didn't have an endpoint that's related to cancer and that includes anything in these 10 key characteristics. So we use a really -- our exclusions are quite conservative I would say. Most things stay in.

Dr. Rusyn
Another question from [Indiscernible]. Can you review -- your review approach be used to guide development of in vitro approaches into maximize -- and to maximize the value for risk assessment?

Dr. Guyton
That's a really good question. We do consider in vitro data. And that definitely comes in. Certainly a lot of the literature we retrieved, these were in vitro studies and we did capture them. I think there's another aspect that I touch on very briefly that wasn't the focus today but let's say high throughput screening test, we did pull in those data from those large databases of the US government. And tried to make some progress. How do we examine those and how can we include those in our decision-making? I think it's quite a challenge for us. We really see because we have these working groups that are really diverse, to have that kind of data in front of them, at the same time they're looking at epidemiology and bioassays and traditional toxicology, this is really a great opportunity to try to understand how those data, what kind of gaps they can fill and those kinds of issues.

Dr. Rusyn
This is from Mr. , primarily about IARC monograph -- part of the question is quite relevant. At which point you bring in the documentation issue of dose and concentration.

Dr. Guyton
Right. So it's a hazard identification program. And we are looking from the epidemiology from exposures that can be high like occupational exposures. We really wouldn't make a cutoff in terms of literature based on the dose was too high. So it's a hazard identification program. We do try to capture the does. From the mechanistic side, it's really about providing that perspective. So a lot of times the animal bioassays, there can be a comparison there between the doses that were looked at across different data streams. Let me put it that way. We do capture that information. Dose and duration are important to try to understand whether something -- there is a causal relationship that was developed and to also look at consistency of studies from the epidemiology. What are the different kinds of populations that were exposed and followed and that were studied?

Dr. Rusyn
Thank you. Last question for literature data. For example as ratios, do you typically take the face value from the paper or reanalyzed to confirm the values?

Dr. Guyton
That's a really interesting question. So our working groups will really look at the primary data reported in a study. Sometimes the author may reach a different conclusion than our working groups based on we will try to get that original data, the working groups may try to do their own analysis, using different statistical tests that we recommend in our preamble. They may do meta-analysis where they've gone back and picked out -- pulled out the odds ratios and so forth. So there is an opportunity to do that. We primarily due meta analyses with the epidemiology data, animal bioassays data and mechanistic data is more about statistical analysis to try to determine significant effect and this kind of thing.

Dr. Rusyn
Great questions. We're right on time. We have a 15 minute break. Those of you online, please reconvene at 10:15.

10:00 AM–10:15 AM Break

Dr. Rusyn
Thank you very much, everyone. Please come back and take your seats so we can proceed with the rest of the agenda.Thank you all again for coming back from time. Hopefully those of you who are on the web also had a chance for a little break. Our next speaker is my colleague from Texas A&M University, Dr. Weihsueh Chiu, who joined us recently a couple months ago. Before that, he had an illustrative career with USPP a winning numerous awards and being a card-carrying risk assessor. So he didn't have much of a schooling, formal schooling, but apparently having 2 degrees in physics, one from Harvard and one from Princeton sets you on the right trajectory. All the heavy number stuff for folks with appropriate education, definitely Dr. Chu has that station. He appeared in one of these colloquia before, the first one on partially hydrogenated oils. So he is the return to spend, but we asked him today to really broaden the issues of dose response because that's another very important part of the risk assessment. I'll try to progress from defining the study question and protocol and literature searches now into more of a dose response organization. So Dr. Chiu?

10:15 AM–11:00 AM Harmonizing Dose-Response Assessment for Cancer and Non-cancer Endpoints in Human Health Assessments– Weihsueh Chiu, Texas A&M University, College Station, TX

I'd like to thank all the organizers. It's great to be back here and to delve into the topics of dose response in more detail. Talking about a lot of recent developments in this field. Although I didn't go to Johns Hopkins, my mom did have a postdoc at Johns Hopkins so there is still a connection there.

[Laughter]

Before she worked here at FDA. I have been asked to discuss harmonization of dose response assessment for cancer and noncancer endpoints for human health. I have to declare I have no conflict of interest to declare. And I want to begin with acknowledging not only the organizers for inviting me back. I didn't do too bad the first time. Also the many colleagues who've contributed to the work that I'm going to describe. So I'm going to be focusing on both some fundamental concepts, first making sure we're all on the same page to how we currently develop toxicity values, talk about a new framework developed through the World Health Organization international program on chemical safety, to develop a new harmonized toxicity value that can be applicable to both cancer and noncancer effects. And then to demonstrate some of the software tools that could be used to implement the harmonized dose

response assessment and particularly the wizard tool that's in Excel, and probe Excel tool developed as part of the WHO project. And using case studies of two compounds that are somewhat relevant to FDA.

So as Juleen and Kate both discussed very eloquently, risk assessment is really in the middle between the research and the risk management that helps informed decisions, but really as Juleen mentioned, should be starting at the risk management side. What risk management options there are, what kind of risk assessment we need to inform those decisions and feeding that back to the research as to what kind of data and analyses are needed to move forward again. To ultimate decision-making. So today I'm going to focus on the aspect of dose response assessment. In order to -- of the between hazard ID and dose response can get a little fuzzy. I want to review terminology to show where -- aware that line is going to be. So adverse effect is really part of hazard identification. What kind of health effects are being caused by agents? And then the dose response relationship is how the incidence or severity of that effect changes with exposure.

From these dose response relationships we derive what are called toxicity values, numerical expression of that relationship that is then usable in a risk assessment context. So I'm drawing the line between adverse effect, assuming adverse effects relationships have already been established and talk about how we go to describe the dose response relationships and derive toxicity values. So the first really important concept is that a point of departure -- so we are presented with a set of dose response data that have been identified through looking at studies through systematically rich research, identifying adverse effects, determining causal relationship and also evaluating mechanistic data. So the first step is to derive this point of departure which is intended to be where the data stop and where some additional extrapolations we need to perform for risk assessment beginning. It's traditionally often done with low AO or no AO approach. Or the lowest dose of the study at which and effect has been observed or the lowest dose at which know statistically significant effect has been observed. Now, there's several limitations although -- it's very dependent on dose facing, depending on the quality of the study, the same dose may be low AO in another study.

That brings in potential lack of consistency. So what becomes common practice nowadays is actually to dose response modeling to fit a curve to the dose response data and then to derive what's called a benchmark dose. And that is the benchmark dose, the dose at which a particular benchmark response level is achieved. So as you can see, where is -- we defined a specific level of response called the benchmark response. Given the overall dose response relationship at well -- the benchmark response level. So then the benchmark dose would be the central estimate. What is called the benchmark lower limit and lower confidence limit as a point of departure. So this is intended to be where the observed data ends in terms of dose and where we're starting to do these extrapolations. So the point of departure itself can be used for characterizing risk assuming a simple ratio between exposure, point of departure and the ratio and that is what is termed the margin of exposure. Basically the larger the margin of exposure, the further below the point of departure you are in terms of exposure in the population. So the less risky you presumably are. Of course the next levels of sophistication is to take the point of departure and divide by number of factors that account for uncertainty and variability in the population and inter-species other aspects to derive what's called a reference dose. Also the ADI, the daily intake, another version of the same thing.

The factors -- since the '50s, dividing by a factor of 100, that factor has been slowly disaggregated into smaller and smaller pieces to account for different portions of these extrapolations steps that we need to perform. So first, now at least at the EPA and environmental applications is the dose a metric factor which I just for the fact that -- experiment animals and humans have different respiratory geometry,

different metabolic rates due to body size and there's a partial adjustment for interspecies differences. Then there's a number of other factors, additional interspecies toxin dynamic factors. There's some issue with study design, sub chronic study but you are looking at chronic exposure, you might have another factor as well. In the end, you multiply or divide these factors together to divide -- to derive the reference value. The reference value again, you can compare to exposure to do a risk characterization to hazard quotient. In this case the exposure is on the top. So the lower the hazard quotient, the lower -- the lower the risk, presumably is.

For cancer assessments, we can begin with the dose response data but we have to make a decision based on mechanistic data or mode of action as to whether we think there is a nonlinear dose response relationship in which we can identify a noncancer precursor event. A key event that if you protect against the noncancer event, it's presumed you protect against cancer. If that can be identified and mode of action established, you would use the noncancer approach and derive a reference value. On the other hand if there is evidence that the dose response relationship would be linear at low doses or if there is adequate information to establish a motive action, then we derive what's called a slope factor. And here, the slope factor is the idea that no matter how low you go in dose, there is some finite risk of cancer. The radiation mimetic kind of idea. So depending on the units of dose, it could be a slope factor of unit risk.

Again, by the symmetric adjustment factor, we usually apply it before each of the doses in the study applying it for deriving the point of departure. And then using the same approach of dose response modeling, to derive the benchmark dose. And because there is a linear relationship being presumed, we actually take the point of departure and draw a straight line to the origin controls. The slope then is the slope factor unit risk. So that presumed to be upper bound estimate as to how much additional incidence of cancer there would be as a function of exposure. So then the slope factor unit risk, we're going to multiply by exposure to get the extra risk in the population of the increased incidence in the population of cancer.

So to summarize, in the current approaches for deriving toxicity values, there's the point of departure, reference values, and slope factors. The first two can be used for cancer or noncancer effects. In the case of cancer only, noncancer precursor has been identified and the latter case cancer effects, they would be -- the slope factors are only used for cancer effects. Then these are then -- the risks are characterized when you combine exposure and different measures of -- to risk characterization, quantitative risk metrics essentially. So you can probably see walking through these approaches, there is some partial progress towards harmonizing cancer and noncancer dose response assessment. First of course is the benchmark dose modeling. Second is the application of these dose and metric adjustments, whether from scaling, from using inhalation dosimetry based on respiratory Joan trees or PB came modeling. There's a number of resources that have been developed to facilitate guidance documents to facilitate these adjustments and these approaches. And there was another one of these colloquial -- colloquia that discussed PB came modeling. I'm mostly going to discuss in a little bit more detail the benchmark dose -- and the use of aloe metric scaling.It's a very parallel approach.

