FOOD AND DRUG ADMINISTRATION

CENTER FOR FOOD SAFETY AND APPLIED NUTRITION

FOOD ADVISORY COMMITTEE

Date: March 30, 2011

Time: 8:30 a.m. – 5:30 p.m.

Location: 10903 New Hampshire Avenue

Silver Spring, Maryland

DR. ACUFF: Good morning. I'd like to welcome everybody. Thank you for attending. I'm Dr. Gary Acuff, acting chair of this committee, and I'd like to officially call this meeting of the FDA Food Advisory Committee to order.

As an introduction, I'm the director for the Center for Food Safety at Texas A&M University, and a professor of food microbiology. I have expertise in microbiological safety and quality of food with emphasis on foods of animal origin.

At this meeting, the committee will consider available relevant data on the possible association between consumption of certified color additives in food and hyperactivity in children, and advise FDA as to what action, if any, is warranted to ensure consumer safety.

Before we begin, I'd like to introduce our designated federal officer, Carolyn Jeletic, and

allow her to provide some introductory comments.

MS. JELETIC: Hi. Good morning. The Food & Drug Administration is convening today's meeting of the Food Advisory Committee under the authority of the Federal Advisory Committee Act of 1972. With the exception of the industry representatives, all members of the committee are subject to federal conflict of interest laws and regulations.

The following information on the status of this committee's compliance with federal ethics and conflict of interest laws, covered by, but not limited to those found in 18 U.S.C., Section 208 and Section 712 of the Food, Drug, and Cosmetic Act, are being provided to participants in today's meeting and to the public.

FDA has determined that the members of this committee are in compliance with the federal ethics and conflict of interest laws. Under 18 U.S.C., Section 208, Congress has authorized FDA to grant waivers to special government employees who have financial conflicts when it is determined that the agency's need for a particular individual's services

outweighs his or her potential financial conflict of interest.

Under Section 712 of the Food, Drug, and Cosmetic Act, Congress has authorized FDA to grant waivers to special government employees and regular government employees with potential financial conflicts, when necessary, to afford the committee essential expertise.

Related to the discussions of today's meetings, members who are special government employees have been screened for potential financial conflicts of interest of their own, as well as those imputed to them, including those of their spouses or minor children, and, for the purposes of 18 U.S.C. Section 208, their employers.

These interests may include investments, consulting, expert witness testimony, contracts, grants, teaching, speaking, writing, patents and royalties, and primary employment.

Based on the agenda of today's meeting and the financial interest reported by committee members, no conflict of interest waivers have been issued, in

accordance with 18 U.S.C. Section 208 and 712 of the Food, Drug, and Cosmetic Act. Dr. Philip Nelson, our chair, has recused himself because of a conflict of interest.

Dr. Gary Acuff, to my right, will act as chairman today. Drs. Edward Kennelly's and Philip Msall's [sic] schedules did not allow them to participate in this meeting. Dr. Xavier Castellanos and Dr. Charles Voorhees have been appointed as temporary voting members. The appointment was authorized by Mr. Michael Landa, acting director of Food Safety and Applied Nutrition in September 2010.

Dr. Sean Taylor, a guest speaker, has reported the possibility of a potential conflict of financial or professional interest. Because his service is considered essential, an acknowledgement of the existence of such interest is being announced at this meeting and made a matter of public record.

We would like to remind members that if the discussion involves any other products or firms not already on the agenda, where an FDA participant has a personal or imputed financial interest, the

participants need to exclude themselves from such involvement, and their exclusion will be noted for the record. FDA encourages all other participants to advise the committee of any financial relationships that they may have with any entity at issue.

Before I turn the meeting back to Dr. Acuff, I'd like to make a few general announcements. Transcripts of today's meeting will be available from Capital Reporting and will be posted on FDA's advisory committee page when available.

I would like to remind everyone that members of the public and the press are not permitted in the committee area. That is in front of these ropes. The press contact for today's meeting is Doug Karas.

Would you please stand? Thank you.

I request that reporters please wait to speak to FDA officials until the committee meeting has concluded. Finally, if you haven't remembered to turn off your phone, please do so now. Thank you.

DR. ACUFF: Now, I'd like to ask our distinguished panel members seated at the table to provide a brief introduction. Please state your

name, area of expertise, position, and affiliation. And we'll start with Dr. Winter and come around this way.

DR. WINTER: Good morning. My name is Carl Winter. I'm an extension food toxicologist at the University of California at Davis. I do a lot of work with food contaminants, food safety, risk assessment.

MR. WALDROP: My name is Chris Waldrop. I'm the director of food policy at Consumer Federation of America.

DR. VOORHEES: My name is Dr. Charles Voorhees. I'm professor of pediatrics in the Division of Neurology at Cincinnati Children's Hospital. My area is neuroscience.

DR. VUGIA: I'm Duc Vugia. I'm a medical epidemiologist, and I'm a clinical professor in the Department of Epidemiology and Biostatistics in the UC-San Francisco School of Medicine. I'm also chief of infectious diseases at the California Department of Public Health. However, I am not here to represent the California Department of Public Health.

MS. MENKE-SCHAENZER: Good morning. I'm Joan Menke-Schaenzer. I am chief global quality officer at ConAgra Foods. I'm one of the two industry representatives.

MS. LEFFERTS: Good morning. My name is Lisa Lefferts. I'm a consumer representative. And I am senior editor at Environmental Health Sciences and a consultant for a number of consumer and environmental groups.

DR. JONES: Morning. I'm Tim Jones, the state epidemiologist at the Tennessee Department of Health and do substantial work in foodborne illness.

DR. GRAY: My name is George Gray. I am professor of environmental and occupational health and director of the Center for Risk Science and Public Health at the George Washington University School of Public Health and Health Services. My areas of expertise are toxicology and risk assessment.

DR. GERBA: Yes. I'm Chuck Gerba with the University of Arizona. I'm an environmental microbiologist, and I have expertise in food

microbiology and risk assessment.

DR. FREELAND-GRAVES: I'm Jeanne Freeland-Graves. I'm the Bess Heflin centennial professor at the University of Texas in Austin in the Department of Nutritional Sciences, and I'm an expert in nutrition.

DR. FERNANDEZ: Hi. I'm Maria Luz Fernandez. I am a professor in the Department of Nutritional Sciences at the University of Connecticut, and my expertise is nutrition.

DR. FENNER-CRISP: My name is Penny Fenner-Crisp. I'm a twice-retired toxicologist and risk assessor. I spent 22 years at the Environmental Protection Agency, primarily in the pesticide program, and a short stint over at the ILSI Risk Science Institute.

DR. CASTELLANOS: I'm Francisco Xavier Castellanos. I'm a pediatrician, child psychiatrist, and professor at child and adolescent psychiatry, radiology, physiology, and neurosciences, at the NYU Medical School, and director of child and adolescent psychiatry research at the Nathan Kline Institute.

My areas of expertise are neuroscience of ADHD and neuroimaging in particular.

DR. BURKS: I'm Wesley Burks. I'm a pediatric allergist and immunologist at Duke University in Durham, North Carolina, and my area of research is in food allergy.

DR. BLAKISTONE: I'm Barbara Blakistone, director of scientific affairs for the National Fisheries Institute in McLean, Virginia. My expertise is food science, and I'm also an adjunct research professor of food science at the Virginia Tech University.

DR. ACUFF: Thank you very much. We appreciate your time in serving on the committee today.

As you can see in the agenda, we have several speakers planning to address the committee today. I will try to keep us on time as much as possible, but we do want to allow time for questions as needed. To help us manage our time, please hold your questions until the conclusion of presentations, and then we'll address them then.

At this time, we will hear the FDA presentation regarding the charge in question to the committee from Mr. Don Kraemer. Okay, just welcoming remarks from Dr. Don Kraemer. Mr. Michael Landa, was originally on our agenda. He has some family issues to deal with this morning and could not attend. So our deputy director will be here today to do the address.

Mr. Kraemer?

DR. KRAEMER: Thank you, Dr. Acuff.

Good morning. I'd like to welcome the members of the committee, the guest speakers, members of the public, and my fellow FDA employees. As introduced, I'm Don Kraemer. I'm the acting deputy director of FDA's Center for Food Safety and Applied Nutrition. As Dr. Acuff just mentioned, Mike Landa, who is the Center director at CFSAN, had planned to be here this morning, but unfortunately, due to a family medical emergency, he was not able to come, which is disappointing.

I would like to give special thanks to our advisory committee members, who have taken time away

from their jobs in academia, private medical practice, state departments of health, and other areas, as they mentioned here just a minute ago, to thank them for being here.

We're very appreciative of your willingness to take time from your busy schedules to help us consider whether available data demonstrate a link between children's consumption of color additives in food and adverse behavioral effects. I also want to thank our invited speakers and other stakeholders who are here today, some of whom have traveled from as far away as Europe. I know you have taken -- you have a very full agenda, so I'm only planning to have brief remarks here today.

FDA prides itself on bringing the best science to the many complicated issues and areas that we face in the areas of food, cosmetics, and nutrition. An integral part of being able to achieve that goal of sound science is seeking opinions, advice, and expertise from the outside -- that is, outside the agency -- and we're doing that here today.

Your contributions over the next two days enable FDA to enhance the expertise that we already have available within the agency, and it also provides a unique forum to air the breadth of different ideas and views in an open, transparent manner. I'd like to thank you again for participating, and I know this does take a lot of time from your otherwise busy schedules, so welcome and thank you.

DR. ACUFF: Thank you, Mr. Kraemer.

So at this time, we'd like to hear the FDA presentation regarding the charge in question to the committee from Dr. Mitchell Cheeseman, the acting director of the Office of Food Additive Safety, the Center for Food Safety and Applied Nutrition.

DR. CHEESEMAN: Thank you, Dr. Acuff. As you mentioned, my name is Mitchell Cheeseman. I'm the acting director of FDA's office of Food Additive Safety. The Office of Food Additive Safety is the Office within FDA that's primarily charged with the review and oversight of the safety of food and color additives.

This morning, I'm going to be giving you an overview of the statutory basis and the process that FDA uses to review the safety of color additives. I'll also provide some background information regarding the regulation of certified color additives, which are the subject of this meeting. Finally, I'll provide you with your formal charge and questions.

FDA's regulation of color additives really began with the 1938 Act, the Federal Food, Drug, and Cosmetic Act. The 1938 Act mandated certification of so-called coal tar dyes. This term, "coal tar" is a term of art that refers to the historic source material used to produce these colors, which are today typically synthesized from petroleum products.

Certification refers to the process of analyzing representative samples of manufactured batches of colors to verify they meet the specifications established for that color by FDA. The 1938 Act prohibited the use of uncertified colors in food, drugs, and cosmetics, with the exception of hair dyes. Finally, the 1938 Act required that food

containing color additives be labeled to declare that they contain artificial coloring.

FDA's regulation of color additives was further enhanced by the 1960 Color Additives Amendment, which required pre-market approval by FDA for all color additives added to food, drugs, cosmetics, and other FDA-regulated products. The 1960 amendment defined color additive, established a provisional list of color additives in use at that time, and established a petition process for listing authorized uses and establishing permanent listings for those provisionally listed colors.

The 1960 amendment also established, as I said, a color additive definition in Section 201(t) of the Act. It's a somewhat broad definition, which defines a color additive as a material which is a dye, pigment, or other substance made by a process of synthesis, or extracted, isolated, or otherwise derived from a vegetable, animal, mineral or other source, and when added or applied to food, drug, or cosmetic, or to the human body, or any part thereof, is capable, alone or through the reaction with other

substance, of imparting color thereto. And color, for purposes of the definition, includes black, white, and shades of gray.

In addition, the 1960 amendments added adulteration provisions to the Act, stating that a color additive is unsafe for use in food unless there is an authorizing regulation or exemption, in effect, defining color additive as an unauthorized color additive. With regard to that, any food containing an unsafe color additive is adulterated and is subject, under the adulteration provisions of the Act, to regulatory action by FDA.

Many people wonder why color additives are used in food at all. Legitimate uses of color additives in food include use to affect or offset color loss, or correct natural variations in color of food, or to enhance colors that naturally occur, but are weaker than those levels usually associated with a given food. Finally, colors can provide a colorful identity to foods that would otherwise be colorless.

The next part of my presentation will discuss the process FDA uses to regulate color

additives in food. As I mentioned earlier, the 1960 amendment created a pre-market petition process for authorizing color additives.

FDA's regulations, in part 71 of Title 21 of the U.S. Code of Federal Regulations, describe the requirements and process for petitioning FDA to authorize a color additive for use in food, among other product areas. The Federal Food, Drug, and Cosmetic Act and these regulations place the burden for demonstrating safety of any proposed use of a color additive on the petitioner. Only after the petitioner has successfully addressed all relevant safety concern for the proposed use, and FDA has completed its safety review, will FDA list the proposed use in the color additive regulations. And absent that listing, the color additive is not legal and unsafe.

As I said, the underlying statute requires a variety of information be included in the color additive petition. This information includes identity and manufacturing information, as well as information on physical and chemical properties that

are necessary to establish a standard for the color, as well as analytical methods to enforce that standard.

In addition, all relevant safety information must be reported, including all raw data when available. FDA also requires information on the intended use and the other information necessary to estimate a probable exposure to the color additive under its intended conditions of use. Finally, proposed labeling and information to inform FDA's decision under the National Environmental Policy Act are also required.

One of the first decisions that is made during FDA's review process for color additives is to determine whether or not batch certification will be necessary for the color additive. I bring this up separately because the colors that are the subject of the question today are all certified colors. This decision on batch certification is generally based on two criteria.

The first is the amount of variation and the composition of the manufactured color additive. And

the second criteria is the relevance of any variation to the safety decision. Color additives for which certification is required will typically have some variation in constituents that must be addressed in the safety assessment in some way. Color additives not subject to certification may have little or no variation across manufactured batches or have variation, which is of no safety significance.

FDA's safety decision regarding color additives must meet the same legal and scientific standards as other food ingredients. That safety decision must be based on a fair evaluation of the data. In plain terms, this means that all relevant data must be considered, whether it supports or contradicts the safety of the proposed use of the substance. When relevant contradictory evidence exists, all data must be considered in a weight-of-evidence approach, taking into account the relative probative value of differing data to reach a conclusion.

In addition, color additives are held to the same safety standard as other substances

intentionally added to food. That standard is reasonable certainty of no harm as stated in the legislative history. Harm within this standard refers to human health hazard.

It's also critically important to understand that the standard does not require proof beyond any possible doubt, that no harm will result under any conceivable circumstances, as such proof is outside the scope of science. As a practical matter, this means that FDA must always make safety decisions under some level of uncertainty.

So, to summarize, some of the significant points regarding FDA's safety decision include that the underlying information and the decision must address questions of a probative nature and may discount questions that are not probative. FDA's decision cannot weigh any possible benefits of color additives and is only based on the safety of the intended use.

The decision is always made under some level of uncertainty, as it is not possible to ensure absolute safety under all possible conditions of use.

And safety decisions are temporal and are made based on all relevant science available at the time. As such, decisions can and should be reconsidered, based on new information, which raises serious questions related to the safety of the intended use.

FDA's review of color additives is generally divided into a review of the manufacture and use of the color additive and a toxicological review. The review of manufacturing and use data is designed to establish the identity of the color additive, and to identify and estimate the likely consumer exposure to components of the color additive. In addition, this review considers adequacy of analytical methods necessary to ensure the safety of the additive and its ability to meet the standard, as well as identifying controls, or specifications, and limitations necessary to ensure safe use.