In deriving of point of departure. There is difference in what you do with the point of departure after you've derived it. What are the benefits of modeling as compared to using a low AOR know AO? It uses all of the data. The information at higher doses and lower doses actually informs the overall dose response relationship. Is independent of dose basing so that if you have some very wide dose spacing, then you can still interpolate between those doses. It's also transparent as to the magnitude of response. What I mean there is that if these error bars are bigger or smaller, no AO for instance, there's

very large error bars could mean quite a high level of response. If the error bars are very tight, low AO could actually mean a very small level of response. What we care about is actually certain levels of severity or incidence in the population and not just whether it is significant on a pairwise basis. In this way it standardizes looking at cost studies you can look at the same level of response, what is the implication, what is the dose associated with that level of recurrence -- response even if they have very different study design?

Finally it rewards higher quality data so that if you have tighter error bars, if you have a less statistical variation in your model dose response relationship, then we lower confidence limit is going to be closer to the central estimate. Where as if you have noisy data, then the uncertainty will be much wider and so you will end up being more conservative which is appropriate from a public health point of view and also provides an incentive structure for generating better data rather than the no AO and low AO approach, there's anti-incentive that if you generate very noisy data, you could get no AO that goes much higher. There are some challenges though. It is more complex. It can be time-consuming to do benchmark dose modeling although software guidance and training are available to facilitate this and I'll demonstrate that later in the case study. Then not all data are always amenable to benchmark dose modeling. Obviously if you only have a single control and a single dose you're not going to do any benchmark dose modeling since everything would just be a line. In that case you can always fall back on know AO or low AO approach.

So some benchmark dose modeling resources that are freely available, the EPA technical guidance document and benchmark dose software have been developed for some time. But also ICF has developed the free wizard tool which is an Excel spreadsheet in which you enter in the dose response data and it actually runs the ICF BMDS software for you. So you don't have to actually go in and go through all the menus there. And then in addition it keeps a record of all the different choices in terms of modeling options that you've made. There's also an -- for those of you that are more statistically inclined and use R, there is a package developed by the Dutch risk assessment agency, RIVM. Many times used more prominently in Europe but because it is in our package, it's freely available and many can access it. We did publish a paper trying to use the wizard and trying to make this a very standardized approach that one could essentially do in a batch process. We went through several hundred dose response data calculations using standardized decision logic -- at logic as to benchmark response levels and which models are selected and demonstrated proof of principle that that can be done in a not too onerous manner. There are plans to incorporate more of these standardized benchmark dose modeling tools into HAWC.

Allometric scaling may be a little less familiar to a lot of the people looking in the audience today. It's really the idea that on average, across species, toxicity -- species appear to be equally sensitive after you adjust for body size differences. This isn't on a milligram, kilogram per day bodyweight basis but more of a power bodyweight less than 1. So a three quarters power, 0.7 power bodyweight. So here it shows species specific maximum tolerated doses for anti-neoplastic agents. This data has been collected over the years by a number of authors in the literature. By showing them here, these are as a function of bodyweight, what is the NPD? These are mice, hamsters, rats, monkeys, dogs and humans. The error bar here is the confidence interval on the median for that particular species. This line here is consistent with the bodyweight of three-quarter relationship as opposed to a strictly bodyweight the first power relationship. This has been demonstrated in a number of other contexts as well. With basal metabolism, and with more detailed statistical analyses. So now this is actually being applied to both cancer and noncancer assessment. There was a recent guidance document from EPA for noncancer in which part of the animal to human factor is replaced by allometric scaling factor.

Noncancer assessment also address the bigger error bar here which is the variability. Depending on the particular compound, there's going to be chemical specific and potentially species specific talk so kinetic factors that are going to increase the scale about the average allometric relationship. This can be addressed in a more quantitative way through accounting for both sides of the certainty when I talk about the WHO work which is next. So there is this dichotomy in terms of the types of toxicity values and risk assessment characterizations that resulted from cancer and noncancer. So really this WHO project was to address the missing harmonization elements doing dose response assessments. Noncancer assessments are missing first a quantitative risk assessment.

Hazard quotient and margin of exposure may be useful in many contexts, but in other contexts it's really useful to have an actual estimate of instance or severity in a population. In addition, noncancerous assessments are mixing uncertainty and variability and motioned them all together. Whereas risk assessment may in many cases need to address there's going to be susceptible populations separately from any uncertainties associated with those. Cancer assessments are missing any explicit -- accounting for human very bloody. Everyone is assumed to have the same dose response relationship. And it doesn't incorporate some of these other interspecies or other uncertainties.

Both types of assessments, output or missing the point estimate presumed to be conservative but really it's not always completely clear how conservative it is and very likely to be differential conservatives and different assessments, different chemicals that there's levels of conservatism likely to be very different depending on the data set because we're using a standardized tenfold approach or straight-line approach is to derive these toxicity values. So this harmonization project on evaluating and expressing uncertainty along with spreadsheet of a tool we developed to fill some of these gaps. So the idea behind HDMI is that dose response is more than just of these two dimensions.

Basically individuals are going to have very different dose response curves and this is also brought up in science and decisions that we need to make decisions and distinctions between individual level dose response if a -- if you or I are exposed at varying doses versus dose response in population which is the aggregate of everyone's individual dose response. So the idea here is that for a given measure of effect, dose at which it is elicited is going to vary across individuals. To specify what we're talking about in terms of negative effect at individual level, and percentiles and population level as well as dose, we need notations, HDMI to specify -- the HD is the dose. This is the dose level here. At which for instance 10% magnitude of effect was like 10% change in hematocrit or something. Or bodyweight. Experienced by the fifth percentile of the population. So this provides more complete picture of what we mean in terms of dose response. This is a number of benefits in that it provides a quantitative risk assessment in terms of incident.

As opposed to reference dose, a quantity that you can define, different doses we're going to have different incidences in the population. It can address the inter-species and other uncertainties quantitatively. It separates uncertainty from variability. So now we have the variability component in this superscript here and then we can talk about the uncertainty in the whole quantity, where we are in the three-dimensional space, we can define uncertainty in all dimensions. And then by doing a probabilistic analysis, we can combine these uncertainties and the ways we can control the levels of conservatism. If we want a 95th percentile confidence limit, we can always estimate that particular and have more consistency across different [Indiscernible]. To illustrate this, I want to have people consider the definition of current reference dose which is a daily oral exposure in human population including sensitive subgroups likely to be without appreciable risk of deleterious effects during lifetime. All the

HDMI is really doing is making these bolded concepts more explicit and quantitative. So we talked about something that's likely, we can talk about specific confidence or coverage. Including sensitive subgroups, we can talk about a particular percentile of population, 1% or 5% or 0.1% or whatever is the most relevant for the particular risk management context. And then talking about deleterious effects, we can talk about specific sports levels. More than 5% decrease in red blood cell counts during my lifetime. Then we have this derived from the uncertainty distribution in the HDMI and taking this 95% constant -- confidence interval to derive the probabilistic RFP's another reference dose is for specific percentile of population for specific level effect and add a specific degree of confidence.

For cancer assessments, this implies a reorientation of what the risk assessment target is. Currently because everyone is assumed to have the same dose response, in terms of increased risk as a function of dose, there is no distinction between individual risk and population risk. This have been disaggregated in the HDMI approach. And so what we are talking about a function of dose different individuals, different increases in risk of cancer. Conceptually, protect data people with genetic polymorphism have increased relative risk of cancer, just in spontaneously. At the same, there's going to be variability in the population as a function in terms of increase in dose would be elicited -- an increase in particular cancer risk. The overall population incidence can still be calculated by integrating overall people. I'm not going to talk about that here because that gets into additional Monte Carlo simulations.

So really, the conceptual model is actually very similar to what's currently done with reference dose. Making each step more explicit quantitatively. You start with what is the dose that causes a particular magnitude of effect in experiments on animals? He derived according to departure and you have to use a benchmark dose because you need to specify this particular level of effect. But instead of using just the be MDL, you take the whole benchmark dose.

Second, you make these interspecies and study specific adjustments to derive what would cause this effect in the median human. So that includes dose metric adjustment factor taking typical animals and humans as you saw on the previous graph. Looking at the median across species. Then these other factors accounting for toxicokinetics and dynamics, again, using distributions rather than fixed values. Then you need to account for human variability. This human variability factor depends on what your target incidence is. Protecting 95% of population is going to be a different factor than 99% of the population, et cetera. So you need this factor is going to very depending on the risk management goal in terms of level of protection. So then you derive HDMI in the similar way multiplying by or dividing by the factors is appropriate but combining these distributions probably stick leasing end up with an overall distribution for the HDMI. So this can be used in a number of ways. The slope -- a couple times on if you have a probabilistic exposure assessment combined with the dose response assessment, you can say what is the distributional margin exposure in the population? People with high exposure and low susceptibility, high susceptibility and high exposure, you get that whole distribution in the population. You can also derive a probabilistic reference dose as described earlier. Then you can look at different levels of incidents effect in the population.

Essentially derive a population incidence dose response function. This harmonization is for the facilities it by the Excel spreadsheet that we developed. Then you have the dose response data, benchmark dose modeling, the MDS wizard, enter the results into the spreadsheet, it is actually automatically populated with preliminary default distributions that the WHO workgroup derived by doing a review of existing literature, and you can even incorporate modeling results and displays these factors if you want. And then it spits out the results, the probabilistic reference dose and the HDMI doing this calculation in an approximate manner using lognormal rather than full Monte Carlo.