The second typical portion of the safety review is the analysis of the available toxicity data to determine whether it addresses relevant safety concerns and whether or not a safe level of use can be established. Steps in this process include a

consideration of the available safety studies and data to determine what information is relevant to the safety decision.

This step is applied both to data presented by a petition, as well as any data that exists in FDA files or in the public literature. Relevant data is then evaluated to determine the adequacy of the data set. This is done through consideration of what data is typically recommended for a given consumer exposure level, as well as through analysis of the relevant data to determine whether or not the data raise additional questions. Finally, FDA's review may also establish an acceptable daily intake or ADI.

Some additional elements with regard to the review that I wish to present is that the review is iterative. FDA places the burden at all times on the submitter to address safety questions until all are resolved. This can result in requests for additional data or analyses, or in the imposition of additional limitations on the color additive through the review process.

FDA guidance documents and specific guidance

on a proposed submission are a starting point to judge whether or not safety has been established, whether or not sufficient data have been presented. But guidance is not binding on industry or FDA to the exclusion of better or more appropriate testing methods. And FDA must issue a regulation permitting the use before such use is legal.

Now, I'm going to discuss two critical aspects of those reviews in somewhat more detail -- specifically, the estimation of consumer exposure and determining what consumer exposure may be safe.

A variety of approaches for estimating consumer exposure may be employed in safety assessments; factors in choosing the approach in a specific circumstance, including applying the principle of effort commensurate with likely risk, as well as the availability of more sophisticated and detailed data necessary for the most accurate methods.

All methods must be applied to produce a suitably conservative estimate designed to provide an adequate margin of safety. The typical approach for

pre-market safety assessments employs information from foods consumption surveys, combined with probable or maximum concentrations in food, to arrive at a range of likely exposures across the consumer population. However, more or less sophisticated approaches may also be appropriate in some circumstances. For example, disappearance or poundage data may be sufficient in circumstances where a substance is used in a wide variety of foods such that it will be present in the diet in a large proportion of the population.

As I mentioned earlier, guidance of safety testing can serve as an initial reference to determine if the relevant safety data is adequate to address the anticipated consumer exposure. However, the review of the safety data itself actually establishes what data is required to establish that question of safety.

For certified color additives intended for use in food generally, the data set would generally be analyzed to determine a point of departure, based on the most appropriate endpoint in the most

sensitive species. This point of departure is typically either a no-adverse-effect level or a benchmark dose level.

Safety factors are then typically applied to this point of departure to ensure an adequate margin of safety, considering inter- and intra-species variations and any limitations on the data set. The result is either the establishment of an acceptable daily intake or an acceptable margin of exposure.

Once the review is complete, color additives are listed in specific areas of the Code of Federal Regulations. Part 73 is our listing of color additives exempt from certification. Part 74 is our listing regulation for color additives subject to certification, and including those color additives that are the subject of today's meeting. Part 81 provisionally listed certified color additives. Each regulation addresses specific criteria, including identity limitations on conditions of use, specifications, and the labeling requirements to ensure safe use.

This table lists certified color additives

authorized for food use. Most are permitted in foods, generally, and the year of approval is shown. As I mentioned, for a color additive to be legally used under a regulation, it must conform to all aspects of the regulation. If batch certification is required, then the batch from which the color additive is drawn must be certified. The color additive must continue to meet the identity specifications, and other specifications, and must be used under the limitations in the regulation.

Because FDA must certify individually manufactured batches of some colors, FDA has annual information related to production of these additives. In 2010, FDA certified batches of color additives amounting to over 15 million total pounds manufactured for all uses.

This table compares the per-capita exposure estimated from the data on amounts certified for colors listed in the previous slide, with the acceptable daily intakes estimated for adults and children, based on FDA's evaluation of the toxicity data for these colors. In all cases, there is a

sufficient margin of safety for the effects identified in chronic animal data on these colors.

This table is a modification of the previous one, which incorporates a common approach to exposure estimation using this type of poundage data for substances used in a wide variety of foods. This approach assumes that all the additive thought to be consumed in a given year is consumed by 10 percent of the population. In this case, although margins between EDI and ADI narrow, the EDIs are still below the ADIs, ensuring an adequate margin of safety for the effects observed in the referenced studies.

The remainder of my slides deal with the committee charge and the questions that FDA has posed for the committee to address. And forgive me; I'm going to simply read these into the record.

The Food Additive Committee charge for today and tomorrow is, the task before this Food Advisory Committee is to consider available relevant data on the possible association between consumption of certified color additives in food and hyperactivity in children, and to advise FDA as to what actions, if

any, is warranted to ensure consumer safety.

Question 1. In the review of published research presented in overview and evaluation of proposed association between artificial food colors and attention deficit hyperactivity disorders, ADHD, and problem behaviors in children, studies were evaluated based on the criteria described in Part 3 of the review.

Were these review criteria appropriate in the evaluation of these studies? Should the criteria be modified in any specific way? And if so, what is the basis for the committee's recommendation? Are there other criteria or other studies that should be considered? And if so, what is the basis for the committee's recommendation?

Question 2. Do the current relevant data support FDA's conclusion, as set forth in the September 1, 2010 Interim Toxicology Review Memorandum, that a causal relationship between consumption of certified color additives in food and hyperactivity in children in the general population has not been established?

Question 3. The National Institute of Health's 1982 consensus development panel on defined diets and childhood hyperactivity concluded that for some children with both attention deficit hyperactivity disorder and a confirmed food allergy, dietary modification has produced some improvement in behavior.

The panel recommended that elimination diets should not be used universally to treat childhood hyperactivity, with or without the presence of food allergies, since there is no scientific evidence to predict which children may benefit. The panel, however, also recognized that initiation of a trial dietary treatment or continuation of a diet in patients whose family and physicians perceive benefits may be warranted.

Are these conclusions and recommendations still relevant today, in light of subsequently published studies, especially as those conclusions and recommendations apply to certified color additives?

Question 4. Under current FDA regulations,

the label of any food to which a certified color additive has been added must declare the color additive as an ingredient by its certified name; for example, FD&C Yellow number 5.

In light of the scientific evidence presented to the committee concerning the consumption of certified color additives in food and hyperactivity in children, what additional information, if any, should be disclosed on the product label of foods containing certified color additives to ensure their safe use in food?

Question 5. Regarding the possible association between consumption of certified color additives and hyperactivity in children, are additional studies necessary to address any questions that have been raised, as to whether or under what conditions, the continued use of the certified color additives is safe? If so, what type of studies would you recommend?

With that, I believe we have time for clarifying questions.

DR. ACUFF: Thank you, Dr. Cheeseman.

Does the committee have questions? Dr. Waldrop?

MR. WALDROP: Thank you, Dr. Cheeseman. I have two questions. One, it looked like on the list of certified color additives approved, that the last one was approved in 1987. And if so, is there any reason why there have not been, or do you have any sense of why there have not been additional color additives approved since then?

DR. CHEESEMAN: I should point out that that list is certified color additives for use in food. I don't have a reason for you regarding why additional color additives haven't been approved. I don't believe -- and I'm looking at someone in the audience to nod his head up and down. I don't believe we've received a petition. It's up to the manufacturer to submit a petition for a color additive for that process to carry through to regulation. We have regulated certified color additives for other uses since 1987.

MR. WALDROP: A second question. Your chart on how color additives, the pounds certified for last

year, and then there was a per-capita exposure. How has that changed in recent years?

DR. CHEESEMAN: The poundage I believe has increased over time. I apologize. I don't have those numbers for you, but I could get them.

DR. ACUFF: Dr. Fenner-Crisp?

DR. FENNER-CRISP: I have a couple of questions, the first being, is the agency obligated to review regulatory decisions for color and other food additives on a regular basis?

DR. CHEESEMAN: We're not obligated to review them on any particular cyclic basis, no. We generally take an approach that's based on a number of factors, including in this case, a petition from external sources, requesting that the agency consider the review.

DR. FENNER-CRISP: So usually, your trigger for review is probably from an external source, as opposed to an internal priority setting for review.

DR. CHEESEMAN: No. I wouldn't say that. We can initiate review of food or color additive safety internally, based on information that comes to

our attention through, for example, the public literature.

DR. FENNER-CRISP: I have another question, two other questions. Is the required data set for color additives the same as for the data set for other food additives?

DR. CHEESEMAN: Generally speaking, the guidance for direct food ingredients, which includes both food additives and grass ingredients, also covers color additives. But I would point out, those are recommendations, and what's required is, in fact, a subject of the review process, a product of the review process itself.

DR. FENNER-CRISP: Then lastly, are your data evaluations available publicly?

DR. CHEESEMAN: Absolutely. Once a regulation is published on a color additive, all information in the files are publicly available with the exception of confidential commercial and trade secret information.

DR. ACUFF: Dr. Winter?

DR. WINTER: Dr. Cheeseman, you were talking

a bit about your process for setting safe levels of exposure, and you mentioned the points of departure, and your examples, and your slide were the no-observed-effect level as well as benchmark dose.

Does FDA make distinctions between a no-observed-effect level and a no-observed-adverse-effect level? And if so, are there criteria for determining what represents an adverse effect?

DR. CHEESEMAN: We make the determination of an adverse effect as part of our review. We don't have a specific criteria, so that determination is made as part of the safety review and review of the safety data. And we don't make a distinction, official distinctions between no-effect levels and no-adverse-effect levels.

DR. WINTER: Thank you.

DR. ACUFF: Ms. Lefferts?

MS. LEFFERTS: Thank you. I have a couple questions. Were any of the ADIs based on neurological or neurobehavioral endpoints? It wasn't listed in the table what the endpoints were that were used to determine the ADIs.

DR. CHEESEMAN: I believe that information was supplied in your background document, but the answer to your question is no. None of those are based on neurological endpoints.

MS. LEFFERTS: Are the studies that are done on the colors, do they contain the same impurities of toxicological concern that are specified by FDA?

DR. CHEESEMAN: You're asking whether or not the colors that were tested are representative of the colors that are regulated?

MS. LEFFERTS: Correct.

DR. CHEESEMAN: I believe they are, yes.

MS. LEFFERTS: Also, I don't know if this is appropriate, but in light of the fact that the legal standard that you explained requires that data establish that the proposed use of the color additive is safe, and that safe is defined as convincing evidence that establishes, with reasonable certainty, that no harm will result from the intended use of the color additive, and also in light of the fact that the review of published research was a little bit broader than hyperactivity . It also included

problem behaviors in children.

I just wondered if it was appropriate for the committee to consider not only the possible association between consumption of certified color additives in food and hyperactivity, but the slightly broader issue of adverse effects on behavior, which was how this meeting was described in the Federal Register, of which of course hyperactivity and ADHD is a very important component.

That's one aspect. And then the other aspect was if it would also be appropriate for this committee to look at that legal standard to determine if there is convincing evidence, that establishes with reasonable certainty, that no harm will result from intended use of certified color additives. And perhaps this is also a question for the committee or the chairman. Thank you.

DR. CHEESEMAN: I think in fact it is a question for the chair. And I'll allow Ms. Jeletic to -- and I'm sure she will correct me if I misstate.

That was a long question, so I may have to ask you to repeat some of it. Our expectation is

that you consider the adequacy of the review that is presented. So to the extent that adverse behavior is covered in that review, I think that is appropriate. And it is, in my understanding of the operations of this committee, within the discretion of the chair to permit that.

With regard to the latter part of your question, with regard to the issue of reasonable certainty of no harm, I think you can certainly consider and make recommendations based on that standard within the terms of the information that you're looking at. But I would caution that you are in fact not looking at the total data set. We've asked you to review a fairly well-defined issue and a fairly well-defined data set that supports that issue, and also of course given you the charge to give us recommendations about whether or not we have adequately addressed that particular question within the safety assessment.

DR. ACUFF: Dr. Jones?

DR. JONES: Tim Jones. There are sort of two issues here. One is assessing the strength of

evidence, which I understand. The second is the standard to which it's applied. And to use the term "no harm" unqualified suggests that you could interpret that as ever, in anyone, which many would argue is unattainable. And there is sort of the unavoidable issue, and you mentioned this, comparing it to other foods which are accepted; some may say caffeine or a host of other additives, which people generally don't argue about.

So how are we to handle that?

DR. CHEESEMAN: Well, no harm is generally applied to the overall population. If there are specific subgroups that require protection, then there are a number of regulatory options that FDA has to address that issue, that range from labeling to not approving or revoking an existing code, an existing regulation.

So I think you would need to address it within the constraints of the general population as a starting point, but you have discretion, I think, under the questions to apply other recommendations to FDA for specific subpopulations, if that is in fact,

something that the committee consensus suggests should be recommended.

DR. ACUFF: Dr. Voorhees?

DR. VOORHEES: Thank you. So I wanted to just ask you a bit about the batch certification process, so that I can understand it a little better. So you're saying that every batch of a certified color additive is tested by FDA, or is it tested by industry under FDA guidelines? And what does it test for? Does it test for the chemical entity itself?

To what extent does it test these batches for contaminants and other chemical entities that might be the product of the synthesis process, or bacterial contamination, or endotoxin contamination, any other kind of impurity? And what is the limit on impurities that the FDA establishes for batch certification?

DR. CHEESEMAN: First of all, it's tested by FDA. Industry in fact funds that program through a user fee, but FDA does the analysis of batch samples. The testing that is done is relevant to the standard that is established in the regulation. And so it's

essentially an analytical chemistry test to compare the batch presented to the specifications, including specifications for impurities, and reaction byproducts, and initial reaction components that are listed in the standard in the regulation.

DR. VOORHEES: So certainly you must set some sort of a margin of acceptability for those analytical standards. Do we have access to information on what that range of acceptable purity are for each of these?

DR. CHEESEMAN: I'm not sure what you mean by margin of acceptability. The standard that the color additive must meet is written into the regulation. And so if there's a limit on impurity, then the batch has to meet that limit.

DR. VOORHEES: So FDA has those data and other items available on the website? Could one see what the batch analytical result is?

DR. CHEESEMAN: I'm not sure whether that's releasable information, and I'm getting a head shake that no, it is not. The batch certification is actually not performed by my office. It's performed

by the Office of Cosmetics and Colors.

DR. ACUFF: Dr. Vugia?

DR. VUGIA: Duc Vugia. My question is about the estimated daily intake, or EDI. I understand that it's based on the pounds certified of each of these color additives. And, therefore, on the slide you show about the high consumer EDI versus the acceptable daily intake, ADI, some of the per-capita exposure of some of the color additives are reaching the level close to the ADI levels listed here, particularly for children.

My question is, you mentioned that in 2010, the poundage, the pounds certified, had been increasing over the years. What would the actions be when these per-capita exposure numbers actually either meet or exceed the ADI for either adult or child?

DR. CHEESEMAN: I wouldn't want to speculate on future FDA action based on that eventuality. There are a lot of options we might exercise, but I wouldn't want to speculate.

Let me clarify a couple of things with

regard to that information presented. It was presented really to provide you with some background information. I would not want to suggest that that would be a final exposure assessment for color additives if, for example, this committee recommended additional considerations for the FDA to take into account with regard to the safety. It was really meant to present background information and give you a notion of how the anticipated exposure would relate to the acceptable daily intakes as currently established.

One of the outcomes of this committee's deliberation could be an alteration of how FDA considers the acceptable daily intake. So in short, that's not the final exposure number. It was provided for background.

DR. ACUFF: Dr. Gray?

DR. GRAY: Thank you. You indicated in your presentation that when a petition for a food color comes in, there's an expectation that all of the data would be shared, including, I presume, the raw data to allow FDA to make its evaluation. Is that also

expected in the case of a petition like this, that all of the raw data are available, and was that the case with this petition?

DR. CHEESEMAN: A petition like what?

DR. GRAY: This petition that you received to look at this question.

DR. CHEESEMAN: No. There is in fact a different standard between food additive petitioner, or color additive petitions in this case and the petition from the Center for Science in the Public Interest, which I think you're referring to. The petition received from CSPI is what we refer to as a citizen petition, and it generally is not required to provide the raw data. But the risk of not providing as much data as can possibly be provided is that the agency won't be persuaded to take action.

This petition is in fact a petition to attempt to persuade the agency to take action on its own, at which point if the agency decided to take action, we would have to gather data that would support that action in our own internal administrative record.

DR. ACUFF: Dr. Blakistone?

DR. BLAKISTONE: I wanted to continue a little bit with Dr. Voorhees's question. I'm interested in batch-to-batch variation. Is it possible that in screening, you might fail to pick up some previously unknown contaminant, and, therefore, something could slip into the food system that might not have previously been there? How do you manage against that, or can you manage against that?

DR. CHEESEMAN: Well, anything's possible, and we certainly can't analyze for all possibilities. I think the best answer that I might be able to give you to that, and perhaps the answer to that question, is the fact that the regulation has to be taken as a whole. And for the colors we're talking about, the manufacturing process is specified in the regulation with some detail.

So our safety review is focused on what is reasonably expected to be in the color additive. It's certainly always possible, and FDA can't write regulations to prevent manufacturers from making errors that result in other unanticipated compounds

or other materials in the color additive. And there are other provisions of the statute, other than the batch certification and the pre-market approval statute that would apply in those cases.

DR. ACUFF: Dr. Freeland-Graves?

DR. FREELAND-GRAVES: For clarification, could you please define the term "iterative," which is used on one of your slides?

DR. CHEESEMAN: I'm sorry. Can you say that again?

DR. FREELAND-GRAVES: You had the term "iterative" on your slide. Could you please define that?

DR. CHEESEMAN: Iterative?

DR. FREELAND-GRAVES: Iterative.

DR. CHEESEMAN: The process of review is generally iterative. A manufacturer will make submissions to FDA in the petition process. FDA will review the data, and in most cases, FDA will have additional questions. And so the first cycle of review will result in questions to the manufacturer, and the manufacturer may have to develop additional

data, do additional analyses, or simply provide clarifying information to address those questions. That can happen many, many times in the course of a color additive review.

DR. ACUFF: Additional questions?

[No response.]

DR. ACUFF: Great. Thank you.

So we'll move on now to the next presentation. Jessica P. O'Connell will address the issue of labeling.

MS. O'CONNELL: I'm Jessica O'Connell. I'm from FDA's Office of Chief Counsel. I'm just going to give a brief overview of the principles of food labeling, specifically as they relate to the labeling of products contained in color additives.

First, I just want to make a brief point about terminology. The Food, Drug, and Cosmetic Act distinguishes between the terms label and labeling, and I've provided the statutory definitions for both terms here. But while this legal distinction is relevant under some circumstances, it's not necessarily important for our discussions today, so

I'm going to be using both terms interchangeably.

From FDA's perspective, the purpose of food labeling is to provide consumers with meaningful and helpful information about a food. The food label has been shown to play an important role in the choices consumers make in purchasing their food. And FDA has regulated the food label for more than a century, but the core of its legal authority to regulate the food label comes from the Food, Drug, and Cosmetic Act.

There are three very general categories of food labeling information, which I have listed here. The first category is mandatory information, and this is information that FDA requires to appear on the label. This includes things like the list of ingredients, nutrition information, the net quantity statement, and any other effects that are material about the food. And I'm going to discuss materiality more in a few minutes.

The second category is optional information, which may need not appear. Some optional information is subject, though, to additional requirements in FDA's regulations, and all information must be

truthful and not misleading.

An example of optional labeling information that additional regulatory requirements is the nutrient content claim, like low fat or high fiber. So a company does not have to make a nutrient content claim. If it does, though, it must comply with additional requirements to ensure that that claim is not misleading to consumers. For example, for food to bear a low-fat claim, the food must contain 3 grams of fat or less per reference around customarily consumed. This requirement was put in place by FDA to ensure that a claim like that wouldn't be misleading. And the final category of label information is prohibited information. And the Act prohibits labels from containing information that is false or misleading in any way.

This slide lists four principles that I'm going to discuss in more detail today, and these principles apply to the labeling of any food, but they are those that are most likely to impact FDA's ability to require additional labeling on products that contain color additives.

First, food labeling cannot be false. Second, food labeling cannot be misleading. Third, FDA cannot require that additional information be included in labeling unless such a requirement is necessary to ensure that labeling is not false or misleading. And fourth, a manufacturer may, on its own accord, include additional information in the labeling of its products, so long as that information is truthful and not misleading, and complies with other FDA requirements.

So the first concept is the most straightforward; food labeling cannot be false, and any food is misbranded if its labeling is false. This is a very basic example. A label must declare the presence of all ingredients in the product. And in the context of color additives, any food that contains a color additive that has been certified by FDA must declare the presence of that color additive in the ingredient list.

A food is also misbranded if its labeling is misleading. And the Act provides FDA with some additional guidance as to how labeling can be

misleading. Under Section 321(n), labeling can be misleading by virtue of either the inclusion or the omission of information.

The Act directs FDA to, in determining whether labeling is misleading, and therefore whether FDA can require additional information, take into account not only representations made, so the information that's included in the labeling itself, but also the extent to which the labeling fails to reveal facts that are material in light of such representations, or material with respect to the consequences, which may result from the use of the food.

Therefore, the omission of material facts from a product's labeling may cause that product to be misbranded. The statute doesn't define "material." However, it does set forth two categories of materiality that FDA has used to guide its interpretation of the word "material." Information can be material in light of representations made or suggested regarding the food, or it can be material with respect to the

consequences that may result from the use of the food. So FDA can require that manufacturers include information that is material in their product labeling to prevent those products from being misbranded.

As I mentioned a minute ago, the statute does not define material. However, FDA has generally concluded that the following information could be considered material, and thus could be information that's required in labeling.

First, information about the characteristics of the food itself. So, for example, if a label says the food is low in saturated fat, but the food does not meet the requirements to be considered low in total fat, FDA requires the label to also disclose the amount of total fact in the product.

The amount of total fat in the product is considered to be material, and this requirement helps to ensure that a consumer is not misled to believe that because the product is low in saturated fact, the product is also low in total fat, if it is not.

Second, information that would identify

certain differences in the nutritional, organoleptic, or functional properties of a food than the food it resembles. So, for example, FDA requires that reduced fat margarine that is unsuitable for frying disclose this difference in its labeling.

Another example would be if a food were engineered to have a substantially higher protein level than its traditional counterpart. FDA would likely conclude that the higher protein level was a material difference that would require additional labeling.

Finally, most relevant to today's meeting, the information about the consequences that may result from the consumption of a given food can be material. For example, FDA requires that the labeling of a number of foods bear warning statements, which I'm going to discuss in more detail in a minute, to identify the health consequences that may result from the consumption of those products or the risks associated with their consumption.

I've listed here a few examples of additional information that FDA requires in food

labeling to provide material information. First, FDA often requires disclosure statements to provide additional nutrition information about a product that is material in light of other statements made in the labeling.

So the example here, if a product bearing a nutrient content claim contains more than a specified amount of fat, saturated fat, cholesterol, or sodium, it must also bear a statement disclosing that that nutrient is present in the food. A product containing more than the specified amount of sodium, for example, would bear the statement, "See nutrition information for sodium content" in its labeling. FDA requires disclosure statements such as this one to ensure that a product's labeling provides material information about the product itself to consumers.

FDA also requires warning statements to convey material information about the consequences of use of a given product. And these warning statements are found, generally, in 21 CFR 101.17. I have listed a few examples here that I am going to discuss a bit.

First, FDA was concerned about the potential health effects of consuming high protein products used in very low calorie diets and determined that information about the potential effects of these diets was material with respect to the consequences of consuming such products.

FDA therefore requires that those high protein products bear the following statement: "Warning: very low calorie protein diets, below 400 calories per day, may cause serious illness or death. Do not use for weight reduction in such diets without medical supervision. Not for use by infants, children, or pregnant or nursing women." And this is an example of a warning statement that FDA has explicitly set out the language that has to be included.

Another example of such warning statement is that which is required for foods containing psyllium seed husk and bearing a health claim about the association of soluble fiber from psyllium husk and the reduced risk of coronary heart disease.

When FDA authorized the use of the health

claim, it was concerned about safety risks that were associated with the consumption of foods containing psyllium husk. FDA determined that the potential for esophageal blockage when consuming certain psyllium husk products, due to not consuming adequate fluids along with those products, was a material fact.

Thus, FDA requires that foods containing psyllium seed husks and bearing the health claim, also bear a label statement that informs consumers that the appropriate use of such foods requires consumption with adequate amount of fluids, alerts consumers of potential consequences of failing to follow usage recommendations, and informs persons with swallowing difficulties to avoid consumption of the product. So the language isn't explicitly laid out in the regulation, but they have to provide all that information to consumers.

The final example I've listed here is the warning statement that unpasteurized juice products must bear, which describes the risks to certain subpopulations of consuming juice that has not been pasteurized. These products must bear the

statement -- and again, this is one that's explicitly directed -- "Warning: This product has not been pasteurized, and therefore, may contain harmful bacteria that can cause serious illness in children, the elderly, and persons with weakened immune systems."

FDA has also determined that some information is not material, and these determinations have been upheld in litigation. First, consumer interests, in and of itself, is not a material fact. Courts have considered this question on a few different occasions and have determined that, under current law, FDA does not have the authority to require labeling based solely on consumer demand or interest.

For example, a group of consumers challenged FDA's decision not to require special labeling for milk from cows treated with RBST. One of the assertions made by the plaintiffs in that case was that FDA should require such labeling because of substantial consumer interest in having that information provided in the labeling. The court

rejected this argument, and relying on the information in Section 321 and of the Act, which I discussed earlier, which contains the material concept, held that plaintiffs were, "incorrect in their assertion that, by itself, consumer opinion would suffice to require labeling."

Second, FDA's concluded that not all effects from customary or usual consumption of a food product are material. For example, in 2003, FDA reevaluated a label statement that it had previously required regarding the possible side effects of consuming products containing olestra. FDA determined that there was widespread consumer awareness about the possible effects, and that the effects were relatively insignificant, and therefore that that information is no longer material with respect to the consequences of consuming olestra. FDA, thus, concluded that the label statement was no longer necessary to prevent those products from being misbranded.

So, to summarize, the following principles are the most relevant to the questions before the

committee regarding the labeling of products containing color additives. First, the food product is misbranded if its labeling is false or misleading in any particular.

Second, FDA considers labeling to be misleading if it fails to reveal material information about the consequences that could result from the use of the food product. And, third, FDA has the authority to require additional information, such as a warning or disclosure statement, in food labeling to ensure that the labeling is not false or misleading, and the product is not misbranded.

Are there any questions?

DR. ACUFF: Thank you, Ms. O'Connell.

Dr. Freeland-Graves?

DR. FREELAND-GRAVES: Yes. I want to clarify that these warning statements, over the consequence of use, are for the general population, not subgroups that might be intolerant or have allergies.

MS. O'CONNELL: Warning statements can be directed towards subgroups. Allergies have not been

something that we've required a warning statement, under 101.17 for, before we require separate allergy labeling. But they can be directed, like the juice example I gave, towards specific subgroups.

DR. ACUFF: Dr. Voorhees?

DR. VOORHEES: Yes. My question is similar to the one you just heard. And that is, so there are children who have severe allergies to peanuts and eggs and other things. So where does FDA draw the line of when they're going to provide a warning for a subgroup population and when they're not?

MS. O'CONNELL: Like I said, for allergies, we have a separate regime under which we require labeling. And so a product containing any major food allergen has to declare the presence of that allergen on its labeling in a separate statement that clearly identifies the allergen is there. And so that's something that is handled under that rubric.

I think in determining whether a warning statement would be required under 101.17, it really would be whether that information is material to consumers. And so if it could affect one person in

the U.S., that probably wouldn't be considered material by FDA. But if it could affect enough people that it would be material -- and again, materiality is a sliding scale. It's not defined in the Act. We use the language of the Act for guidance, so it really is a case-by-case determination.

DR. ACUFF: Dr. Gray?

DR. GRAY: Thank you. I'm hoping you can clarify something for me, back with your example from RBST labeling. I believe you said that the court held that consumer interest does not require labeling. My question is does it allow it? That is, can FDA decide to do it, but not be required to do it because of consumer interest?

MS. O'CONNELL: FDA does not have the authority to require labeling solely based on consumer interest. And so the court in that case upheld FDA's decision that it didn't have the authority to do so. Once there has been determined there's material information, then FDA can then consider consumer interest as an additional factor,

but consumer interest isn't material in and of itself.

DR. ACUFF: Dr. Freeland-Graves?

DR. FREELAND-GRAVES: The previous question, you would not require warnings for one person, but it would vary according to the information that would be presented. But, I mean, is there any baseline? Does it have to affect 2 percent, 10 percent, 50 percent, before you would have a warning?

MS. O'CONNELL: There's no hard baseline. Like I said, it really is a case-by-case basis, but it really comes back to whether that information is so -- "material" is the word that's there -- is material that it would be necessary to provide that information to consumers about the product or about consequences of consuming the product.

DR. ACUFF: Dr. Jones?

DR. JONES: I guess presumably, in addition to that, the severity of the effect comes into play?

MS. O'CONNELL: Yes. And so like I said, in the example about olestra, FDA initially required a warning statement about products containing olestra.

And then after those products had been on the market for a while, and FDA was able to obtain more data about the actual consequences of consuming those products, FDA then determined that the possible effects of consuming those products weren't significant enough to require a statement.

DR. ACUFF: Dr. Blakistone?

DR. BLAKISTONE: I think it's getting covered, what I was thinking about. Thank you.

DR. ACUFF: Okay. Thanks.

Additional questions?

[No response.]

DR. ACUFF: Thank you, Ms. O'Connell.

MS. O'CONNELL: Thank you.

DR. ACUFF: We're a little ahead of schedule, but we're going to go ahead and take a break prior to the next presentation. So we'll take 15 minutes and come back at 10:00.

(Whereupon, a recess was taken.)

DR. ACUFF: Okay. I'd like to call the meeting back to order again.

Next on the agenda this morning is the

presentation by the Center for Science in the Public Interest, CSPI. And Dr. Jacobson will begin the presentation, followed by Dr. Weiss, and then followed by Ms. Edelkind.

So, Dr. Jacobson, if you're ready.

DR. JACOBSON: I'm not going to be using any slides, if that affects which lights you have on.

Good morning. I'd first like to thank the FDA for holding this meeting on food dyes and children's behavior. For too long, the agency failed to examine the research, but simply stated flatly that there is, "no evidence that food color additives cause hyperactivity or learning disabilities in children."

While this committee is charged with reviewing the scientific evidence on dyes, I'd like to start by emphasizing the legal standard for judging the safety of dyes. In implementing the 1960 Color Additives Amendment, the FDA apparently had special concerns about the safety of colorings. The FDA's standard for color additives states that, "safe means that there is convincing evidence that

establishes, with reasonable certainty, that no harm will result from the intended use of the color additive."