So to close, I'm going to walk through a demonstration of these software tools. The MDS wizard, it -- using the APROBA spreadsheet, using deoxynivalenol. So these compounds, deoxynivalenol and methyleugenol, Don is also known as vomit tocsin -- vomitoxin, highly prevalent in food and in cereals, there are obviously limits as to how much can be present in foods that are derived using risk assessment. So the noncancer effects include decreased body wait which you can guess why the animals were not eating as much. The prenatal development and male fertility, decreased body weight is the one I'm going to use as my case example.Methyleugenol is a natural component of a number of essential oils and a flavoring agent, they combine this with a pesticide so that the insects are attracted to it and killed by the pesticide. It's the group -- we were talking about this in the previous hesitation, possibly carcinogenic with multiple issues and multiple sites. I'm going to leave -- to use liver tumors in rats as the example here.

The first step in calculating the HDMI is to select M and I. For Don, 5% change in body weight and say we want to keep that less than 1% incidence in the population. For methyleugenol, 1% or 10% extra risk just to show the difference in that 1% incident. This is individual risk. We're saying less than 1% of the population is going to have extra risk, individual risk of cancer of 1% or 10%. So the next step is to do the benchmark dose modeling. This the MDS wizard is a tool developed by ICF International with funding by US EPA that streamlines the use of be MDS for modeling dose response. In terms of data entry, modeling, the different model options, running multiple models, simultaneously. Someone else is controlling -- controlling the mouse here –

[Laughter]

So disregard the random fluctuations there.

[Laughter]

The decision logic for selecting models. A similar functionality is being developed to be incorporated intothe -- HAWC. The alternative is to use the R package. This is a screenshot of the be MDS wizard. I know it's difficult to read but the basic point is that in the one section over here, you enter study endpoint information. What is the endpoint, what is the dose unit? Some textual description. You click here, and then it pops up another window in which you enter the dose response data. What is the mean response, standard deviation, number of animals? Then here is all the different models that are in the MDS on the side and you click and select them. The ones you want to run for the particular modeling exercise. I just selected them all. Just click on them so it's not any more burden to select more models. Then for each of these models, you need to select a model option. For the benchmark response level, so for Don, it was a relative change of .05. 5% decrease in body weight. For methyleugenol, BMR was extra risk and 10% or 1% extra risk.

So this is all in a spreadsheet, very civil to enter. Then you just press run. And the results pop out. So in terms of I'm going to skip some of the details of selecting the appropriate model and for this exercise, illustration just say we're going to select the model with the lowest information criterion that balances the model, the model fit with the number of -- the complexity of the model. So for DON, the lowest AIC was linear model. The data don't look too different from linear. The benchmark lower limit and the benchmark upper limit are pretty close together. So really, there's not a lot of dose response uncertainty in this particular data set.

For methyleugenol, there's a little bit more uncertainty about factor three. The lowest AIC model was a logarithmic model. Log logistic model. You can see as well that the amount of uncertainty if you go to a 1% extra risk gets larger which makes sense because you're getting further from the data and so the model uncertainty is expanded as you go closer and closer to the origin.So then you can enter this information into APROBA. APROBA is the Excel-based tool that we've developed as part of the work group that conducts an approximate probabilistic analysis. We showed that the approximate analysis is very close to doing a full Monte Carlo analysis with 10 to the seven samples. So given that there is uncertainty in the distributions et cetera, so this is really -- you have to compare with what we're doing now which is dividing by six factors. This is definitely a step forward from that. We're actually putting together a training course for everyone who might be at euro talks -- EuroTox in September. We're also developing a new version of this that is going to try to incorporate probabilistic exposure assessment as well.

Back to the actual tool, the general layout is there's these different sections in which you enter in different inputs. So the first input section, there's the output section which gives the results and then another important aspect is there's this section in blue which is contributions to uncertainty. Which of the uncertainties is contributing most to the overall uncertainty?So the first set of inputs is related to the study endpoints and the protection goals. So just like in the MDS, what endpoint, what the study design is et cetera, and then there's the risk management goals related to what's at the level of effect, what's the population incidence target and what's the level of competency want? And then you can enter in the be MDS wizard results here as well. So then a second set of inputs is related to these uncertainty distributions.

For interspecies, there's first of this element of scaling component in which there is uncertainty as you can see from the graph, there was still residual uncertainty in the allometric scaling. The scatter, chemical specific toxicokinetics or toxicodynamics might make one species more or less sensitive to that average relationship. And then there's also intra-species which is human variability. So as opposed to a single factor of 10, this actually is based on data that Dale had his had collected over the years on real human variability based on individual data from clinical data et cetera or some of it was epidemiological data as well on how much variation there is in human population. These of course are editable. You can replace with this commission. The outputs here are the HDMI confidence interval, the lower confidence and upper confidence interval, how wide this confidence interval is, how much higher might this HDMI be? The human dose that causes this amount of effect? Adverse effect? And then a probabilistic reference dose which is the lower confidence limit as well as textual interpretation of what this probabilistic reference dose means which in this case, the dose in these units, with 95% confidence at which no more than 1% of the population will have a decreased body weight magnitude of 5%.

In addition, the spreadsheet has a graphical output in which they probabilistic reference dose is this point here, these are essentially population dose response curves but with different levels of confidence. Here is the incidence in the population which for the probabilistic reference dose was set at 1%. And here is the dose associated with that little incident. So here it is very easy to see that for this 1% incident, this is the lower limit, but the confidence is quite -- the uncertainty bounds are quite wide. We all have an intuitive feeling with the uncertainty factors, but this current reference dose gives a one sided confidence bound at which you don't actually know what percent confidence that is. Whereas this gives you a more fuller characterization of where -- what the acceptable dose might be. So for DON, the probabilistic reference dose was .0005. Milligrams per kilogram per day. With uncertainty -- for methyleugenol, the result was .05-milligram per kilogram exposure at which there is 95% confidence

that 1% of the population will only have extra risk of 10% or more with 47 fold uncertainty. If you go down to the 1% extra risk, then the uncertainty goes up as with the ex--- as would be expected.

The last thing I wanted to talk about was but a characterizing these uncertainties. It's good to have an uncertainty range but how might you reduce that? Some of the questions during the break was like, what kind of additional information or studies could you do to reduce this uncertainty? This is where the part about the percent contribution can be really informative. So here is the percent contribution to the overall variance from the point of departure, from the allometric scaling, from interspecies differences and from [Indiscernible]. As you can see there's different options as to reduce these uncertainties that are dose response data, PPK modeling or toxicodynamics studies, intra-species, really the driver of this uncertainty is that we don't know that much about human anybody. And this is beyond the scope, but there are beginning to be population-based experiment experiments systems in which we might be able to interrogate this experimentally and potentially reduce these uncertainties.

So just to summarize, there is some limited harmonization of the current approaches to developing toxicity values in terms of benchmark dose modeling and allometric scaling. WHO, we developed this HDMI harmonized toxicity value which is a redefinition of the RFP to be more quantitatively explicit. There's two Excel-based software tools that can facilitate implementation of these. I walk through a couple case studies of dawn and methyleugenol. I'd be happy to take some questions. Thank you very much.

[Applause]

Dr. Rusyn
Let's hold the questions to the discussion. Just one technical question. You use a some data some of these tools are available with URLs. Maybe this is what this is.

Dr. Chiu
This is what --
The publications with the data for DON and Matt, if you can include those or where the actual raw data came from.

Dr. Rusyn
Maybe one quick question.

Ron Lawrenson, US FDA, College Park, MD
Ron Lawrenson from FDA. I have one quick question. I don't know how long the answer will be, but the question is why do risk assessment methodologies always want to model more variables than the data can handle? As a corollary to this, is that not a form of bias? Or is it just for risk communication?

Dr. Chiu
I guess in terms of dose response modeling, the AIC is supposed to account for how much data -- how much does results modeling can be supported by that data. I think one of the ideas behind moving towards this probabilistic approach is that we want to draw upon what we know from other chemicals, from the universe of chemicals like allometric scaling that information -- from historical studies, we should be able to draw upon our historical knowledge which I think we do already in the heuristic sense.

But more in a quantitative sense drawing upon the information we already know from other chemicals to inform our dose response assessment for the current chemicals.

Dr. Rusyn
Thank you. Our last speaker before our discussion session will be -- come to us via video even though he had the longest distance to travel. Getting outside of the beltline is difficult business in this town. Dr. J. Vincent Cogliano is director of the iris program at US EPA. Integrated risk information system. That is doing most of the risk assessments that people are using for risk management decisions. And Dr. Cogliano has been was EPA a very long time but he also was on detail to WHO where he served as director of the IARC monographs program. Vince, the floor is yours.Vince, can you hear us? We're giving a signal -- we are given a signal to wait a second.

11:00 AM–11:45 AM **The Use of the Mechanistic Evidence in Human Health Assessments–J. Vincent Cogliano (by webinar), US EPA, Crystal City, VA**

Okay. You don't need the telephone anymore, right? Can you hear me?

Dr. Rusyn
Yes.

Dr. Cogliano
Hello? Can people hear me?

Dr. Rusyn
Yes, we can.

Dr. Cogliano
Okay. Thank you very much. Again, sorry for not being very in person. I would have liked to be. It's rather ironic that the closest person is doing it by webinar but we have a science meeting on PCBs that have long been scheduled. I want to thank you very much for inviting me. This is a topic I really enjoy. So let's see. We advance the slides -- so I'm going to talk about using mechanistic data in human risk assessment. There will be examples from both the iris program and from the IARC experience I had. Seems to be a delay in advancing slides.

Okay. No conflicts of interest. Here we go again. Okay. I'm going to talk a little bit about health assessments and systematic review. I know you've had some talks about systematic review. What we ask of mechanistic evidence, several examples of how we can use it not only in classification but other places. Then I'll talk about how we're thinking about organizing mistake -- mechanistic data here in the IRIS program. Systematic review, IRIS assessments are multiple systematic reviews. We ask does the agent we are assessing cause cancer in humans, animals? Does it cause neurotoxicity in humans and animals? This can be formulated into systematic review questions. We're expanding the boundaries to do much more. Analysis of mechanistic evidence and exposure responses relationships.

Now, even though those aren't in the domain of systematic review, we tried to use structured replicable processes -- I'll have to say that about 80% -- that should come up soon, 80% of the issues we deal with and the complexity deals with the mechanistic data and exposure response are raised. When we work

on systematic reviews, it's probably a lot less of the controversy. On the next slide, when it comes up, at IR, mechanistic data are used quite often when human data are not conclusive. -- at IARC. You might have seen this already. If the human data are not sufficient or show lack of carcinogenicity, mechanistic data can affect the classification. And the IARC has many examples of chemicals going up or down, classification based on analysis of mechanistic data for various reason -- reasons. I'll say more about that in the future. EPA also uses mechanistic data. The same is true. When mechanistic data are not conclusive, mechanistic data can move something up to be carcinogenic or can take positive animal results and say something is not likely to be carcinogenic. So there's a wide use of mechanistic data and those tend to be some of the most difficult and interesting questions that we deal with in an IRIS assessment.

On the next slide, IARC asked several questions of the mechanistic evidence that deal with their classification rules in the preamble. So for example, you can upgrade something to carcinogenic if you have strong evidence in exposed humans. Strong mechanistic evidence. So would you say do we have strong evidence of an operating carcinogenic mechanism? Is this evidence in exposed humans, in vitro human cells or in animals? Another question IARC asks is, does the mechanism operate only in animals? Would it be relevant to humans? IARC also classifies agents based on similarity to other agents that have already been classified. And I'll say a little more about that as well. At IRIS we ask similar questions. That's going to come up and we ask, is the motive action sufficiently supported in test animals? Do we have strong evidence in test animals? Is it relevant to humans? Then we go on and ask which populations and lifestages might be particularly susceptible to the mode of action? Because we do dose response analyses and we would like to be able to reflect susceptibility in those dose response analyses when we can.

Mechanistic data gives us a good way to do that because when you understand the mechanism, you can start to say which populations or lifestages might be more susceptible to this particular mechanism.Okay. On the next slide, in IRIS assessments, there are uses of mechanistic data that we talk about. One is to augment the human evidence by establishing the occurrence in humans of precursor events. So human data are very rarely definitive in the types of chemicals we look at. Sometimes we do have some human data that suggests an adverse outcome in humans. But it's not really definitive. It's not compelling data. If we had mechanistic data that showed some of the precursor events also happening in humans, that's corroborative data that can increase the weight of evidence. We also can determine the relevance of animal results to humans. That's probably the most talked about use of mechanistic data at least in government agencies in the US. Where there's a lot of effort to understand the results happening and experimental animals -- and to determine whether or not they are relevant to humans. Then as I said earlier, identifying susceptible populations and lifestages, that feeds into dose response assessment.

In dose response assessment, we use mechanistic data to determine whether we should go for linear or nonlinear extrapolation to lower doses.Mechanistic data can also give us data for dose response modeling. To perhaps extend the dose response curve below where we have typical endpoints and where we have some of the precursors at lower doses. And also mechanistic data are often used when we have relevant potency analyses. For example looking at different dioxin-like conjurers or different pH. Those are relatively derived from mechanistic data. The relative potency of different congeners to produce a particular effect.

On the next slide, I'm going to talk about classification for the next few slides. The strong mechanistic evidence can lead to the highest cancer classification classification both at IARC and at EPA. And you see

there, the language from IARC preamble or from EPA cancer guidelines. EPA cancer guidelines structures it through four points. We still want to have strong human evidence. We want to have extensive animal evidence. We want to have mode of action identifying in animals and then strong evidence that that precursor in the mode of action occurs in humans and our anticipated to progress to tumors.

So the next slide should be an example of how we apply that for benzyl pyrene. It's going to come up very, very soon.I know I hit it. I'll hit it again. Okay. Okay.There's a couple other slides. We can counter positive animal results that if we have positive animal results that you show are not relevant to humans, EPA will classify as not likely to be carcinogenic and IARC will classify something as not classifiable and there you see on the slide, some language from the IARC preamble or EPA cancer guidelines.

In IARC they've gone one step further. They have in their preamble, the strong mechanistic evidence can't by itself identify a possible carcinogen. This was set up by IARC scientific publication in the late 1990s about consistently positive results in several models that establish several stages of the multistage process of carcinogenesis can be considered. In the preamble revision that at IARC went through in 2006 they said an agent may be classified as Parsi -- possibly carcinogenic solely on the evidence from mechanistic and other relevant data. This is opening the door to the future when we'll have fewer bioassays and a lot more mechanistic data that this can actually be used to classify compounds in the absence of human and animal studies. Okay. So here's the case study in IARC on benzyl pyrene. This is an assessment under review. It was reviewed by our scientist advisory board at a meeting last month. So there is extensive demonstration of Carson no tenacity in animal species. Cancer risk in lung and skin and human exposed to many pH mixtures, strong evidence linking BAP to DNA reactive metabolites in events that lead to human development and specific benzopyrene DNA additives have been found in oncogenes and tumor suppression genes in humans exposed to pH mixtures. The PA has -- EPA has proposed this based on those data. And -- why did that happen? Okay. So I'm through with that one.

Let's go to the next slide, slide number 16. And we'll see ethylene oxide. Similar four points. Strong human evidence but not sufficient of breast cancer and lymphoma hematopoietic cancer, expensive -- extensive evidence of genotoxic tea, and strong evidence that these key events occur in humans because we have evidence of damage in exposed humans. This idea of seeing this mechanistic evidence in exposed humans is something that is consistent between the IARC and IRIS guidelines for mechanistic evidence to get to the top category of carcinogenesis.

The next slide -- next three slides, 14 or 15, group one classifications and IARC based strictly on mechanistic evidence. First is about half a dozen with genetic toxicity in exposed humans. One of the characteristics things you see running across all those agents in the middle column, mechanistic rationale, is the genetic toxicity is happening in exposed humans. Each one of those agents we see the genetic damage in exposed humans. The next slide shows based on structural similarity, pentachlorodibenzofuran. Extensive evidence of mechanisms that is identical to what has been extensively studied with TCDD. The last slide is a bunch of miscellaneous agents that were classified in group one. Based on different types of mechanistic or other relevant data but not sufficient evidence in humans.

Now, the most common use of mechanistic data in cancer classifications has generally been to discount animal results. We see positive results in multiple animal studies, sometimes multiple strains, sometimes in two species. What mechanistic evidence tries to do is understand how the animals are getting cancer. And then finding out whether that is similar or different from what would be expected in

humans, and different concluding that the animal tumors are not relevant to humans. There's a trap in that I would like to talk about for a couple minutes. Suppose we have tumors that are operating caused by two separate modes of action, either mode of action one or two could lead to a tumor. Suppose only mode of action one is investigated and it does lead to a tumor.

The framework for evaluating motivation data here generally looks at what's called the modified who criteria. They look at strength and consistency. So do we see whenever we see motive action one happening and the precursor, do we see the tumors? Yes. You probably would see consistency there because mode of action one does lead to tumors. Do you see dose response? Yes. You probably would see that. The more mode of action one is happening. The stronger a tumor response, do you see a temporal relationship? Probably yes you would see that because tumors are going to appear generally after the precursors. Do you have biological plausibility and coherence? You probably do because you developed this motive action and you've shown there's a good correlation. This could fool us into thinking that motive action one is the only cause of the tumors. And if we do have good evidence that mode of action one is not relevant to humans, that shouldn't necessarily make you discount those tumors unless you really know that's the only mechanism. Because mode of action two which either you don't know about or haven't studied as well could be causing those tumors and motive action two could be relevant to humans.

So on the next slide when it comes up, there are several questions that we can ask. Basically will you want to have strong data to establish the mechanism but you'd like to see different experimental systems because you have more of a chance that something else that might be going on is going to be visible there. The most important point I think is the second one. Has each mechanism been challenged experiment of the? That is, if you suppress the key mechanistic processes, do you suppress tumor development? How would you suppress those key processes? You might have it in knockout animals that are not going to produce a particular metabolites or that are not susceptible to a particular mechanistic step. If you reduce the tumor incidence, it shows it's not necessarily the only mechanism going on. If you suppress that mechanistic event and suppress the tumors, then you do have good evidence that perhaps the only mechanism going on. You have to think very hard about alternatives. Could there be multiple mechanisms and could the different mechanisms operate in different species or ranges or in susceptible groups? When we look at mechanistic data, here is -- not a case where we can do systematic review, look at what the overall data tell us. An uneven support for one mechanism versus another might just reflect this disproportionate resources spent on investigating one mechanism or family of mechanisms. And another one might be equally valid but not as well supported in the literature.

So you can't really go with overall prevalence in the scientific literature. You have to think about this in a different way. I'd like to say there's so much more we can do with mechanistic evidence. It's really exciting, more so than just classifying chemicals more than discounting the results of animal studies. The first case study from IARC is about ingested nitrate and nitrite. It shows how you can resolve difficult to interpret results.

The epidemiology study is a very perplexing, the highest consumers of nitrites and nitrates are people who eat lots of fruits and vegetables and actually didn't observe cancer risks in those populations even though they had the highest intake. Nitrate in drinking water wasn't really informative because that's well-regulated and there's really not a lot of large exposure contrast between exposed and nonexposed groups. With nitrites in food we did see consistent positive associations with some stomach cancer and the strongest association when people with high nitrate -- nitrite but low vitamin C intake. But what is

perplexing about this is ingested nitrate is reduced to nitrite by oral bacteria. So there shouldn't be so much difference about whether you are ingesting nitrate or nitrite. You are getting actually eventually nitrite in your body. So this was very perplexing.