The term "convincing evidence" is not in the definition of safety for non-color additives, preservatives, thickening agents, and the like. Anything short of convincing evidence should disqualify a color additive from being used. And as you hear presentations today and tomorrow, I hope you'll keep in mind that benchmark of convincing evidence of no harm.

The FDA has prepared an exhaustive review of the several dozen studies on the topic. However, the more I read the report, the more I thought it was a terrible indictment of the peer-review system for publishing scientific studies. It would appear that almost every study is so flawed, that it's a wonder that any of them were funded, any of them were published. But I do agree with some of the reports' conclusions. The literature is confusing and inconsistent, with some studies finding an effect of dyes and others not.

Also, I agree that dyes are not the underlying cause of behavioral reactions, but rather trigger some pre-existing predisposition in some children. And, third, dyes are not unique in disturbing children's behaviors. Other foods or ingredients have similar effects, but that certainly shouldn't let dyes off the hook, especially because dyes, unlike some of those other triggering ingredients like wheat, eggs, milk, are totally unnecessary additions to the food supply, and indeed are used largely to fool consumers into thinking a food contains fruit, egg, or other natural ingredient that it doesn't.

While the body of research is inconsistent, a number of studies identify children who clearly were affected by dyes, and I'll just mention a few studies. In the one study sponsored by the Food and Drug Administration itself, Dr. Bernard Weiss, who will talk just after me, and colleagues, tested 22 children who were kept on a diet free of dyes and certain other additives in foods, and then they challenged the children with dyes on certain days.

One child reacted dramatically, according to her parents, and a second reacted more moderately.

In one of the few studies that used high doses of dyes, 150 milligrams, researchers in Toronto challenged 40 kids, half of whom were considered hyperactive. After eating a diet free of dyes and certain other ingredients, the children were challenged on one day with a mixture of dyes and on another day with a placebo. Compared to the placebo, the dyes decreased the attention span of the hyperactive children, but not the other children. Seventeen of the 20 hyperactive kids suffered impaired performance in a learning test, though other tests didn't detect an effect.

Finally, a British study tested the effect of dyes on 19 children. Their parents suspected that they were affected by foods and then put the children on restricted diets. In this study, the children were kept on their restricted diets, but then challenged with 125 milligrams of a mixture of four dyes. Seventeen of 19 sets of parents rated their children's behavior as worse, and in several cases,

sharply worse, when their children consumed the dyes.

There's a large body of research, or moderate body of research, on kids with hyperactivity or other behavior issues. In the last few years, there have been some studies that helped clarify that body of research and also extend it.

First, in 2004, David Schab and a colleague published that meta-analysis of 15 controlled studies in which children with hyperactivity were administered mixtures of most dyes used in the United States. And a meta-analysis, of course, is never definitive because studies never use identical protocols, and the design of some could be flawed. But the Schab-Trinh study concluded that, on the whole, the studies found a clear effect of dyes on behavior, especially in the eyes of parents.

The authors characterized the magnitude of the effect as being about one-third to one-half that of children ceasing the use of stimulant drugs; in other words, a significant impact. And that average obscures greater and lesser individual responses.

You'll also hear a presentation today from

Dr. James Stevenson, who with his colleagues conducted the two largest studies ever done on dyes and behavior. And unlike the previous studies, these two studies looked at ordinary British kids, kids not suspected of being sensitive to food dyes or other food substances.

The two placebo-controlled studies administered mixtures of four dyes, plus, for some reason that I don't know, the preservative sodium benzoate, and they were given to young children. Between the two studies, six dyes were used, including the three dyes that account for 90 percent of all dyes used in the United States, Red 40, Yellow 5, and Yellow 6.

Even though the children were not being suspected of being sensitive to dyes or other food ingredients, a battery of tests detected a small but significant effect of one mixture of dyes. There are two mixtures of dyes. So they detected an effect of one mixture of dyes on the behavior of three-year-olds, while both mixtures of dyes affected the eight- and nine-year-old kids.

Now, as I'm sure you've deduced, the FDA staff report downplays the findings of most of those studies, partly on the grounds that blinding wasn't proven, or that only parents and not clinicians or teachers associated the dyes with problems, or that the findings in a given study weren't internally consistent.

I recognize that some studies didn't find any effects of dyes. And whether that was due to the luck of the draw in terms of including sensitive kids, inadequate dosages, inadequate testing methods, or other factors is unknowable. Still, some of the studies demonstrate clearly that normally-consumed amounts of dyes can impair the behavior of some children with hyperactivity or other behavioral issues, while other studies are equivocal, as the FDA report acknowledges.

The FDA report states that, "The effects of dyes appear to be due to a unique tolerance to these substances and not to any inherent neurotoxin property." However, that unique intolerance is not terribly unusual, with possibly millions of children

destined to be adversely affected over the years. And remember, we have 300 million people who have moved through the childhood years, and we're going to be getting 300 million more kids, moving through those years where this may be an effect. I should mention it's totally unknown whether dyes have effects on adults. There are really no studies on that.

The FDA staff report says that this latter scenario of unique hypersensitivity can best be addressed by continuing to understand the biomolecular factors that may predispose an organism, meaning a child in this case, to this type of unique disruptive behavioral response, to otherwise non-neurotoxic chemical substances. I maintain that from a health perspective, the supposedly unique hypersensitivity would be best addressed by getting dyes out of the food supply.

The staff report portrays dyes as being similar to allergens. The problem is that some people are sensitive, rather than dyes having some impact. In the case of traditional food allergens,

because only a small minority of people are sensitive, the FDA holds that declaring those ingredients on labels offers sufficient protection. Susceptible people then could figure out if they have an allergy and avoid the allergens. And often, that works.

Learning that one is allergic to foods that cause obvious physical symptoms like hives, diarrhea, or vomiting is a lot easier than identifying substances such as dyes that disturb behavior. After all, how many parents even think that dyes might affect a child's behavior? And that's why Ben Feingold's announcement 35 years ago raised such public interest, that nobody had ever thought that common chemical additives in food could disrupt the behavior of children.

Moreover, no one has suggested that traditional foods be outlawed, like milk or peanuts or wheat. Those have been in the food supply, they're basic foodstuffs, and they're not going to be outlawed. Nobody's pushed for that. But synthetic food dyes have no nutritional value, no other health

benefits. They're not preservatives. They're used primarily to trick people into thinking that a food contains fruit or other valuable ingredients. Dyes would not be missed in the food supply, except by the dye manufacturers. Europe is managing quite well with minimal use of dyes.

While FDA's staff clearly worked hard to identify shortcomings in the studies on dyes, as well they should, they did not highlight why some of the studies may have underestimated the effects of dyes. First, most of the studies used doses much smaller than what many children consume. A dose of 15 to 30 milligrams is less than what children may get in an individual meal or snack. In fact, in 1976, an FDA nutritionist estimated that 10 percent of children between 1 and 5 years old consumed more than 121 milligrams of dye per day. Moreover, according to FDA's certification data, dye usage is now about twice what it was in the 1970s. It's possible that more children would have reacted or reacted more forcefully if higher doses had been used in these various studies.

Most toxicological studies use exaggerated doses to increase the likelihood of seeing an effect in humans or animals because they know that they're not dealing with the most sensitive strain of rat or the most sensitive individuals in a small clinical study. They use, what, 20, 30, 40 people in some of the larger ones. Even the highest doses, 150 and 250 milligrams per day, were no higher than what some children consume on some days.

While the staff report acknowledges a study by Rowe & Rowe, they acknowledged that study was a good dose response study, one of the only dose response studies, which found a linear dose response relationship. The FDA staff report argues that dyes' impact on behavior does not increase linearly with dose, but plateaus somewhere between, perhaps, 10 and 100 milligrams, somewhere where most of the studies -- doses that most of the studies used.

Also, several factors may account for why it can be difficult to detect the effect of a dye challenge. For one thing, kids' behavior is normally erratic. These are kids. So with that kind of a

noise level, it's much harder to detect an effect. Also, children may control their behavior in a clinical testing environment or a classroom better than they do at home. Aren't we all on better behavior here than we are at home and at a doctor's office?

Moreover, while the FDA staff report concludes that foods or ingredients other than dyes, such as wheat or milk, affect some children's behavior, most studies make no effort whatsoever to control the consumption of these other substances, foods that can affect children's behavior. So is it a combination of dyes plus a piece of bread or a glass of milk? Most studies didn't consider that at all.

I started my talk by discussing the legal framework for judging the safety of dyes. I think that a fair reading of the evidence is that there is not convincing evidence of no harm. On the contrary, I believe that there is convincing evidence of harm to at least some children, and then one could debate how many kids -- what is the magnitude of the harm to

some of the kids, either judging by these controlled studies or in some cases the anecdotal evidence, which I think deserves some attention.

If you conclude that dyes significantly impair some children's behavior, or that dyes have not been adequately demonstrated to be safe, the logical next step would be to advise the FDA to bar them from the food supply. However, I suspect that banning dyes would be a challenging and probably futile legal process because, except for Yellow 5, the dyes have been tested only in mixtures, not individually. So there's no way of knowing from all these studies if there is an effect, is it from Yellow 5, or is it Red 3, is it Red 40, or a combination of Yellow 5 and Green 3? No evidence on that.

Hence, as a weaker measure, the FDA could simultaneously require a health notice on labels, as the European Union has done, and urge companies to stop using synthetic dyes, as the British government has done. A warning notice might state something like, "Warning or Notice, the artificial colorings in

this food cause hyperactivity and behavior problems in some children."

A warning label would not be nearly as protective as a ban because each parent would still bear the burden of recognizing that their child is affected by dyes, and then try to prevent the child from consuming dyed foods. And, moreover, restaurant menus likely would not bear warnings. If you buy food from a vending machine, you don't read the label before you put in your money. However, warning labels certainly would help educate consumers and spur companies to reformulate their products without dyes. Thank you.

Dr. Weiss I believe is the next speaker.

DR. ACUFF: Dr. Weiss?

[Pause.]

DR. WEISS: Thank you for the opportunity to address this audience and the committee.

Dr. Jacobson I think is incorrect by saying that the evidence is inconsistent. I think it is very consistent, and I will tell you why, and give you some background information for that conclusion.

He talked about the study I conducted in the 1970s, published in Science in 1980. It happened because Dr. Feingold had called me and asked me if I knew how the Soviet Union addressed the toxicity of food additives. In lieu of my participation at the US-USSR Environmental Health Exchange Agreement, I found out the Soviets paid no attention to food additives.

He also asked me if I believed his hypothesis. I told him that the biomedical community would pay no attention to it unless he conducted double-blind, controlled clinical trials. Well, he said, "I'm 75 years old, and I can't do that." I also put him in touch with Sheldon Margen, who was chair of the Department of Nutritional Sciences at UC Berkeley, who told him the same thing.

But then later that year, 1975, at a meeting in Berkeley, Dr. Margen and I asked Dr. Feingold where all these children whose dossiers he was waving in front of congressional committees resided. He pointed us to the Kaiser Medical Center in Santa Clara, California, where we talked to the

pediatricians on the staff, who told us that they believed there was some benefit to taking kids off dyes.

Incidentally, Dr. Feingold and I were in the harbor at Okinawa in July of 1945 during one of the last kamikaze attacks, so we're bonded in that way.

We decided to do a single-subject design study. My purpose was not to prove or disprove the Feingold hypothesis. My purpose was, as a toxicologist, to see whether or not there was any indication that exposure to artificial food colors could induce any kind of behavioral reaction.

We calculated the amount of dyes in the foods consumed by California children. We came up with the same amount, practically, as the Nutrition Foundation, for their studies that they had sponsored in Wisconsin and elsewhere. Notice that our doses were about 1/50th of the allowable daily intake, which of course depended on endpoints that do not include behavior.

As Dr. Jacobson noted, one of our children was a kind of spectacular responder. Now, the

criteria we used for behavior was to choose a set of behaviors from the parents' selection from a series of inventories that characterized their own child during infractions. And for the endpoints -- acts as if driven by motor, runs away, short attention span, throws and breaks things, whines -- we found sharp differences between the days, seven days during an 11-week period, that the child drank a blend of colors versus other days in which she consumed a placebo drink or a controlled drink.

This is a chart I drew later, showing the comparisons between color-challenge days and placebo days for this child. You can see that there are irregularities during the placebo days at the bottom, but during challenge days, the scores on a version of the Connors scale were much higher.

In the same study and same issue of Science, in which our study was published, Swanson and Kinsbourne published theirs. They administered 150 milligrams as a challenge to hyperactive boys, boys who had been clinically diagnosed as hyperactive. And what you can see here is that a

couple of hours after they consumed the color challenge, they showed an increase in errors on a paired associates learning task.

There are two questions that to me are elicited by these data. One, we know nothing about the pharmacokinetics of food colors. We know nothing about the duration of the behavioral effects that they induce. That's important because a number of other studies, conducted at about the same time, reported no effects, but in those cases, the investigators only scored behavior at the end of a week or several days. So they missed the acute phase of the response.

A study conducted in Toronto by Williams and his colleagues also looked at kids who were on medication because they were considered to be hyperactive. They concluded, well, maybe there was an effect, but I was able to obtain some of the raw data. I replotted them, and if you compare condition 1 and condition 2, you will see that there is a subgroup, that I've circled in blue, that apparently responded in a very forceful manner to the challenge,

which was in the form of a cookie.

An interesting study by Kaplan was conducted in Canada. She provided all the food for the families that participated in the study. And she concluded that some children within this group showed improvement by eliminating food dyes and other constituents labeled by Feingold as potential triggers from the diet. As she noted here, relative to baseline phase, 10 of the 24 children were designated responders.

Now, Dr. Jacobson referred to Rowe & Rowe. Here's a chart showing differences on the days that the challenge itself, the tartrazine was administered, versus placebo days. They decided that certain children could be called responders because of the reliability of their response to the tartrazine challenge. They used several different doses, which Stevens then graphed.

As you can see here, there's kind of a rough dose-response function here. There are virtually no dose-response data here, which we would require if we were to say anything about the risk. So we are

exposing children to dyes about which we know very, very little.

There is a statistical problem here, which I'd like to review with you. At the upper left, you can see two distributions. That is a distribution or a population in which 30 percent of the children, or respondents, are susceptible. Without a challenge, this is the kind of distribution you would get.

If, to a challenge, the susceptible subgroup shifts the mean by one standard deviation, what you would see, as in C, is almost no change in the mean. That is, you would come to the conclusion that there was no effect of the challenge on this population. And what you see on the lower right is a random sample drawn from the population when it's not challenged and when it is challenged. Again, with a very small population, you would see virtually nothing. And then Dr. Christopher Cox, a biostatistician at Rochester, calculated the number of subjects you would need to get a probability value of .01 for different effect sizes.

Now, look, if the proportion of respondents

say is 30, to get an effect size of 1, you would need an N of 265. That tells you why so many studies have concluded that there is no effect.

If you weigh the evidence, which is what we do in risk assessment, to look at the conclusions based on all of the literature, this is what you find; synthetic food dyes at levels found in foods consumed by children can evoke adverse behavioral responses. Such responses include disruptive behaviors in learning and performance impairment. Hyperactivity is not the question. The question is adverse behavioral responses.

My conclusion is that, generally, these responses are acute, not chronic conditions, although their impact on the child's adjustment may be long-term. Not all children appear to be sensitive to levels of food dyes found in those foods, although dose effect functions have only rarely been explored.

Finally, clinical trials must be sensitive to statistical issues arising from sensitive subgroups within a larger sample, but we really don't know if there are especially sensitive subgroups

without more information on dose-effect relationships.

I am bewildered, truthfully, by the response of the FDA to the study from Southampton. They said we have no reason at this time to change our conclusions that the ingredients that were tested in this study, that currently are permitted for food use in the United States, are safe for the general population.

In contrast, you can see the U.K. response, which tells parents to consider eliminating coloring from the diet, and the European Union response, which says eliminate these dyes. Now, the fact that they existed at all is ridiculous to me, it seems.

This is my version of a study published in 1972 in pediatrics. A mother came into the emergency room believing that her child was bleeding or had blood in his stools. It turned out to be a red dye in this cereal. And dogs have only rudimentary color vision. As Dr. Jacobson pointed out, food colors are a marketing device.

But I think the committee should pay close

attention to what Philip Handler said about risk, "A sensible guide would surely be to reduce exposure to hazard whenever possible, to accept substantial hazard only for great benefit, minor hazard for modest benefit, and no hazard at all when the benefit seems relatively trivial." Thank you.

DR. ACUFF: Thank you.

Ms. Edelkind?

MS. EDELKIND: Thank you. I'm very happy to be here, representing the Feingold Association, a non-profit organization helping families avoid synthetic food dyes and other additives for better behavior, learning, and health. At this hearing on synthetic food dyes, I will present our view of the relevant research.

This is a graph of double-blind studies in which children were subjected to synthetic food dye to see if their behavior deteriorated. These are called challenge studies, but it's not a meaningful graph, as results are completely scattered, and critics rightly say they are all over the place.

This is a chart of diet studies in which

children with ADHD were put on a Feingold-type diet, eliminating food dyes and other additives. In most of them, more than half the children got better. Diet studies are often used to prepare children for a challenge study. The procedure is to first recruit some children with ADHD, next, put them on a diet free of food dyes and other additives, do it right by making sure that vitamins, toothpaste, and medications are also additive free, and about 70 percent will respond favorably.

Now, give them some food dye. That's your challenge. But how much? The Nutrition Foundation, which is an industry organization, told researchers the average daily intake was 27 milligrams per day. Nevertheless, many researchers used far less, some as low as 1, 5, or 13 milligrams.

So look again at the challenge studies. Now, graph the percent of children reacting against the amount of food dye used in the study. It's not scattered anymore, is it? When little dye has little effect, researchers report that children are not sensitive to dye, so additive-free diets don't work.

But wait, it was the challenge that didn't work. The diet worked, so well in fact, that 5 or 10 or 20 milligrams of food coloring didn't overcome it for most children. The more dyes ingested, the more people affected.

Almost 40 years ago, Dr. Ben Feingold said "The reaction to synthetic dyes is pharmacological and dose related." The 1994 Rowe & Rowe study verified it.

But how much food dyes do children really eat? It's really hard to find out. Amounts are not listed on product labels, and companies consider it proprietary information. Well, I tried measuring it myself. 13 milligrams of powdered red food coloring on an electronic scale, and 3 tablespoons of white frosting looked like this. But more accuracy was needed. So the commercial red frosting was sent for analysis. I got the report. I did the calculations. The actual amount of dye in 3 tablespoons of commercial red frosting is 58 milligrams.

Now, 58 milligrams of Red number 40 was in one cupcake, but researchers had been told to give

the children only 27 milligrams for a whole day. Maybe a toddler eats one cupcake. A teenager might eat two. Either way, they're getting two to four times the amount of dye the Nutrition Foundation recommended for studies.

Where are the studies on products like these or all these? Yes. Children like the bright colors. But, unfortunately, an increasing number of children are caught up in the growing epidemics of learning disabilities, hyperactivity, depression, and violence. And these children need help.

But medicines come with side effects. In fact, the FDA has mandated black-box warnings for most drugs used in ADHD. Now, there are actually many other good reasons why a child may have symptoms of ADHD. Some are medical problems that can be treated. But with so many possibilities, why focus on food dyes anyway? Aren't they only a small part of a large problem? Well, no. We believe they are a large part of a large problem, but this is one the FDA can fix.

Most synthetic food dyes are petroleum

derivatives, containing lead, mercury, and arsenic. They add no nutritional value and can lead to behavior, learning, and health problems. In Europe, synthetic dyes are already being replaced by natural colors, and the same American companies providing products colored with natural colors in Europe won't do it for us unless the FDA says they have to.

Back to the numbers. Won't removing synthetic food dyes really help only a small number of children anyway? Actually, as long ago as 1986, Dr. Stephen Schoenthaler published a study on 803 New York City public schools. The dotted line here represents the national average of California Achievement Test scores in the '78-'79 school year, and New York City schools were not doing very well.

By implementing what was essentially the Feingold diet, the schools achieved dramatic academic improvement -- remember, this is average over a million children -- with scores rising almost 16 percentage points. And what's more, two-thirds of the children who had been over two years behind caught up to grade level.

In several prison studies, a more natural diet led to a decrease in anti-social behavior, which prisons measured, by almost 50 percent. And 20 percent of inmates improved so dramatically that if they had had a better diet all along, they may never have been there in the first place.

Here are 179 studies on food dyes, and I know it's tiny print, but your handout includes a copy at the very back of it. Food dyes can cause a variety of symptoms; headaches, stomach aches, sleep disorders, speech disorders, behavior problems, learning problems, neurodevelopmental changes, DNA changes, bronchoconstrictions, sperm abnormalities, dopamine changes, serotonin changes, and loss of zinc.

Until recently, Blue number 1 was added to the tube feedings of patients unable to eat normally. In 2003, the FDA asked doctors to stop doing that since patients were dying, not from their disease, but from the Blue number 1, which apparently caused refractory hypotension and metabolic acidosis, and also, incidentally, turned their colons bright blue.

The FDA told the doctors, "in-vitro evidence that Blue number 1 can be a mitochondrial toxin lends plausibility to the idea that Blue 1 could cause these kinds of serious adverse effects." Remember, the FDA itself said this. And, well, what's a mitochondrial toxin doing in the food?

We are now in the third generation of Americans exposed to significant amounts of food dyes. These amounts are averages. They're not what any particular individual eats in the real world, but see how the average daily consumption has increased since 1955, from 12 milligrams to 62 milligrams, a fivefold increase. But we still wanted to know what real people are really eating. So at the CSPI, we sent several products to an independent testing laboratory to determine how much food dye they contain.

This slide shows how easily a child can consume hundreds of milligrams of dye in a day. Such levels of food dye have never been studied in children. Realistically, not every child eats like this every day. On the other hand, some may eat far

more. And this is just the tip of the iceberg. Thousands of other products contain food dye, and this slide doesn't include the blue toothpaste, green mouthwash, and children's vitamins many children take every day.

So does this mean these studies are irrelevant or useless? No. It means you have identified that some children are so sensitive, even this small amount can affect them, with many more affected as you increase consumption.

Children are exposed to a lot of harmful chemicals these days, but food dye is one that can be controlled. If we don't, if we continue to expose American children to increasing amounts of synthetic food dyes, are we setting them up for a lifetime of frustration, illness, and failure? The FDA recently wrote that the food dyes' effects on behavior was due to a, "unique intolerance, and were not the result of any inherent neurotoxin properties of the dye."

When we started the Feingold Association 35 years ago, we believed that, too. Today, we disagree. After working with hundreds of thousands

of families, we now see that it's not just the ADHD child who gets hyperactive when he eats artificially colored fun foods, but his normal siblings and his parents are also affected, just to a lesser or different extent.

So we heartily support what Dr. Stevenson has demonstrated, that it isn't just a small select group of children who are affected by food dyes; it's all of us. And we're honored to join with this eminent group of scientists to ask for a ban on synthetic food dyes in this country. Thank you.

DR. ACUFF: Thank you.

Now, we'll have questions of the committee members. So you can address who you'd like to address the question to. Dr. Blakistone?

DR. BLAKISTONE: Either Mr. Jacobson or Ms. Edelkind. I'm wondering if you have a sense of what the percent population is. It sounds like it's almost systemic, so I was just looking for, perhaps, some data. Maybe CSPI has done some studies to collect that kind of data.

DR. JACOBSON: It was a long trip up here

for not much information to give you.

[Laughter.]

DR. JACOBSON: Is your question what percentage of children are sensitive to food dyes?

DR. BLAKISTONE: Yes. I just wondered if you had a sense of what value are we talking about here.

DR. JACOBSON: I think it's very hard to know. Some studies found 80 to 90 percent of kids affected. Other studies found a lower effect, and then that's with kids with behavior problems. If you give credibility to Dr. Stevenson's studies, then it's a much larger percent because it brings in the whole population of children.

So I think -- and then with all the studies on hyperactive kids, the FDA says they're all flawed, inadequate blinding, or we don't know what the blinding is. And on the other hand, some studies may have used the wrong tests. They may have looked at the kids at the wrong time, inadequate dose. So it's really hard.

I come to the conclusion that there are an

awful lot of kids who are affected, and there's a price to pay, and there's no benefit. Benefit doesn't matter under the law. The law just says, if there's a risk, get rid of it.

MS. EDELKIND: I would just like to add that in the background information you've all heard today already, what's called the high user, the high consumer numbers for the various dyes were listed. What wasn't done is they weren't added up. Of course, I added them up because I add things up, and it was well over 400 milligrams a day.

No study has ever, ever been done on that amount of food dyes on children, adults, or anybody else. And until they've done that kind of study to push the envelope a little bit, they're not going to know what it can do to people. If we already know that 10 percent of people, let's say, are high-end users, we really need to know what's happening to them and what's likely to happen to them.

If we have an ADI, which by definition means that it won't have an effect, it's safe, where are the studies to show that the ADI levels really are

safe? And I'm not talking about hyperactive kids; I'm talking about anybody. If 10 milligrams makes Johnny hyperactive, maybe 400 milligrams makes somebody else hyperactive who normally wouldn't be. And we really need to know that. But if you don't do the studies, you don't know.

DR. ACUFF: Dr. Jones?

DR. JONES: Tim Jones. And I have two related questions for Drs. Jacobson and Weiss.

First, Dr. Jacobson, most of the studies that we've been presented with are very small, preselected populations, and the number of affected children is in the single digits. But you estimated or said that you thought there were millions of children in the country potentially affected.

Could you just talk a little bit about how you came to that estimate?

DR. JACOBSON: It's not just a single-digit percentage of kids who are affected. Some of the studies find much larger percentages of this subgroup. And hyperactivity affects, what, about 10 percent of boys, 4 percent of girls, perhaps. And

every year, we have a new crop of kids. So if you look over at what percentage of all people have been affected, I think, surely, it's many millions of kids. Many millions of Americans will be affected.

But I don't know how many four-year-olds that are at a given time. But because of the weaknesses of the studies, I think it's hard to know just what percentage of hyperactive kids, and certainly hard to estimate what percentage of normal kids are affected by maybe in lesser amounts.

But why accept any reduction? Why accept any impairment of kids' behavior whatsoever? And hyperactivity isn't just running around. It affects the kids' abilities to have friends, to study, to have a happy family life. Why impair that for no benefit?

DR. JONES: Then my second question for Dr. Weiss is, I guess, in light of that explanation, when you showed your statistical data and suggested that 200 to 300 subjects would be necessary to show any statistical effect, what do you think the barriers were that made, essentially, none of the

studies that we looked at or we've been presented with had even 10 percent of the number of subjects that would be required to show such an effect?

DR. WEISS: Sure. That's a problem in experimental design. If you're looking for an effect at a low dose, you need a larger population. That's the reason, say, for the cancer bioassays. The FDA and the EPA look at small numbers of animals at high doses, because to determine the impact on the population as a whole for, say, a risk of 1 in 100,000, you need 2 million rats.

We haven't done those studies for food dyes. They were put on the market before we had an adequate appreciation of their adverse effects, and the fact that a few studies have turned up susceptible subpopulations could be attributed to either a low dose or the fact that they have selected a particular population. But the risk question still persists.

Remember that regulations in EPA are based on the identification of susceptible subpopulations. That is a factor that they have to take into account when setting the acceptable daily intakes or

reference doses.

Isn't that true, Penny?

DR. FENNER-CRISP: Yes. May I add amend that comment? In fact, it's true across the government because there was an executive order that President Clinton signed back during his tenure that required every federal agency that dealt with children, one way or the other, must make special findings for them.

DR. ACUFF: Dr. Voorhees?

DR. VOORHEES: Let me start with Dr. Jacobson.

So I'd just like to get Dr. Jacobson to talk for a moment about -- he talked about the law requiring no harm. And how do you make the distinction between harm and effect?

DR. JACOBSON: Well, you look at what are the effects.

DR. VOORHEES: So if a child is having sleep problems, is that harm? If it takes an hour to go to sleep rather than 15 minutes, would that meet your threshold for a definition of harmful effect?

DR. JACOBSON: Yes. Yes, it would. And I recognize your question. You need permanent, severe harm or something short of that.

What about vomiting? Is that harm? I think most parents would think it is. I'm not saying food dyes cause that, but there's a range of concerns. And here, to have something that can reduce a kid's -- we would contend that it reduces a child's attention span, increases irritability, those are certainly harms. Impairments of a kid's life; does it change for the worse the way a kid goes through the day? And I think, certainly, yes.

DR. VOORHEES: So in your view, all of the effects that are described in this literature, you see as significant harmful consequences of these studies on the food additives and the elimination diets.

DR. JACOBSON: I'd have to think about every single thing, but to see a significant change in a child, I would say yes.

DR. VOORHEES: Let me ask you another question. So it is my understanding that the FDA has

the legal authority to determine what kinds of studies are submitted before or even after they prove. So in your advice or your petition to the FDA, I didn't see that you asked the FDA to require industry to go out and conduct further studies at higher doses, at better blinding. You didn't make that recommendation to the FDA.

DR. JACOBSON: That's right. What we would like to see is for the FDA to revoke the approvals of these dyes. If a company, then, wanted to do the tests, it could do the tests, and go back to the FDA, and say, we've proven that Green 3 is safe. Approve it. But the FDA – I suppose the FDA could say that that it wants to have more tests before it makes a decision. But we don't think it needs -- it shouldn't take on that burden.

DR. VOORHEES: But, in effect, what your petition does is it indicts a whole collection of food colors. But we don't actually have studies, color by color.

DR. JACOBSON: That's right.

DR. VOORHEES: It is entirely possible,

would you not agree, that some might have effects, and some might not, and your petition is asking the FDA to blanketly withdraw them all?

DR. JACOBSON: That's right. But that's the reason why I think the FDA didn't ask the committee should we ban food dyes, because industry would say, well, you can't prove that Green 3 or Yellow 6 or Red 40 is dangerous. I think it's a legal challenge. Congress could do that. Congress can just say get rid of them all. And I think there's a much lower threshold for FDA action for a warning notice, where it could say, may cause hyperactivity.

DR. VOORHEES: I wasn't addressing the warning part of your petition, but rather the proposal to have them ultimately removed. For that, doesn't it strike you that we don't have enough information about the dyes individually?

DR. JACOBSON: I would say let the companies prove that their dye is innocent. But meanwhile, almost all the studies going back -- almost all the studies going back to, what, 1976, 35 years, have used mixtures.

Do you throw all that out or do you say, hey, it looks like it's a suspect category of additives. We shouldn't expose our kids to the risk until companies prove that they're safe?

DR. VOORHEES: If the evidence were to reach the level of saying there's a probable effect of these mixtures, then isn't the next logical step to try to deconstruct it into what are the offending agents?