Until you go further into the mechanistic data, and determine that ingested nitrate is reduced to nitrite but the nitrite reacts in the stomach to form nitro so compounds and these reactions can be incented by the presence of vitamin C and antioxidants. There's a lot of studies with nitrate or nitrite with and without vitamin C. And you see results that support this hypothesis and it's been explaining the MS -- epidemiological results. If you are getting your nitrate and nitrites with fruits and vegetables you are probably getting antioxidants at the same time and suppressing the reactions. But if you're getting the nitrites and nitrates from processed meats, there's less likelihood you are getting adequate antioxidants at the same time and that's where you might get a stomach cancer. So IARC concluded ingested nitrate or nitrite under conditions that result in endogenous -- the conclusion is not for nitrate or nitrite alone.

The next case study is circadian disruption. It helps you specify the agent that you're evaluating. There were a lot of studies of shift work. The title of the monograph is shift work and that's what we were looking at. But shiftwork is not all the same. There's some people who just alternate between dayshift and easing shift but they sleep the same at night. There's other people that have rotating shifts and they're working at night sometimes, sometimes during the day, sometimes people work night shifts for long periods of time but go back to their daytime schedule on their off days. To join the rest of society. So we did have limited human evidence but there's a lot of really interesting animal studies that show effects from dim light at night, simulated turning the lights on and off and the rooms at different times of the day, removing the pine needle gland that makes melatonin and there were lots of hypotheses but they generally put it to circadian disruption is The Act of -- the agent that's causing these cancers. So we named the agent not shiftwork but shiftwork that involves circadian disruption.

On the next slide, another use is to facilitate preventive actions. This was a really exciting thing that in 2002, IARC evaluated classified plants of the genusaristolochia and aristolochich acid. Out of Belgium, weight loss pills that inadvertently got contaminated with plants of the genus aristolochia. And women who were taking these pills were developing severe renal failure and renal cancer to the point that they were getting preventive removal of their kidneys because this was just such a compelling risk that they had. There are also reports from the Balkans where cornfields are infiltrated with this plant. And IARC was able to come up with the plants of the genus aristolochia workgroup one but it's hard to go examine a pill or some round up leaves and know what it's really coming from. So there was a lot of research afterwards.

Just in about six or seven years, through published studies that showed ADT and T to A conversions in TP 53 genes from aristolochich acid. So that allowed us just a few years later to put aristolochich acid in group one and that gives you a good test for whether there might be contamination. You can look for a particular marker chemical. The next example is on susceptible groups. So IARC looked at alcoholic beverages. You probably mostly know about metabolizing alcohol with Alki -- of -- detoxifies that into acetate. And the variant allele that is highly prevalent in some populations and people who are heterozygotes for that allele have about only 10% of the metabolic rate that other populations would have. And people with that allele were found to have 10 to 10000 times higher risk of esophageal cancer and other upper digestive tract cancers. So not only did IARC classify alcoholic beverages as carcinogenic to humans but they were able to say that acetaldehyde associated with alcoholic beverages is carcinogenic and that gives you now a susceptible population as well based on the mechanistic evidence.

The last example I'd like to give is coming up looking forward, this is a page from our draft benzopyrene assessment where we looked at transcriptome it's. We followed a systematic approach to identifying the studies looking at the raw data assessing the data call it and the genes that are active across multiple studies. So you have this very interesting diagram there. You can find it in the draft benzopyrene assessment on our website. We are going to be doing that in future I risk assessments to see when these data can be useful to help us interpret or modify a classification.
On the next slide, I want to talk a little bit about organizing mechanistic database. When we look at human studies, animal studies and mechanistic studies, even for well studied chemicals, generally it's tens of studies maximum that we find epidemiology studies with particular effect. And the same for animal bioassays. Mechanistic data can have hundreds to thousands of studies and endpoints to be evaluated. So doing a systematic review looking at study quality is almost prohibitive to the number of studies we have. So other approaches need to be taken there. And so we do the same thing as in a systematic review. We try to identify all the pertinent studies. What we do at first is developed an organized inventory of studies to facilitate subsequent analyses. Then we will evaluate key groups of studies using uniform criteria. We want to be able to go through the database efficiently. So there might be dozens of mechanistic endpoints evaluated into all the mechanistic studies as a whole. But the ones we really want to identify and look at more closely are the ones that might be associated with susceptibility, that might be associated with the key events and the motive action. So that gives us a limited number of events.

If we have our inventory, now we can go and get just those studies that are evaluated -- the parameters associated with the motive action or with susceptibility. Now, to do this, we're building on some work that was done at IARC where in 2009, they looked at human carcinogens. Rather than just classify the human carcinogens and say yes, benzene is still carcinogenic, asbestos is still carcinogenic, we decided to do something that would be really useful for the field of risk assessment. We looked at what tumors are caused in humans and animals so we can begin to answer questions about whether there is site concordance. Probably not overall but maybe some subsets of studies or subsets of agents or subsets of tumors might have some site concordance that we could use. But more extensive part was to look at mechanistic events.

Many human carcinogens are classified as 1970s and 1980s before there were a lot of mechanistic data. And in looking at them all now, we were able to identify what were the established and likely mechanisms of events? Subsequent to doing this for all 100 some human carcinogens, IARC convened working group meetings to look at the mechanistic data and try to organize it in some way. What's coming out of that -- it's on the IARC website at the address you see at the bottom was there are 10 key characteristics that explained that were evident in all of the human carcinogens. I don't know that any one of them had all 10, but with that subset of 10 key characteristics, each of the humans are -- human carcinogens operated through one or more of those pathways.

The next slide takes those key characteristics and shows some of the specific types of assays, so for example number three altering DNA repair, you could have alteration and replication, repair base exclusion, double strand break repair, epigenetics, DNA methylation, histone modification, micro RNA, specific events associated with key characteristics. So the organized inventory is around those 10 key characteristics. Each one includes studies that evaluated an array of related endpoints. What you can do with those key characteristics, some of them do lead to others. So the next slide should be a network that shows for example, electro filling or metabolically activated to a genotoxic -- cause genotoxic the.

These are the key characteristics, some of which are related and cluster around going through mutagenesis. And others don't necessarily go through mutagenesis to get to new pleasure. There's a lot of back and forth feedback through these diagrams. Some things -- the dichotomy between mutagenic and non-mutagenic is not all that clear. Some compounds are directly mutagenic. This network can help you understand that. What we can do when we look at carcinogens in the future, is to take the data on a, coal, and where the color sampling is more red, the ones where there's strong evidence of the key event, the key characteristic is operative. The color with orange, the ones that are not -- there's some evidence in the gray, the ones there's no evidence for, this gives you a somewhat limited number but still multiple pathways to get to neoplasia. Sometimes through mutagenesis, sometimes not through eugenicist. This will help us understand for example carcinogenicity. I think the challenge is to try to identify similar pathways for other types of adverse outcomes for example, for reproductive and developmental toxicity. For neurotoxicity. For other kinds of adverse health outcomes, can we identify what the key characteristics might be, construct networks like this and then look at the compounds for which of those key characteristics are active? And then perhaps use this to identify networks that will lead to adverse outcomes and what modes of action should be evaluated as we look at the mechanistic data?

Them in my final slide, just a few observations. Many agents do operate through multiple mechanisms. We found that out at IARC and also other chemicals so we shouldn't stop when one mechanism is described. The dichotomy between genotoxic and non-genotoxic agents is overly simplistic. There's really a lot of feedback loops there. Some mechanisms other than mu coach in the city -- mutagenicity would go right to threshold mechanisms but there are mechanisms particularly hormone disruption or other receptor mechanisms that can be active at lower doses. Then there is some populations highly susceptible. So I want to leave you with the mechanism -- the message that there's a lot more than a discounting of animal studies and it's really quite exciting to think about what we're going to be doing with mechanistic data in the future. I hope I've shown you a few examples that go beyond the traditional classification and that we're going to be doing more and more of that in the future. That's my last slide so I'd like to turn it over to you, Ivan.

[Applause]

Dr. Rusyn
Can you hear me?

Dr. Cogliano
Yes, I can. I'm going to turn my volume up.

Dr. Rusyn
Is a couple questions from the web, briefly [Indiscernible -- echo] we're having a challenge with the feedback loops here. So hopefully we got that resolved. So a couple of questions came in on the web asking that you elaborate a little more on in vitro data and -- you showed the example of benzopyrene in a very complex network diagram. There was a little bit of a growl here in the room because I'm not sure how many people found that to be very informative or easy to communicate. Can you perhaps expand on that a little bit just to give us an idea of how you or others in the agency or outside our thinking about where that data can come into the evaluation process?

Dr. Cogliano
I think that this is something that we're just beginning to think about. Benzopyrene was a chemical where there is a lot of this -- these data. So what we would like to do is include this in an appendix in all of our assessments and ask. Reviewers are asked the public, is there something that we can take out of this? I'd probably -- someday we're going to be able to make a classification of a compound based on observing a pathway through these data. We might make an evaluation based on similar profile between a compound that we're evaluating and another compound that's known to be carcinogenic or have some health outcomes. So for example, some metals that are getting used in semiconductors for nanotechnology, you might get the genomics profile this way and then say, this looks a lot like some other metal that we've studied very well and know what it's adverse outcomes are. So that would give us hypotheses and perhaps even strong enough data to classify. I don't think we're ready for that yet, but I think what we would like to do is present those data so that people can start to think about it and I think that as soon as people are ready to use these and we know enough about how these data will correlate with adverse health outcomes, we will be in the forefront of using them to identify chemicals and adverse outcomes.

Dr. Rusyn
Thank you. And another question from the web was about the different uses of mechanistic data, different levels of Hazzard classification. So for example, in the possible or probable category, the mechanistic data can show that adverse events can occur if the exposure is high enough. But it's quite different, at least the question posits, from Bruce one known carcinogen where the healthy fat should have been observed and established rather than probable or possible. So can you comment on that particular question?

Dr. Cogliano
Yes. I think that's why both IARC and EPA have pretty rigid criteria for using mechanistic data to get into group one. You do want to have extensive animal evidence and you want to understand what's happening in animals and show that those precursors could be happening in humans even if you don't have sufficient human evidence. The slide I showed about IARC classifying something as possibly carcinogenic from mechanistic data alone, at this point in time, IARC I don't think will go beyond that. I don't know that we are really ready for that. I think we really have to understand a bit more about the mechanisms to go beyond that. Now, IARC will go beyond that with structural classes of compounds, for example all the dies metabolites to Benzedrine. They don't have to all be tested. You're not going to get human did on those. But it essentially equivalent to exposure to Benzedrine if you are exposed to a dye metabolite. That was enough for a higher classification by IARC. It's basically you have to understand a lot more to get to a known carcinogen or a probable carcinogen. And we right now have the door open to possible carcinogenicity with mechanistic data alone. I think until we understand this better, I would hope that in a generation we will understand it a lot more and we can make any kind of a classification based on mechanistic data that our understanding will be strong enough to be confident with those kind of predictions but it's not there now.

12:10 pm-1:00 pm Moderated Roundtable Discussion – Ivan Rusyn, moderator

Invite our presenters from this morning and Susie to join us. And we'll have some lingering questions that we have punted from the question and answer but we invite folks to come to the microphone and introduce themselves and ask questions. I would like those of you on the web to do the same. We're monitoring your questions and trying to group them the best we can.Thank you again, all, for great participation. One of the comments that was coming in from the web and something that we already

tried to tackle a little bit is the time and effort question because the traditional way of doing risk assessments in different agencies for different contexts, we pretty much understand what the time and effort is involved. For some of these new challenges and new exciting methodologies, there is less confidence and less understanding. So maybe we can take from Juleen, from you and try to put this into context, so when the clock is ticking versus this is a research problem, and we can do it until we run out of money. So how would you then provide more context in terms of advised folks here in the audience and on the call and what to expect in defining their question and being really systematic or systematic light?

Dr. Lam
Okay. So that's a very loaded, challenging question. For our initial case study, to date we've completed and published two case studies implementing the systematic review approach for environmental health questions. We've completed one other one we are currently working on and currently working on two more. I think the approach that we've taken coming from more of an academic standpoint from UCSF where we have collaborated with both EPA as well as California State EPA in working through these case studies to get that point of view from the agency, where they can say, see the process and say, these are things we can do, cannot do in the given timeframe that we have to make our decision. Working through these case studies has been most informative to see exactly how long each part of the pieces take. And it also depends on the study question at hand which is why we've gone through five case studies now where we're looking at when we broadened health outcomes and exposures, when we are taking a narrow focused study question versus a more broader one, how does that change the time constraints required for actually implementing these reviews? One of the things we've done along the way is not only documenting for transparency purposes, each step of the way but also to be fully informative about how long each of these steps might actually take.

One of the examples is going through the screening of the studies. So if we have 2000 studies that we're screening both title and abstract and full text which might seem a lot to some people and it might seem a not a lot to other people depending on what field you're coming from, some of the software programs we've been using keep track of people's time and I think Kate was talking about how one person can go through 90 studies in an hour. We have software that does keep track of that. How long you are open looking at the title and abstract, how long it takes you to come to a decision. And also for fulltext. How long does it take you to go through the fulltext PDF and come to a conclusion about whether you are including or excluding? Going to that and keeping track of time is one component of it as well as keeping time of how long the other steps take. Helps to identify which parts are time constraints, which parts take a lot of manpower because for most of these steps you are doing it to people independently doing it and then comparing and contrasting and coming to consensus for any disagreements between the reviewers. Identifying which step takes the most a matter of time and where we can draw from that to learn how we can actually apply this to implementation in the real world when we're actually making decisions under a very short timeframe. I think that's where I would have recommendations going through these case studies. But I will turn it over.

Dr. Rusyn
Let me maybe make it a little more specific. I think it's important to point out that this is not from start to finish. So I think the entry point and exit points can be at different levels depending on what the question is. It's one question dealing with thousands of studies in title and abstract and trying to identify relevant literature dealing with inclusion-exclusion criteria. Not necessarily looking at fulltext very detailed into every study. And then there's a different time and effort commitment to actually extract information from the study and make sure you are ready for decision. So it would you agree that is really

something we need to Medicaid better as we talk about systematic review, that there are multiple parts of it and these different parts have different levels of requirements and then you can juggle the effort and personnel and time commitments accordingly to what your actual task is?

Dr. Lam
Absolutely. I think that was the point I was trying to make. I was focusing more on the screening because our software program has explicit documentation of the actual time that was spent on that process. Other parts of the process, we internally try to keep track of. But it's a very challenging thing to do because I think like most people, you're working on data extraction. And then e-mail comes in or you go to a meeting and to be able to say I spent five weeks on this, is not really five weeks full time. Five weeks plus or minus all of the other things you are juggling at the same time. So keeping track of that time and presenting that as part of the case study is something that we're really trying to focus on doing, so people can get an idea of just how long each step, each part of the step actually took us to do.

Dr. Rusyn???
So I think this is a really great question and certainly for us in the monographs program as we organize meetings, really thinking about what is the workload, our working groups are essentially volunteers, not paid an honorarium. We do cover their travel and lodging. Many of them find that it's an honor to participate. We certainly are honored by their participation but at the end of the day, it's a volunteer exercise. So we do try to balance the workload across a working group when we compose them, but I also think about for us, the better that we can prioritize their work, what are the key factors that are going to be in the decision? I think looking at some of the examples that Vince showed of when does the mechanistic data really make an impact? This issue of does a mechanism really operate in humans and do you have that evidence? That is a really key question and that is going to come up every time and we need to be sure that we can organize our literature and make sure our working groups are looking at that particular question. Then there's other similar questions for each line of evidence. I think in part it can be addressed by prioritization, making sure people see the forest for the trees and not just start at A and go through Z and come out dizzy but saying, here's all this literature, let's try to organize it for you. Some of this is going to have a bigger impact than others. That's the way our process is set up and many other decision processes are set up as well.

So very important part is to know when enough is enough. New paragraph are getting published, new date is always coming out. Very important point, also very clear guidance in the silver bullet. That you know when you are done with this particular decision when you are moving on with the actual decision. So for those of us who participate in the IARC process, we know the end of the monograph meeting, the white smoke has to come out of the chimney on that last day. I don't think many other processes are like that. That's why I think IARC is very unique. But Kate, can you comment on again, how in your previous experiences with the EPA, consulting days, where the stopping rules, how they can still be very clear? Because you can go on.

Dr. Guyton
Right. And I think someone made a very good comments during the discussion. I don't recall who it was. It's always a snapshot in time. When we evaluate an agent today, I give the example of parathion. Previously evaluated into group three in 1987. In 2015 is classified into group 2 B. This is because there's no data available and those data were not available at that previous time. So you have to appreciate one of the key strengths of the IARC process is the timeline. It's a year from announcement of the meeting until that decision is made. If you have a very open-ended process, I actually will posit it's impossible to

do a systematic review. Because new papers are constantly coming in. If it takes to -- 10 years, your literature base could be very, very different. And if you can think about the year, a year is really a long time and a lot of studies are going to be published during then. It's very important to be in touch with the research community to try to understand what is going to come out, but that's imperfect. And you do the best you can. We have an open call for data. And people do submit new papers as those, available. Because -- I think managing your timeline and also the expectations, what are you going to build to decide upon? I also would say, don't be afraid of group three. That means it's not classifiable. Approximately half of the agents we have ever classified are in group three. This is because there's a big interest to know, does this age and cause cancer? The answer is we just don't know right now. I think that's an okay answer. Coming to that decision, it can feel frustrating for working group to have looked at these data for a year and then the end of the day, it's not classifiable. But that is a mark you put in the sand and maybe next time you come back and you are going to be able to reach a decision.

Dr. Rusyn
Thank you. Weihsueh Chiu, in terms of the dose response, that's a natural concern for many toxicologists because we are not trained quantitatively. So we are faced with a Linux based system in the original EPA benchmark dose software. So -- or even worse. What can you -- can you comment on so speak, the learning curve and the use of actual comments? Did bring the manuscript which was over 800 done by a student that was learning how to do that and a couple of months -- so can we also demystify some of the dose response assessment in terms of time and effort as well?

Dr. Chiu
Software tools and the training and tutorials have really evolved a lot in the last 10 years.At this point, I think benchmark dose modeling is eminently learnable by someone who's not a trained statistician. I think a lot of it is actually more of a conceptual transition from low AO to know a lot to the concept of benchmark dose. That actually might -- in terms of the harmonization might be a bigger barrier than the actual technical Park. In the same way that in the old days people thought of know AO as the threshold, but really it is not the threshold because it's just based on a pairwise comparison. Statistical variation, et cetera. So going from that type benchmark dose in which we are trying to be more explicit about this is the uncertainty, I think it's very analogous to the transition that hour WHO group was hoping will happen between the RFP which is the safe level, to being more explicit about what exactly do you mean? How safe? How much of the population? What level of confidence? I think really the conceptual transition that is going to be more time-consuming than the technical one.

Dr. Rusyn
Thank you. A related question is something that folks started commenting on when Juleen, you were presenting, which is the study question is probably the most important in terms of the time, the effort, and the literature or evidence that you're looking at. And the HDMI is very similar to that. The finer you can define what exactly you're trying to derive, the better in the long run this is. However, there is a downside I think, to this. If your study question is too narrow, then answering that particular question, how effective that is in the context of a particular exposure, because the counter argument can be mounted that you can always find this acceptable subgroup or disadvantage or socioeconomic status group that is acceptable to this particular exposure. And that may inadvertently drive the decision on that particular exposure beyond the very specific narrow study question. So how do you guard against something very narrow in showing hazard or risk not being misinterpreted in terms of a larger population?