DR. JACOBSON: As a scientific issue, yes, but as a public health issue, no. These kids are being affected, we think -- there's obviously arguments about that. We think kids are being affected every day.

How many years do you think it would take to get these chemicals tested individually, and in combinations, with high-quality tests that pass muster by FDA's staff with adequate blinding, sufficient number of kids, adequate dosages, and all the rest? You're talking decades. And industry has had decades. They've had 30 years of warning that food dyes may affect kids' behavior. Why didn't the

companies plan ahead and sponsor these tests? They haven't sponsored one single test since Dr. Feingold has done this.

DR. VOORHEES: Thank you.

Could I ask Dr. Weiss a question?

DR. WEISS: Chip, let me reply to that question. The EPA has been dealing with complex mixtures for decades. I think the first report published under Academy auspices came out in, what, 1978, and they're still dealing with it. So there's nothing new about this question.

We are exposed in the environment to a combination of chemicals. If you have enough information to say this one class of chemicals, say organophosphates, is harmful, then you try and reduce exposure.

DR. VOORHEES: I wanted to ask you about your study, your 1980 study. So you took a population of preselected children who were believed to be susceptible. That's how they got enrolled. That's what you said.

DR. WEISS: Yes.

DR. VOORHEES: Of those, you got one that you report as a dramatic responder. Could you comment on why you think a population, preselected to be responders, would show one responder under the blinding condition? So in other words, why under open label, do you get 22 responders, and then when you go under blind conditions, you get one dramatic responder out of that set?

DR. WEISS: No, no. The studies that have been reviewed by Dr. Jacobson and myself were all double-blind, controlled clinical trials. I wouldn't deal with an open label setting.

DR. VOORHEES: No, no. I'm saying, you preselect them to be responders before you enroll them, and then you put them in the double-blind.

DR. WEISS: Let me explain why we did that study. The question that I addressed was, is there any evidence at all that such an effect takes place? Period. It was a simple toxicological question.

DR. VOORHEES: But in a group thought to be responsive.

DR. WEISS: Well, thought to be.

DR. VOORHEES: Yes. That's what I'm asking you. If they're thought to be responsive, and then you get one dramatic responder under the double-blind study, I'm just asking you how you understand how that happens that way. Why didn't you get 22 under the blind conditions?

DR. WEISS: Limitations of time and budget, like other studies.

DR. VOORHEES: If you look across this whole literature, you see that there is a general pattern that children are selected for the studies based on the concept that they were pre-known by their parents to be responders. Then they're enrolled in a blind study. And in the blind phase of the study, you see a fairly consistent pattern in which fewer responded in the blind part, whereas most or all of those enrolled in many of the studies, were thought to be responders.

DR. WEISS: That's not true.

DR. VOORHEES: No. In many of the studies, it is true.

DR. WEISS: But a number of studies chose a

population that were classified, say, as hyperactive. And then the question was, are these children responsive to certain elements of a diet, say, food dyes? These were double-blind, controlled clinical trials on a selected population. Now, you're asking whether you can extrapolate from those studies to the population at large. And my response to you would be, well, as Dr. Jacobson said, do the studies.

DR. VOORHEES: No. I wasn't really asking that. I was asking --

DR. JACOBSON: Let me try to get at this, because I had the same question. Why don't you see 90 percent of kids responsive of these preselected kids? And I think some of it is, the parents think that the kid is sensitive to food, but they're not certain that it's the food dye, maybe it's something else.

As Shula mentioned, in a lot of these studies, kids are put from their normal diet onto that oligoantigenic diet or the Feingold diet that removes not just food dyes, but flavorings -- BHA, BHT, wheat, soy, milk, and so on, very restricted

diets. And very high percentages of children respond, 50, 80 percent. Their behavior improves -- 50 to 80 percent of the kids, their behavior improves significantly.

I suspect some parents are wrong, that it's not the food dyes. It's probably something else, and may not be anything in the food. It might be something else. You can't expect 100 percent to respond. And maybe it's the dosages also.

DR. VOORHEES: But the petition is really about whether kids respond to food colors and whether they produce harm, which would justify their removal. So if you're doing a study, and you put them on a special elimination diet, and they all improve, let's say, because they're responsive to that, and then you try to test whether it's food colors by challenging them with food colors, and you get a very small percentage of those who then respond to those challenges, doesn't that present a bit of a problem for the hypothesis that it is the color that's causing the problem, and should therefore be acted on?

DR. JACOBSON: No. I don't think so. As the FDA says repeatedly in their staff report, food dyes are not the only substance that causes a problem. But it's the one that the committee is dealing with -- and we're not contending that a hundred percent of kids are affected. We're not saying that food dyes are the only thing. We're saying this is a category of substance that causes these reactions in some percentage of kids, and I don't know exactly what percentage it is, and the FDA can easily eliminate them.

DR. WEISS: Chip, can I make one other point? Look, in laboratory studies, we do this all the time. In fact, the FDA is about to embark on a study of bisphenol A. And they have specified that in their study, to be conducted at NCTR, they will be using a phytoestrogen-free diet because phytoestrogens have estrogenic properties. They are doing it to try to isolate the toxic effects of bisphenol A during gestation.

That's another point. We have no idea about the neurodevelopmental toxicity of these food

additives. They have not been tested according to the protocols that the EPA has been working on for years. And I just don't understand. I can't grasp why this has not been done.

DR. VOORHEES: I think you just made my point, which is, when you're trying to isolate an effect, what the FDA is doing in the BPA studies is that you eliminate estrogens from the diet first. Then you can determine the effect of that item which you're particularly interested in.

That's what a lot of these studies are conceptually trying to do. First, they eliminate what they think are all the offending agents. Then they challenge with the agent what they think is suspect. And under those conditions, many of these studies find no effect or diminished effect. And that's what I'm trying to understand, why that should be true, if that is the offending or one of the offending agents.

MS. EDELKIND: I'd like --

DR. JACOBSON: But we're not saying that it's the only offending agent.

DR. VOORHEES: No, I know that. I'm just saying, if it's an offending agent, why would you not get a --

MS. EDELKIND: I'd like to comment here, if I may. One of the things that we find, from a very long history of being in contact, of course, with the parents, as well as studying the research, when the children are on the Feingold diet, which is -- by the way, we don't eliminate milk, wheat, and soy unless it's a particular allergy or something for the child. So it's a livable -- very nice, livable diet, the way people used to eat, pretty much, before 1950.

In any case, when they've been on that diet for quite a while, they can have some coloring. They can have some of any of the additives, and they usually really are okay. We call it a washout. They tolerate a small amount, sometimes, for a while. But if they now think, I'm all cured, I can do this, and they begin eating at Taco Bell or something, which has BHT, TBHQ, and everything, and they think they're okay, after they've done this for a few weeks, they're not okay anymore.

So there is some tolerance that's built up. And these children, once they are off the diet -- remember, you're taking away thousands of additives when you put them on either the Feingold diet or an oligoantigenic diet, then you're picking one thing and testing them.

It may not be the one thing this child is sensitive to, or it may be that this child's only sensitive to Yellow number 5 when he's also getting MSG. And there has been some toxicological testing, in-vitro testing, but only one that I know of. And that was a Lau (ph) study in 2006, which showed that when the additives are used as an inhibitor of neurites, and they test MSG, Blue number 1, and I think it was aspartame and Yellow number 5, one of them, individually, they got one result. But when they put them together, it was a far more toxic result.

So, obviously, even -- well, maybe not obviously, but they haven't done the studies on human beings to see if when you get these additives in mixtures, if it's more toxic or not. But in the real

world, we're eating them in mixtures.

But I just wanted to add that. That may be why not all those children reacted. Many of them had been on the diet for quite a while. And you give them 10 milligrams, it doesn't do anything. Maybe if you gave them 100 milligrams -- remember, those tests that used 100 milligrams, 125, 150, had a much higher percentage of children responding.

DR. VOORHEES: Can I ask you a question?

MS. EDELKIND: Me?

DR. VOORHEES: Yes.

MS. EDELKIND: Yes.

DR. VOORHEES: So you mentioned the fact that some of these dyes have been eliminated in Europe. Is there any evidence that problem behaviors in children are declining in Europe?

DR. JACOBSON: They are no studies on that.

MS. EDELKIND: If you don't study it, you won't see it.

DR. JACOBSON: But you might ask Dr. Stevenson later today. And also, we talk about Europe getting rid of dyes. They never used dyes as

widely as are being used in the United States. They've always -- they've stuck with a more natural diet, suspicion of fabricated foods.

DR. VOORHEES: Thank you very much.

MS. EDELKIND: One of the things that they have done in the U.K., they did some studies on children with ADHD and with autism, and found that they were low in an enzyme called phenol-sulfur transferase. This enzyme is needed in the brain for cognitive function, but it's needed also to metabolize the exogenic food dyes, and salicylates also, and other things. So it's needed to metabolize things with sulfite.

If you have a hypothetical person who is low in this enzyme, and you put them on a diet that doesn't need the enzyme very much, like the Feingold diet, which is considered a low-phenol diet as far as needing the PST or the phenol-sulfur transferase -- and I'm not a biochemist. But if you have taken the stress off that system for a while, it makes sense that the child will have some kind of recovery, will be more tolerant. And if you take that child and you

test him, you may not get a result in these short-term tests.

DR. ACUFF: We have some additional questions piling up, so we'll move on for a little while. And we can come, Dr. Voorhees, if you have additional questions.

Dr. Gray?

DR. GRAY: Thank you. I have a question for Dr. Weiss. And before we start, I should acknowledge that I learned my neurotoxicology from Dr. Weiss at the University of Rochester just a few short years ago.

Bernie, I've got a question for you. All of us looking at this were struck by what Dr. Jacobson calls what appears to be confusing and inconsistent literature. And as I learned, as a toxicologist, looking for that dose response is something that I do a lot. And I guess I'm interested in how you personally think about this, because if I look at the FDA review, if I look at the review from the European Food Safety Authority, they tend to say, certain children, or some children, sort of almost implying

that there may be some kind of an idiosyncratic response. Yet, I was very struck by Dr. Edelkind's graph that showed if you put the trials that are available, organize them by dose, you actually get something resembling a dose-response relationship.

So I guess I'm trying to understand that as you look at this body of literature, these data, is this a situation that you think there is a dose response and we're just not testing it properly, or is it something where, because we're already selecting a group of kids that we think, in many cases, not all the cases, may already be sensitive, and then challenging them, we're actually dealing with a subpopulation that is somehow unique?

DR. WEISS: To tell the truth, George, I'm not sure. I haven't done the kind of analysis that she has done, which seems to imply a dose-response relationship, which also implies that the reason we see subgroups is because they're the ones who appear to be the most sensitive.

But either way, we're dealing with a situation in which some children are sensitive to

dyes at low levels prevailing in our food supply. And I think that's enough to provoke questions about whether or not the risks are worth the benefits. I think that's the ultimate challenge of this committee.

DR. GRAY: I'd like to follow up. Excuse me. I was also struck with your argument that there are kids who are responding at very low levels. And what I'm trying to understand is, can we identify those kids ahead of time, or are we in danger of getting ourselves into kind of a tautologous argument that says, the ones who respond are the ones who responded, even though randomness would suggest that some fraction of them will go up a little bit, some fraction will go down. And how do we avoid a situation where we simply say, those who go up, those are the ones we really need to pay attention to, even if it's not real. It's sort of the hypothesis testing side of trying to identify sensitive responders.

DR. WEISS: Those of us who do research on animals will purchase animals, say, from the Sprague

Dawley. They are not genetically uniform, but they're fairly homogeneous, unlike the human population. And then we subject them to an exposure, and we test the effects, and we find a wide distribution of response. That's biology. You can't have evolution without variation.

So we're dealing with a natural situation in which there is widespread variation and susceptibility or responsiveness in the human population. And it's possible that the variation is not uniform. It doesn't follow a Gaussian distribution. But on the other hand, we know that there are dose-response relationships possible. And any random population of rats from Sprague Dawley at Charles River, if you administer a substance, you will find this variation in response. So why is this different?

I mean, we just finished the Swan study with bisphenol A, looking at changes in weight gain, and in fact, there is a subgroup of animals, not chosen beforehand, that got fat. That's because bisphenol A is an obesogen. I don't know why. But it's the same

question that confronts us every time we do a toxicological study. In many ways, it's hardly any different.

DR. ACUFF: Ms. Lefferts?

MS. LEFFERTS: First, I'd like to thank all of you for very excellent presentations. They've been very illuminating, very helpful. And I'd like to ask your help with one thing that the committee will be grappling with, and that's the criteria for evaluating studies.

I've heard, in a number of your presentations, some suggestions of additional criteria that the committee should be taking into account when evaluating both positive and negative studies, if I'm interpreting your talks correctly, such as the size of the study, the dosages used in the study, the duration between when the challenge was given and when the effect was assessed. So I'd like to hear your comments on the criteria that the FDA has proposed for evaluating studies, if you are familiar with those.

I'm also thinking about, when I went to

graduate school, more than a few years ago, we learned about the seminal paper by Dr. Austin Bradford Hill, that was about environment and disease association or causation. And he puts forth a number of criteria, which he's very careful to say not all of them are required to reach a conclusion of causation. But I have always found them very helpful for myself in looking at a body of evidence and trying to weigh it to understand what it says.

Could you comment on those criteria, whether you think they would be helpful to the committee? As I reflect on them to some extent, in thinking about your presentations -- for example, one of his criteria is about experiment. When you remove the suspected agent so it's no longer being exposed, do you see an improvement in the disease or the consequence? And it sounds like from many of these studies, that has resulted when children have a cleaner diet, so to speak, they respond positively, as Dr. Edelkind has shown.

So I'd like to just get your reactions about the criteria, what's missing, what criteria we should

be using. And, also, perhaps Dr. Weiss would like to respond in more detail to some of the criticisms that the FDA has raised; for example, verification of double-blind, and others, of his study, and other studies in the database. Thank you.

DR. WEISS: Well, remember, the principles put out by Bradford Hill apply to epidemiological studies. And there is no perfect epidemiological study, which is why he set forth those principles.

The FDA responses puzzle me. It's as though they were saying, you have to conduct a GLP study in humans in order to verify this hypothesis, good laboratory practice studies. Well, you can't do a GLP study in humans unless you're an Arab dictator. It's absurd. You can't hold those studies captive to that criterion.

What the studies have shown is that a challenge of a blend of food colors, two cases, a single food color, can provoke responses that generally we would consider adverse.

So, again, is the evidence firm enough to lead you to believe that in some children at least,

at some dose, adverse effects on behavior are elicited? Take that information and weigh that against the benefits of food colors in the diet. I think that's your challenge.

MS. LEFFERTS: I also just wanted to draw attention. I thought that was a very interesting graph, Dr. Edelkind, that you presented, showing how the data points align, depending on dose. If either of the other speakers would like to address that, that seems a really important point for us to get at.

DR. WEISS: Yes. That should be published in the Scientific journal.

Go ahead, Shula.

MS. EDELKIND: What exactly is the question?

MS. LEFFERTS: I guess I wanted to hear from the other speakers about if they had -- what they thought of that. That was just very compelling, and I just wanted to get your reaction.

DR. JACOBSON: I'd really want to look through the body of all the studies and take a possibly more careful look at how those other studies fit on the curve.

MS. EDELKIND: On that particular chart, I put effort into finding only studies that were similar to each other, that first put the children on a fairly clean, as you put it, diet, so that they got better, took these children who had already improved, and then challenged them. There are other studies that do things differently, but they wouldn't have been appropriate on that chart.