Dr. Lam
That's a great question. Also very important. And I think the way that that can get Incorporated is throughout the process of formulating that study question. So in planning and scoping and problem formulation, you're talking to the people who know the science, who know the decisions who are at hand, who know some of the potential study questions that you might be interested in as a risk assessor or risk manager. Having that conversation between all of the interested stakeholders who might know some of the scientific evidence, who might know some of the quantitative tools for assessing the data, who know some of the decisions to be made, those are the kind of questions that need to be discussed among the stakeholders. So focusing on the most susceptible population is not something you should start out with. You should, as Weihsueh was showing that chart, starting with the risk management options. What are the things that we are considering that are feasible to do? As a risk manager or as a decision-maker? Then working backwards to identify what some of the risk assessment methods are and come to what the actual study question is. That's when, if the susceptible population is something that we truly do need to be concerned about, that's when it would come out of something to target your study question on. But it shouldn't start out as the default. To focus narrowly on that, there has to be this conversation among people who are invested in this process, who can suggest whether or not that's something that should be done in this particular situation or not.

Dr. Rusyn
Thank you. Weihsueh, from the standpoint dose response, in the BMD guidance, there is a discussion of continuous data versus categorical data and what the typical benchmark response should be. But -- typical doesn't apply in every situation. In the IPCS guideline, how much in that particular document, if any time is spent on recommending -- is it 1%? 10%? Three standard deviations? The people are usually -- do not have a preconceived notion as to what that level of adversity should be for them to start doing these analyses. So what would be your best advice in where to start? And the default is the best advice, that would be good to hear as well.

Dr. Chiu
I think this probably will undergo some evolution. I think when for instance in ECA when they were transitioning to benchmark dose, they set the benchmark response levels so you would get a similar answer on average. So a little bit circular in that sense. But I think what our working group was hoping was that as we go into the future, we engage risk managers and decision-makers as two, what level of effect is acceptable in this context? Or what is the cost benefit trade-off between potential effects in the population versus the mitigation options? I think having a more dynamic discussion about what those protection goals out to be, I think is something -- it gets to Juleen's point about getting people at the table, I think is something we should aspire to in the future.

Dr. Rusyn
Thank you. I see there are a couple questions. Please introduce yourselves and if you want to direct a question to the panel, please do so. If you have a question to a particular participant, please state that clearly as well.

Center for Tobacco Products
Center for tobacco products. So obviously in tobacco we deal with a pretty complex product. So a lot of discussion we've seen up here deals with chemicals that are either data rich or we have information on. And so this may sound like two questions but I'm posing it as one. And that is that I'm not hearing a lot of conversation about the complexity of mixtures and trying to deal with the analysis of these mixtures

with regard to how to respond on risk assessment, manage these complex measures. And to make it even more difficult, in modernizing toxicology we're trying to make a move away from minimizing and reducing animal testing. And in making that move, we're losing a lot of traditional data we rely on to do these assessments. From benchmark dose to even formulating the problem. Which we're making decisions on computational software QSR. So my core question is basically, we're seeing individual chemicals. How are we tackling these complex mixtures to try and arrive at decision and how do we modernize that with reduction in animal studies?

Dr. Rusyn
I'd like to ask Dr. Guyton to comment because IARC doesn't only consider individual chemicals.

Dr. Guyton
I think this is a great question. And Ivan is correct. I think part of the reason that IARC does deal with complex mixtures also, occupation for example, occupation as a painter. That is a classification. And that may be atypical, but you have to understand the epidemiology data, maybe that's exactly what the endpoint was in those studies. So when you have an expert working group, they may say we can classify tobacco smoking. Now, a separate classification might be something in the tobacco, tobacco carcinogen. XYZ. So these are different things that can be evaluated and have been evaluated. So that's what I would say to the mixtures issue.

Dr. Rusyn
Let's stay on this for just a little bit. Because there were a couple of questions on the web as well. Staying on the hazard part, I think we agree that on the hazard, it is not as complex or insurmountable to think about a complex exposure because frequently there would be studies on complex exposures, epidemiology or animal studies. But then moving that into the dose response and actually doing the quantitative part of risk assessment, so Weihsueh, I know this has been a challenge for you in your previous slides. And there are no easy answers but would you like to comment?

Dr. Chiu
Well, it remains a big challenge. Similar to assessment has sort of -- there's been probably not nearly the amount of progress people wanted or anticipated in the last 10 years. So one thing in moving to in vitro is it becomes more possible [Indiscernible] et cetera to actually test mixtures and subsequent dilution and things like that. And then in terms of moving the dose response, the HDMI concept that I illustrated is actually very similar to what Richard published in the BPD concept in terms of biological pathway activating dose in terms of using a probabilistic approach to look at the dose response and extrapolating it to what using human variability extracting to what might happen, so I think there's potential for path forward, to be explored there.

Dr. Rusyn
I agree. I think that specific issue of how to deal with complex exposures and mixtures is actually something that is driving a lot of the in vitro approaches. And how to translate these in vitro approaches into actual decisions is the process that will take a couple of years at least. But at least that's the data that we can collect even if in dose response and population ability, different types of model systems, from different organs we can do these experiments now. We just need the time and the feedback from how best to let go of human epidemiology and animal studies and start making some decisions. As important a question as decision context, defining what is the context for that decision, whether it is an ozone level integrated science assessment or something for that particular Superfund site. We really

need a decision and it doesn't have to be very complicated. There are examples of that in their efforts to that effect. So next question?

Christy Jacobs Office of Food Additive Safety
Christy Jacobs, office of food additive safety here at six same. Thanks, everybody for a really interesting morning. I question relates to some very impactful papers that have been published recently talking about the overwhelming number of papers in the peer-reviewed literature's that are not reproducible. And I'm wondering whether or not the panel things that a system that we have in place especially looking at systematic reviews pulling in all the literature we can find or using mechanistic information to inform our assessment, do these processes already deal with this problem or does this really represent a new uncertainty in our risk assessment that we have to start considering? What are your thoughts around how we deal with potentially non-reproducible data in the literature?

Dr. Rusyn
Dr. Guyton, can you comment on this -- on the synthesis of information and then Juleen I would like to ask you how you deal with that in your case studies.

Dr. Guyton
So I think this is a great question and it's certainly an issue that I think every scientist really thinks about. He published a study, it's great, you're the first one out there and then that results in an isolated event. I think in our process of data synthesis, we would look at that in terms of particularly for the mechanistic data, any time was the mechanistic data would actually come to play, they would really need to be a coherent -- this study would have been replicated, they would be -- Vince alluded to findings on perhaps multiple types of test systems. These really shouldn't be based on isolated events. So I think personally, science is somewhat self-correcting. So I'm not sure this is as big a problem as some people make of it. We certainly look at agents that are very few studies, but I would say many of them have many, many studies. And in this case, you're certainly going to have your outlier effect, that you could perhaps do a whole meta-analysis to show everything coming out of this group goes one direction and everything coming out of this group goes another direction. And maybe the truth is in the middle. And these are issues that can be debated.

Dr. Lam
That's a great question. And I think in the application of our systematic review procedures, to environmental health literature in particular, this is addressed in a few different ways. First we look at what we refer to as risk of bias. Looking at internal validity. We evaluate that on an individual study scale. So we look at each study and we have anywhere from eight to nine domains depending on if it's handled -- animal or human looking at the study's internal validity in terms of its quality in how it's addressing the study question that was posed by the study investigators. So we look at for epidemiological studies, obviously there's a lot of methodological differences in how they measure exposure. How they measure outcomes. And we look at the validity of those measures that they've used. Have they've been validated or used in other studies? What is the evidence base for using these particular methods? We do that on an individual level scale. Than what Kate was talking about at IARC, we do similar things where we integrate the evidence and look at the quality and strength of the overall body of evidence and do this separately for humans versus animal and then we haven't done this quite yet but NTP is starting to look at this in vitro evidence base as well and evaluating that as a separate body of evidence. So this is where we combine all of the studies from the same animal human body of evidence and we look at it collectively.

So this is kind of where Kate is talking about the self correction where if you one negative study in an overwhelmingly number of positive studies, if there are no internal validity issues with the overwhelmingly positive studies or vice versa, this is where you would self correct that and see those outlying pieces of evidence. We don't include or exclude or upgrade or downgrade based on whether they have positive findings or negative findings. It's all based on the internal validity of the study design, how they implemented their study and then looking at a comparison of the collection of the body of evidence. Obviously when you have a lot of studies, that's something that's going to be easier to plot. If you're working with a smaller number of studies, might not catch that but I don't know if there's a lot of ways you could actually do that. It really is waiting for more evidence to come out and then one last thing I will highlight that we do do in the navigation guide systematic review approach is look at documenting financial conflicts of interest within the study. This is something that's been discussed a lot in Cochrane and grade and some of the other systematic review methods for the clinical field. And clinical literature. So this is kind of a controversial topic that's just starting to develop and figure out the best way to capture this element. It is something we're documenting within our review approaches.

Dr. Rusyn
Thank you. Just to make a point of this is not necessarily would qualify as a study. Just an important point you'd like to document. In case they are inconsistencies, that may be one factor you may take under advisement one way or the other. This brings accurate -- one of the points discussed earlier is having the rules before the evaluation. So you're not changing your rules or not changing your rules without documenting those rules while you're doing the evaluation. Not to introduce the bias in the bias evaluations of speak, but there's also another point which is what is some of the advice you can give in terms of negative studies? Usually don't get published? I think personally this is less of a problem with epidemiology in animal studies because this is a conference of study and you asking a question about possible tumor sites obviously you are documenting with which organs you have looked at or which tissues you have looked at or found no evidence. It's less of a clear-cut issues for the mechanistic and other evidence because if you did a study and found nothing, not always you can publish a study where we had a hypothesis. It was wrong, but let's show the data. So what are some of the thoughts that Juleen, Kate, Weihsueh that you have in terms of negative studies that don't get published? As Dr. Cogliano mentioned, lack of evidence may be lack of resources or some other reason for it.