So I was trying to compare apples to apples, and when I graphed them against the amount of dye I used in the challenges, it just came up that way. I mean, it wasn't engineered like that. That's the way it is. For some reason, which I really don't understand, nobody seems to have looked at it like that until now.

DR. ACUFF: Dr. Fernandez?

DR. FERNANDEZ: Hi. My name is Maria Luz Fernandez. I am in the Department of Nutrition, and I am very interested -- it's a totally different question -- on the mechanisms. For example, it is clear that some of the children respond to this colored dyes and others don't. And there some

evidence and some suggestion, at least brought up by you, that maybe regular children might have an effect.

But I am very interested in mechanisms. For example, we know, just to give you an example, the dietary cholesterol, not everybody responds the same, but we know exactly why; because people can suppress synthesis, or decrease absorption, or whatever.

So I would like to hear some comments about the mechanism. What is the mechanism that triggers that hyperactivity in children, in specific children?

DR. WEISS: There are just a handful of animal studies, a couple of which have indicted disorders of dopamine function. But other than that, we really don't know. We don't know anything about the molecular mechanisms by which these effects are evoked. I guess Dr. Stevenson has some data on possible genetic predispositions. But we don't know anything about the epigenetic possibilities now, which may be more important, even, than the genome itself.

Now, I agree with you that it would be

interesting to pursue the question of mechanism. But let's look at the NIH budget and tell me what the probability is that you would get funded to do such a study. It's hopeless.

DR. FERNANDEZ: Yes. It is hard, because I think the mechanism would clarify everything. Then we could be able to understand exactly what we are dealing with here.

DR. WEISS: We've dealt with this for a long time with cancer. There was always the promise that if we understood the mechanisms by which carcinogenesis occurs, we could set standards for exposure. We would know which chemicals in the environment were hazardous and which were not.

But in the meantime, we have set standards for exposure, based on things like a two-year bioassay because the pursuit of mechanisms is so illusive, the target keeps changing. But I'd love to have some money to do studies on mechanisms.

DR. JACOBSON: But I don't think we should wait for the mechanism before we decide what to do about food dyes. How many years intervened between

when British sailors were given Lymes and when the mechanism of their action was understood? It's, I think, an interesting scientific question.

Yes?

DR. FERNANDEZ: But do you have like a partial answer? What do you think is going on?

DR. JACOBSON: I'm going to leave that to Dr. Weiss and Dr. Stevenson, because I think it's really murky. It's really murky. But that shouldn't be a pre-condition for taking action.

DR. ACUFF: Dr. Burks?

DR. BURKS: Just to follow up on her comment, first, thank you for the presentations. I think it would help the, as you say, murkiness of the field to have a little bit better understanding of your thoughts about the mechanisms, so we can understand; maybe if we understood what a mechanism might be, why one study might be one better than another study. And not having even an idea about what that is, I think it still leaves the field pretty confusing.

Just for your comment, the second thought is

that in any scientific medical field, there are often people that have the same background, that have ask the same scientific question, but then they develop a hypothesis. And some people really believe the hypothesis and try to prove it's true, and others try to prove it's negative. So they may design their studies differently.

For the Feingold Association that came out of Dr. Feingold's studies almost 40 years ago, why hasn't, out of that group, there been studies that are really consistent in the results? Why haven't there been -- or consistent, the results come out that are even inconsistent out of people that are proponents of the hypothesis?

As an example, if you look at 20 different outcomes, statistically you're going to have one that's different. So every study has a different outcome that's positive. You say they're positive, but it's not the same one that's positive every time.

MS. EDELKIND: I do want to make one clarification. The Feingold Association is a volunteer-run organization. We're not an educational

– I guess we could be called educational, but we're not a scientific organization. We don't do research. We're down there in the trenches, helping the parents of the children of America.

Now, how many times it happens -- I mean, this is not once in a year; it is probably more than once in a day that somebody comes to us, after they have been to their doctors and to their specialists. They have been through medications. Their children have been suffering for years. And then, when they finally find out about this kind of a diet, which is really a very simple, and probably the cheapest intervention of all treatments for ADHD, and they put their child on the diet, in a week, he's a normal child, I mean, they are ecstatic.

But this is not science. Again, we're in the trenches. We're not running scientific experiments. It would be nice if somebody did that, though.

DR. WEISS: I always like to look at other parts of toxicology in relation to this question. Now, let's take lead. Lead is neurotoxic beyond

belief. At the time that Feingold wrote his book, the level of concern for lead in children was, what, 60 micrograms per deciliter? And then in 1979, when the seminal study by Needleman came out in the New England Journal, showing effects in children at much lower levels, based on assays of lead in teeth, the field changed. But even then, there was a lot of dispute about whether or not lead was really toxic at those levels. And now the level of concern is 10, and we know now that there really is no threshold you can calculate for lead exposure. That's how the field has changed.

I think for this question, we have to look at the history that we've benefitted from over the last 50 years. Remember, the 50th anniversary of the Society of Toxicology is this year. That was the theme of the meeting held earlier in March in Washington, and we've come a long way. And we've come a long way in two ways. One, we know a lot more about mechanisms, but, two, we understand that instead of looking at LD50s and organ pathology, we should be looking at much more subtle effects, subtle

effects like what we see here, behavior.

Behavior itself is quite a challenge to study, as Chip knows. And your responsibility here is to decide, again, whether or not, whatever risks have been uncovered are worth whatever benefits are provided by food colors.

DR. ACUFF: Okay. We have just a few more minutes, so we need to keep the answers as concise as possible.

Dr. Burks, follow-up?

DR. BURKS: To follow up to what you said, the risks that you've shown aren't consistent, so I don't see how that's a risk, unless it's a consistent risk, because the risk in each study is different; like for one, it's a parent's outcome; for another, it's a teacher's outcome. If it was really a risk, it'd be the same in all of them, wouldn't it? I mean, you have a cause and effect, and if it's the cause, then that cause should be consistent, and that's not what I hear you talking about.

DR. WEISS: We don't do that for cancer. We do a two-year bioassay, and we find a dose-response

relationship. We set a risk and dose at a risk assessment. We don't know the mechanism. We don't know the cause. But here, we really have a set, I think, of consistent data. That is, you have to weigh the evidence, which is what EPA does all the time. There are very few perfect studies, including mine. I've never done a perfect study. I look back and say, ah, this is what I should have done. It happens all the time. We all know that.

DR. ACUFF: Dr. Castellanos?

DR. CASTELLANOS: Thank you. I also want to commend you for putting this information before us with all of the flaws that are part of the literature. One of the novel pieces of information that you put before us is this graph that appears to show a relationship between dose, on the one hand, and response.

The question is, can we have more details about how that's put together? Because you've described briefly that you've gone through the literature and used certain criteria to decide which studies are appropriate, whether or not they had

enough of a washout period, et cetera. But were those explicit? Do you have a framework of studies you've accepted into this analysis and those which you have not?

In other words, what are the inclusion/exclusion criteria for the way in which you created this figure, first? Secondly, what are the response criteria that are used in the studies, and are those comparable, or did you simply use whatever the authors described in those studies, in which case that's also acceptable. It's important to know that.

I guess more importantly, what we have here is percent response, which could be 5 children or 50 children, and is there a numerical quantitative sense of whether or not this very interesting arrow or regression line really does have some numerical heft to it.

MS. EDELKIND: Well, the classic way of studying whether the food dyes affect children has been, from the beginning, to take them off of all of these additives, and then choose one, and see what happens. Sometimes, a child may not respond because

you're testing Yellow number 5, and maybe that child is only sensitive to Red number 40. So those things happen in the studies to complicate things. But it's the same pattern of the studies that I looked for, and I wasn't able to discuss it with the researchers. I was going by the actual studies themselves, and trying to make sure that they looked similar.

For example, there were a few studies, which Dr. Schab is going to be analyzing, but they were only studies of two children. Now, you're not going to get a percentage when there are only two children. You've got 0, 50, and 100. So I didn't include a study like that. If there was one child and it was an ABA, where he was his own control, that was obviously not appropriate.

DR. CASTELLANOS: I guess my question is, do you have the table of studies that you've included? I mean, we just have data points.

MS. EDELKIND: Yes.

DR. CASTELLANOS: And do you have the table of studies that you've not included? Because if we're going to really make sense of this, we have

to --

MS. EDELKIND: I have a table of studies that were not included?

DR. JACOBSON: Could you just provide it to the committee over lunch or something?

MS. EDELKIND: Yes.

DR. CASTELLANOS: You know, the study of two people that you didn't include, it's worth knowing that that study exists and you didn't include it. And you used decision rules, in other words, to make this. And it's important to have those be explicit if we're really going to evaluate this as an important contribution. It has that potential, but I can't tell whether or not this is meaningful unless I know what these four points on the right represent, and what other points might be part of that distribution that were not included.

DR. JACOBSON: Maybe it's a question for the chair. If Shula can prepare this over the next few hours or something, could you get it to the committee?

DR. ACUFF: Yes. If you can get us the

list, then we can distribute that.

DR. JACOBSON: It may be handwritten or whatever.

DR. ACUFF: Right.

MS. EDELKIND: Okay.

DR. ACUFF: Dr. Gray?

DR. GRAY: Thank you. Another brief question in our discussion of the Bradford Hill criteria triggered in my mind, and I'd just like to get your thoughts. What I often look for in a weight-of-evidence kind of evaluation is notion of corroboration. So that is, how often do I see the same result when I'm essentially trying to measure it in different ways.

One of the things that strikes me in this literature is real inconsistency in measurements, the situation in which parents may find an effect, but a clinical or a teacher evaluation doesn't, or some other combination of that. And I'm just interested in your thoughts of why it is that we don't see that kind of corroboration, even within a study, let alone across studies.

DR. WEISS: I think there are two reasons. One is that I think parents are better observers of their children's behavior in general than clinicians in a very restricted setting, and sometimes, in schools where the limits of the behavior are enforced.

There are a couple of studies -- I think the one by Williams in 1978, and another -- that showed corroboration between teacher ratings and parent ratings. I'm not sure about Kaplan. I agree with you that, in the best of circumstances, it would be desirable to have that kind of reliability built into the study. Sometimes, it's not possible.

Now, with one of the original studies from Wisconsin, there is a great deal of agreement between mothers and fathers rating the kids independently. But they were young children, and I think the scores were based on a modification of the Connors scale. But then the investigators concluded that their results might not apply because the Connors scale was not really built for young children. I think we need more -- I'd like to have more data. That's not the

question that confronts this committee, as I said before.

We're not going to have more studies unless industry goes out of its way to fund them, because they are not going to be funded by government agencies. So what you do now with the data you haven't had, to me, the weight of evidence tells me, that, yes, there is a significant risk of adverse behavioral effects evoked by exposure to artificial food dyes.

Now, for you, the question is, again, balance these risks, these adverse effects, against whatever benefits the food dyes confer on this population. Then it's almost like a clinical trial, isn't it? When I go before my institutional review board, do a clinical study, they want to know what are the risks, is there informed consent, do the benefits balance the risks? That's your question. You're the IRB.

DR. ACUFF: We have one final question from Dr. Voorhees.

DR. VOORHEES: Each of you has brought up,

in one way or another, the weight of evidence as a way of thinking about this. And a number of regulatory studies, agencies, and independent organizations have spent a lot of effort in developing systematic ways of going about determining weight of evidence. It's not an impression. What it is, is a very systematic review of a set of studies, and each study is looked at for what are considered to be the integrity of the study. And then studies are generally divided into those which are not considered sufficient basis on which to include in the weight-of-evidence analysis, those that are, and sometimes there are those that are put in the middle as maybe contributing.

So one of the struggles I see that we're going to have to tackle here is whether the studies that are being reviewed meet individual criteria of sufficiency to be included in a weight-of-evidence analysis, because some of these studies have some fairly significant limitations.

So in your review, have you gone through that sort of systematic process of making a

determination of which studies meet minimally sufficient standards for inclusion and which ones do not, in a formal weight-of-analysis type of evaluative process? Thank you.

DR. WEISS: I don't think the body of evidence is sufficient enough to do that kind of detailed weight-of-evidence analysis. It is sufficient, as I said before, to tell me that there are adverse behavioral effects incurred by exposure to artificial food dyes. It doesn't allow me to set an RFD, which is really what the WOE is meant to do. I could only give you a more general assessment of the weight of evidence, but there's a risk. You're the IRB.

DR. ACUFF: Thank you very much. We appreciate your time and responses, and we'll move on to the next speaker.

DR. JACOBSON: Thank you very much.

(Whereupon, a recess was taken.)

DR. ACUFF: Our next presentation is by Dr. Andrea Chronis-Tuscano, an overview of attention deficit and hyperactivity disorder.

I apologize if I messed your name up.

DR. CHRONIS-TUSCANO: Thank you. So my charge today is to provide an overview of the research on attention deficit hyperactivity disorder, or ADHD. And before I do that, I was asked to provide some background on the Maryland ADHD program. We conduct clinical research that advances our knowledge about evidence-based assessment and treatment of ADHD. We provide comprehensive evidence-based assessment and treatment of ADHD to children and their families in the community. We're a training clinic for the clinical psychology PhD program. And we educate parents, schools, health professionals, in the community about evidence-based assessment and treatment for ADHD.

I want to provide a brief overview of the talk today. I'm going to begin with providing a definition, a DSM-IV definition of ADHD and features. I'll move into ideological factors, then evidence-based assessment and treatment practices, and then I'll present the professional practice parameters for the treatment of ADHD.

First, the prevalence and impact. There's a prevalence rate of 6 to 10 percent, with ADHD being more prevalent in males than females. The male-to-female ratio is 3 to 1 in epidemiological samples and ranges from 3 to 1 to 9 to 1 in clinical samples. There may be some bias in terms of boys being referred more often than girls.

Fifty percent of children who are referred to mental health clinics are referred for ADHD-related problems. And the annual societal cost of illness for ADHD is estimated to be between 36 and $52 billion, which goes to 12 000 to $17,000 annually per individual.

The DSM-IV defines ADHD in terms of two constellations of symptoms, inattention symptoms and hyperactivity impulsivity symptoms. Inattention symptoms, of which at least six symptoms are required for a diagnosis, include failing to give close attention to details, or making careless mistakes in schoolwork, work, et cetera, difficulty sustaining attention, not seeming to listen when spoken to directly, failing to follow through on instructions,

or failing to finish schoolwork, chores, et cetera, difficulty organizing tasks and activities, avoiding tasks requiring sustained mental effort, losing things necessary for tasks or activities, being easily distracted by extraneous stimuli, and being forgetful in daily activities.

Within the constellation of hyperactive, impulsive symptoms, again, at least six symptoms are required. These symptoms include difficulty playing or engaging in activities quietly, being always on the go or acting as if driven by a motor, talking excessively, blurting out answers, having difficulty waiting in lines or waiting turn, interrupting or intruding on others, running or climbing about inappropriately, fidgeting with hands or feet or squirming in one's seat, and leaving seats in the classroom or other situations in which remaining seated is expected.

Other DSM criteria include that symptoms have been present prior to age 7, that there's clinically significant impairment in social or academic occupational functioning, and that symptoms

that cause impairment are present in at least two settings, which for most children include home and school. And finally, there's a requirement of differential diagnosis, that the symptoms are not due to another disorder.

There are three subtypes of ADHD. The first is the combined subtype, which is the most common subtype. And this indicates clinical levels of both inattention and hyperactivity impulsivity.