Dr. Lam
This is something that we tried to address in -- in our case studies specifically. So we do try to identify as much gray literature as possible. Meaning unpublished -- full peer-reviewed studies. The reason we do this is I think it is still an issue in epidemiological or animal studies where they might have started out with one study question looking at a particular health outcome or exposure and they are not finding anything statistically significant. You might shift years and look at something else that you've measured in that population. If you do find something statistically significant, report and publish on that. I think it's a report for human and animal studies. One of the ways we tried to address that is searching through gray literature databases. Set up from after species, as well as conference abstracts. That's one of the best ways -- where you will see somebody reported that they're going to start the study or have preliminary results from the study and you realize these people never published anything. This is something we've been trying to capture. It's indicative of publication bias. Something that is important to address because you have negative studies haven't captured. Obviously this does take some time. We've been able to do that in our academic setting contacting study authors and trying to get -- identify what the reasons are, but it's been a challenging thing either for us to do. Because we do have limited resources available for these case studies. So it's challenging. And I think it hasn't really come out quite

yet because it wasn't really an issue in our first two case studies but for some of the case studies we're working on now, trying to find out whether or not it's really truly makes a difference. Also one last thing I want to mention, NTP also does explicitly state in their handbook as well as protocols, that they will attempt to obtain raw data and analyze it themselves or have it peer-reviewed independently. Something that takes a lot of time with this gray literature database but something that can be done. And for a government body, this is what NTP is proposing to do. So it can be done.

Dr. Guyton
Those are all really good points. It's a difficult question. To address. I think we had some recent experience that I mentioned with our last two meetings for volume 112 and volume 113 where we actually went into the Tox21 database and pulled out all of the results there. And it's a really interesting database because every chemical is screened through the same battery. So I think anytime you have a battery of assays, this can be very valuable. Because you can begin to say, where is something actually negative going to be captured? This is one of the areas where it might be. Our initial experience is we actually took all the battery of assays and mapped our 10 key characteristics to say, where is the data actually going to be informative? Not actually going to be informative on genotypes is it the. A separate battery outside of this domain but there is this area of receptor-mediated assays where this particular battery is quite strong. This is an area where perhaps there haven't been literature studies or academic studies looking in the same systematic way across. So I think we're going to begin to identify opportunities where these other databases can be important for filling gaps and for identifying where you may have negative studies are positive studies, or you may have a blank space and this is one area where I think I can anticipate this could fill in in the future.

Dr. Rusyn
Thank you. We'll take another question.

Dr. Chiu, one of the struggles we have when we're doing risk management is trying to figure out the difference between the risk and uncertainty in order to make our decision. And the scheme that you propose is -- seem to abstract unnecessarily. For example the benchmark dose and the -- people have argued circuits because of the way we design experiments. And so particularly when you're doing a cancer risk assessment and we are projecting several lines below whatever your point of departure is, the difference between the maximum tolerated dose, the LOEL and benchmark dose is tiny compared to the other uncertainty in your analysis. The advantage of the benchmark dose is that it's giving you some uncertainty measurements because your model is telling you something about uncertainty and that's the advance over using the MTD or the LOEL. But then by taking the lower confidence limit of the benchmark dose, you are giving up the entire advantage you spent the effort to generate. Similarly if you're going to hell metrics scaling, you have now reduced the uncertainty in your projection from human to rent. Your uncertainty is shrinking. And your central estimate is either going up or down depending on your allometric scale. When you are looking atmetholugenol, the difference between high-dose and low still -- logos, helps you reduce the uncertainty. So what would be much more useful for risk management tool would be a much clearer articulation of where -- where is the central estimate of the cancer risk? And what are the factors that contribute to that uncertainty and display that around that central estimate, rather than doing that as part of the risk assessment. Because those are really risk management tools. Which areas of uncertainty to use and how to wait them in the overall risk management decision. That's what makes risk management so difficult and so all of those things need to be displayed in the risk assessment rather than simply a symbol number that says, with 95% confidence, here is the level that will protect 95% of the population.

Dr. Chiu
I would clearly agree. Preference in the future would be to show the entire distribution both in terms of population variability as well as the distribution uncertainty. Then that distribution can be disaggregated into what are the major contributors to that uncertainty?We thought it would be an easier transition from the RFT to a lower confidence bound on the this tradition in terms more screening level decisions. Certainly for more sophisticated risk management decisions you would want the entire dose response curve as well as full distribution in order to figure out how you want to balance those various assays.

Dr. Rusyn
One other point that came across some of the comments on the web was doing this systematic review when trying to define the question the best we can and the most appropriate dose response assessment. The issue of adversity versus adaptive responses is something that frequently comes up. And this is less again important for epidemiological studies or animal studies where the end result is disease. For the mechanistic evidence, some of the other information that goes into these decisions, what are some of the tools or approaches to guard against mixing the two? Are there any thoughts that the panel would like to offer on that?

Unidentified
The adverse events versus adaptive?

Dr. Guyton
From the WHO work, the idea is that your.
-- can specify what are your critical magnitude of effects? The critical size of effects you care about? For DON, 5% change in body weight. Bodyweight in that context is more of a surrogate for nausea or whatever else is being caused. So maybe that's not the best example. You would probably want to get down to more lower level of biological organization as to what type of effect you're talking about.
I think the other point that comes to mind, a lot of the in vitro data we are not looking at particular adversity, the pathway perturbation, and at least some of the information that is available right now from the study at the Hammer Institute, they did try to compare fairly short-term toxicology gnomic -- toxicogenomic, informs us that we can't start making some decisions being quantitative without fully understanding whether we're dealing with complete adversity or just response that may or may not lead to adversity.

I think another approach that we had as Vince described of looking at the known human carcinogens during the volume 100 review, and then you are in a domain of these are agents that we know cause cancer in humans. And so the idea of trying to figure out are there key key -- key characteristics that these agents have that would help you classify agents in the future? And I think this is the idea that we're beginning to explore with these key characteristics to try to say, you don't really need to prove that immunosuppression, that you have to fill in all these gaps and that what is adverse and how much and so forth? We know that we have immunosuppressive agents that do cause cancer and therefore there is a much more direct link by coming to this tendency characteristics mindset. So I think that, as we gain more experience with this framework, for organizing the information, and analyzing it, hopefully we can make some good progress in that regard.
My other comment is that adaptive adverse size of effect, there is a hazard or causality part of it, which is what kinds of early events lead to potentially later events that are clearly adverse? Then there's the quantitative dose response part as to what level of perturbation is associated with what level of

increased risk of some event? So the example that comes to mind is long to DC or increase in QT which small increases in QT lead to small increases in the risk of myocardial infarction. So the causal pathway, there's a clear causal pathway between small effect that might not be adverse in itself, but leads to an increased risk of adverse event later. And then the task is how to quantify that relationship for risk management.

Last question before close.

CFSAN
I've just passed them from the Office of Food Additive Safety here at CFSAN. This is for Kate, following on your previous answer. I believe that I read somewhere that regarding thresholds and carcinogens labs carcinogens acting through non-mutagenic mode of action indeed have thresholds. And carcinogens acting through mutagenic mode of action don't. Did IARC look at this group of 10 characteristics in the context of whether or not we thresholds are involved in specific ones or groups of them?

Dr. Guyton
Thanks for that question. One of the great -- it's a hazard identification program. We don't really have to get into that threshold, non-threshold but it's still important because it can play into hazard identification. And I think there is a long-standing belief that mutagens do not have a threshold in -- and other carcinogens do. I think the hormonal agent is one that really -- Vince pointed this out where you can have receptor-mediated effects that actually can occur at very, very low exposures. We have certainly seen that for some of the group one known human carcinogens. These are very low exposure events. They can actually be isolated events that maybe occurred during a specific development window or other sensitive time period of development. And these are different challenges that -- in terms of interpreting that hazard data to reach your conclusion. So Vince actually made the opposite point that this threshold non-threshold dichotomy is weakened because you can have immunosuppression, receptor-mediated, some of these are very, very complex, not well understood in terms of dose with respect to later outcome and whether there is a threshold that does operate. So I think that's an area for a lot further discussion and for the development. And it does play into the hazard. It's more for leaving it to Weihsueh.

CFSAN
Thanks for the explanation.

Dr. Rusyn
That brings the end of our roundtable. I would like to ask you all to give another round of applause to our speakers for their time.
[Applause]
And I would like to welcome Dr. Goering to say a few words to close the symposium.

Dr. Goering
On behalf of SOT again and on behalf of FDA, I think this morning has been a big success for a number of reasons. I was talking to my FDA colleagues at halftime and some of them remarked to me how even the first two talks this morning were valuable to them. Referred presentations today about how the science and risk assessment has made great strides just in the last few years. I think I'm safe to say that with those strides some new questions. We've heard about some of our questions from the audience today. And I think we could expand the lessons of Dr. Lehman to three lessons. And maybe push this out to

each 110 years. We're now at 30 years into -- instead of two lessons in 20 years. I think there is plenty of challenges that lie ahead that we can be confident that we are making significant advances in risk assessment with the goal of protecting public health and promoting public health. I want to thank the speakers again for taking the time out from their busy schedules, coming today. I want to thank the key organizers, Ivan and Susie, for this specific session today. I also want to thank our staff from SOT headquarters. Betty sitting in the back. And Amy and Rachel from her staff, they did a lot of background work that helped pull this together. So I want to acknowledge them. I also want to thank the attendees. We had 80 or so here at our peak today. We had upwards of 300 online. Thanks to the technology, we had upwards of 100 individuals internationally representing 21 different countries that were plugged into this session today. So that's spreading the word out as far as -- as well as we can. So again thanks to all. And thanks for a successful colloquium today. We hope to pursue this next year. So look forward to announcements about this continue to successful collaboration. Thanks.

[Applause]

[Event Concluded]