The predominantly inattentive subtype, which is the second most common, in which there are clinical levels of inattention only, and oftentimes, the predominantly inattentive subtype is not identified until middle school, when academic challenges and planning and organizational requirements increase. And it's been proposed that a sluggish cognitive tempo characterizes at least a subset of children with a predominantly inattentive subtype.

The least common subtype is the predominantly hyperactive impulsive subtype, in which clinical levels of hyperactivity or impulsivity only

are present. And this subtype is more common among very young children prior to school entry. In fact, research that's been conducted on preschool children with ADHD, as they age, have shown that most children with the predominantly hyperactive impulsive subtype eventually are diagnosed with the combined subtype.

There are many controversial issues with the DSM-IV criteria, which I'll discuss very briefly. First, that the criteria are developmentally insensitive, and particularly the symptoms of hyperactivity. Given that the symptoms are based on field trials conducted with elementary school-aged boys, we often have a difficult time applying these criteria to adolescents or adults with the disorder. Not many of us know adults who run about, or climb excessively, or get out of their seat in situations in which they need to remain seated.

Also, the DSM-IV criteria imply a categorical view of ADHD, whereas there are studies that exist that show that continuous levels of ADHD symptoms are associated with increased levels of impairment.

The requirement of onset prior to age 7 is viewed as somewhat arbitrary. The requirement of six months' duration is considered to be too brief. And the requirement that symptoms be demonstrated across two settings is unique to ADHD and doesn't characterize any other psychiatric condition in the DSM-IV.

A host of associated problems have been found in children with ADHD. First, we see peer problems. Often times, children with the inattentive symptoms are ignored in social situations, whereas children with hyperactive impulsive symptoms tend to be actively rejected. And this level of peer isolation is quite devastating for children and their parents.

It seems that children with ADHD aren't deficient in social reasoning or understanding, but rather, the execution of appropriate social behavior. Oftentimes, family dysfunction or parental issues are present. However, these factors are not causal. Rather, family problems such as parental psychopathology, maladaptive parenting, or negative

parent-child interactions can impact the severity and the developmental course or outcomes of ADHD.

Research with regard to self-esteem of individuals with ADHD has been mixed. In general, individuals with ADHD have inflated self-esteem. They tend to take credit for their successes and externalize blame for their failures. However, there are a subset of individuals with ADHD who have low self-esteem, and those tend to be the individuals who have co-morbid depression.

In most cases, ADHD is persistent across the lifespan. However, methodological issues within any given study impact exact estimates of persistence. A recent study, however, has suggested that baseline characteristics, including ADHD severity, psychiatric co-morbidity, and parental psychopathology predict the persistence of ADHD into adulthood.

In most cases, it does appear to be persistent across the lifespan.

In general, inattention remains stable across development, whereas hyperactivity declines with age. And, again, this may be an artifact of the

DSM-IV criteria, more accurately reflecting the manifestation of ADHD in young children.

Adult outcomes include many different forms of psychiatric co-morbidity, most notably conduct disorder and depression. When ADHD co-occurs with conduct disorder, we see very, very serious outcomes, which may include chronic criminality and serious substance abuse. And some recent evidence suggests that when ADHD co-occurs with depression, there's a high risk of suicidal ideation and suicide attempts. So this is a very serious disorder in many cases.

Ideological factors have been widely studied, and it's thought that there are multiple factors which may contribute to the development of ADHD, most of which are brain based. First, there's an average heritability of .8 to .85, with environmental factors thought to not be the cause. But they are or have been shown to contribute to the expression severity course, and co-morbid conditions that develop.

There is evidence suggesting dysfunction in the prefrontal lobes in individuals with ADHD, and

these areas of the brain are involved in inhibition and executive functions, which I will define in a few moments.

Genes which are involved in dopamine regulation have been implicated. In particular, the dopamine transporter gene, DAT1, and the seven repeat of the dopamine receptor gene, DRD4. However, genes by environment interactions have been found, whereby the presence of a gene in the presence of environmental adversity is specifically associated with more negative outcomes than either gene or environment alone.

Research has also suggested possible differences in the size of brain structures, including the prefrontal cortex, the corpus callosum, and the caudate nucleus. And finally, abnormal brain activation has been found among individuals with ADHD during attention and inhibition tasks.

In terms of brain structure and function, there have been differences in brain maturation, structure, and function, particularly abnormalities in frontostriatal circuitry, including the prefrontal

cortex, the basal ganglia, and the cerebellum. And these areas of the brain are associated with executive functioning abilities, so things like attention, spatial working memory, and short-term memory, as well as response inhibition and set shifting.

Neurotransmitter differences have been found, particularly in the levels of dopamine and norepinephrine, with less support for epinephrine and serotonin. And dopamine has been associated with approach- and pleasure-seeking behaviors, whereas norepinephrine plays a role in emotional and behavioral regulation.

I've alluded to executive functioning deficits. Executive functioning deficits are defined as the cognitive processes which activate, integrate, and manage other brain functions. Some examples in the cognitive domain are things like working memory, planning, use of organizational strategies, and the language domain includes things such as verbal fluency and communication; in the motor domain, response inhibition and motor coordination; in the

emotional domain, self-regulation of emotion and frustration tolerance. And all of these things have been implicated in the characterization of individuals with ADHD.

A few things to keep in mind is that some of these executive functioning deficits actually overlap with ADHD symptoms. Executive functioning deficits are not unique to ADHD, and not all children with ADHD have executive functioning deficits.

Barkley's theory is that ADHD is not a problem with knowing what to do. It is a problem of doing what you know. So as I mentioned in the peer domain, individuals with ADHD can often tell you how they're supposed to behave in social situations. Yet, when they're in the social situation, they act before they think. And so Barkley's theory assumes that behavioral disinhibition is the basis of the executive functioning deficits in ADHD. And, again, this is a performance rather than a knowledge deficit.

Mash and Wolfe propose a positive developmental pathway model for ADHD, which begins

with genetic risk, which may be activated by prenatal exposure to things like tobacco, et cetera, which may lead to disturbances in dopamine transmission, abnormalities in the frontal lobes and basal ganglia, which contributes to failure to adequately suppress inappropriate responses across situations, which may be associated with cognitive deficits and working memory, self-directed speech, self-regulation, which manifest as behavioral symptoms of inattention, hyperactivity, and impulsivity, which are then associated with impairments in social and academic development, which contribute to disruptions in parenting, and eventually we see co-morbidity in the form of oppositional and conduct disorder symptoms.

So this is just one possible developmental pathway, but I think the main point here is that ADHD is viewed as multiply determined with complex interactions between these various factors that have been shown to be risk factors for its development.

The next part of my talk is related to the evidence-based assessment and treatment of ADHD. Evidence-based assessment for ADHD necessarily

includes a multi-method, multi-informant approach, first, where teacher and parent completed questionnaires are gathered, with somewhat more emphasis placed on teacher ratings because teachers have a better sense of what's developmentally normative at a particular age. They have a wider sample of children with which they interact, and can therefore tell when one child's behavior is outside the norm for that age group.

Also included in an evidence-based assessment for ADHD is a structured clinical interview with parents, IQ achievement testing to screen for learning disabilities, which occur in about 50 percent of children with ADHD. And then, to the extent possible, it's recommended that clinicians do behavioral observations of children at home and at school. And each of these methods contributes to the ultimate diagnosis of ADHD.

It's important to note that no medical screen, cognitive test, or brain imaging technique can at this point be used diagnostically to detect ADHD. And also, the children with ADHD can indeed

focus long enough to watch TV, play video games, or to sit still at the pediatrician's office. So these are not necessarily useful bits of information in terms of making a diagnosis of ADHD.

Our well-established ADHD treatments include stimulant medications, behavioral interventions consisting of behavioral parent training, behavioral classroom management, and intensive summer treatment programs. And I'll talk about the criteria used to establish what well-established treatments are.

Medication with stimulants is the most well-researched, effective, and commonly-used pharmacological treatment for ADHD. And the most commonly used stimulants are methylphenidate and dextroamphetamine. These medications reduce ADHD symptoms by blocking the reuptake of norepinephrine and dopamine and facilitating their release in the synapse. This enhances norepinephrine and dopamine availability in brain regions, including the prefrontal cortex and the basal ganglia.

Stimulant medications have been shown to be highly effective in reducing ADHD symptoms in the

short term, such as including decreasing disruption in the classroom, increasing academic productivity and on-task behavior, and improving teacher ratings of behavior. Different formulations work best for different children, and common side effects include insomnia and decreased appetite.

More recently, Straterra or atomoxetine has been proposed as a non-stimulant alternative that works well for some children, but Straterra has not been studied as long or as intensively as the stimulants. And, in fact, smaller effect sizes relative to the stimulants have been found.

There are certainly some limitations to stimulant treatment; first, that there are individual differences in response, with approximately 20 percent of children not responding adequately to any of the various stimulant formulations. Very importantly, stimulant medication alone seems to have a limited impact on domains of functional impairment. And functional impairment is very important to consider because it is the primary reason for treatment seeking. Most people don't go to their

primary care doctor or a mental health provider because of specific symptoms, but rather because they're having trouble functioning in their day-to-day life. Stimulant treatment doesn't normalize behavior, so even on stimulant medication, children with ADHD still look different from their non-ADHD peers.

Other problems, such as family problems, extend beyond the scope of medication. No long-term effects of stimulant treatment have been found. Long-term use is rare. In fact, the modal number of stimulant refills that are filled range between one and two, even though this is a chronic disorder which requires lifelong treatment. And there is limited parent and teacher satisfaction with stimulant treatment alone. And finally, as we all know, some families are not willing to try medication for their children.

So the question is, how do we identify evidence-based non-pharmacological treatments for ADHD? The American Psychological Association has, for about the past 12 to 15 years, really been

focusing on identifying evidence-based non-pharmacological treatments for various disorders. And I was involved in the task force to identify evidence-based non-pharmacological treatments for ADHD.

Calling a treatment an evidence-based treatment implies that studies have been conducted with the following features; first, that there's been a very careful specification of the target population, which includes their membership in a particular diagnostic group, demographic characteristics, the manner in which they were recruited for the study, and specific inclusion or exclusion criteria describing who was and was not included in the study, whether they were preselected on the basis of any factor.

There must be random assignment to conditions. And in many cases, the experimental condition is compared to a wait-list condition. Ideally, the treatment is compared to a placebo, and under ideal circumstances, if an already-established treatment exists for a disorder, the best case

scenario is that your experimental treatment is compared to the already-established treatment.

There's also required the use of treatment manuals, and this is for psychosocial treatments, but this could apply to a variety of different types of treatment. Essentially, the goal here is to ensure reliability of administration of the treatment. And this type of knowledge being available in published studies facilitates replication of the study by other investigative groups.

Again, this emphasis on a multi-method approach to evaluating outcomes is required with raters being blind to treatment condition. There must be statistically significant differences between the experimental treatment and the comparison group post-treatment. And to go a bit further, in recent years, studies have required not only statistically significant differences, but also clinically significant differences. So we need not only to know that the experimental treatment is statistically better than no treatment, or to the alternative treatment, but also that that difference is

meaningful in a clinical way. So are these differences associated with differences, for example, in life functioning or in quality of life?

Then, finally, it is recommended strongly that replication by independent researchers be done. And of course, the weight of the evidence depends on the extent to which other studies, and particularly those conducted by other researchers or other research groups, replicates the finding of the study.

So as I said, there are two well-established, non-pharmacological treatments for ADHD, behavioral parent training, which has 33 well-conducted studies which meet the criteria that I presented on the prior slide. Behavioral classroom management has 45 well-conducted studies, and this was as of 2008. Behavioral treatments modify the environment such that rules and expectations are clear and that there are consistent consequences provided in the environment for both positive and negative behavior. And here, I present the various components of these behavioral treatments.

Some considerations in terms of the

implementation of behavioral treatments are, first, that we need to address the cross-situational impairments that define ADHD. In general, because ADHD is a deficit in performance, we often see poor generalization from the treatment setting to the real-world setting. And, therefore, it's recommended that treatments be implemented in all settings in which the child shows impairment, which means that, in terms of school behavior, oftentimes 504 plans are individual. Education plans, IEPs, are required for children with ADHD. It's also been recommended that academic interventions are needed in addition to behavioral interventions for children who have specific learning problems.

One major issue with behavioral treatments is that environmental contingencies must be delivered consistently in order for treatment to be effective, which is difficult to maintain. And, in particular, parental psychopathology such as parental ADHD or maternal depression, can interfere with the consistent implementation of these treatments.

I'll talk for a couple moments about the MTA

study, the multi-modal treatment study for ADHD, which is an NIMH-funded study across six sites, which was looking to evaluate the separate and combined effects of medication, intensive behavior therapy, and combined treatment, compared to treatment as usual in the community for seven- to nine-year-old children with ADHD, combined type.

The overall results of the study showed that all groups, regardless of what type of treatment they received, showed reductions in the ADHD symptoms over time. And ADHD symptoms were designated as the primary outcome in this study, although there's been a lot of controversy about the appropriateness of that designation.

But on the primary outcome measure of ADHD symptoms, medication alone and combined treatment did better than behavioral treatment alone and treatment as usual in the community. But on many measures, combined treatment was not significantly better than medication alone.

So one lesson we learn here is that depending on the outcome measure that you're looking

at, you can get a very different answer. Only combined treatment was better than treatment as usual on outcomes including oppositional symptoms, aggression, depression or anxiety symptoms, social skills, parent-child relationship, and reading achievement. And, finally, higher medication doses were needed in the medication-only group, relative to the combined treatment group.

Combined treatment was superior to the other treatments, in terms of parent and teacher satisfaction with treatment, normalization of child behavior, and improvements in functional outcomes, including family interactions, peer relationships, and academic functioning.

Now, the recently-published six- to eight-year follow-up of the MTA sample was quite sobering. This report concluded that the original treatment assignment was not associated with any of the 24 outcome measures six to eight years later. So there didn't seem to be any lasting effect of the original treatment assignment, and in the MTA study, treatment was administered for 14 months.

ADHD symptom trajectory in the first three years predicted 55 percent of the outcomes, whereby children with the best initial treatment response and the most favorable clinical presentation at baseline fared best over time. And these were children with behavioral and sociodemographic advantage, who had the best response to any treatment in the acute phase. These were the children that had the best long-term prognosis. So as a group, children with combined-type ADHD continued to exhibit significant impairment in adolescence, which suggests a need for sustained treatment over the long-term.

I'm going to conclude with presenting you with the professional practice parameters of the American Medical Association, the American Academy of Pediatrics, and the American Academy of Child and Adolescence Psychiatry.

The American Medical Association encourages the use of individualized, therapeutic approaches, which may include pharmacotherapy, psychoeducation, behavioral therapy, school-based and other environmental interventions in psychotherapy, as

indicated by clinical circumstances and family preferences.

The American Academy of Pediatrics guidelines say that the clinicians should recommend medication with the strength of evidence good and/or behavior therapy, strength of evidence fair, as appropriate, to improve target outcomes in children with ADHD.

Finally, the American Academy of Child and Adolescence Psychiatry states that treatment may consist of pharmacological and/or behavior therapy, but that pharmacological intervention for ADHD is more effective than a behavioral treatment alone, and that behavioral intervention alone might be recommended as an initial treatment if the patient's ADHD symptoms are mild with minimal impairment, or parents reject medication.

If the child has a robust response and shows normative functioning, then psychopharmacological treatment alone may be satisfactory. If the child does not show a real robust response to all FDA-approved medications, then clinicians should behavior

therapy and/or the use of medications not approved by the FDA for the treatment of ADHD.

I should note that these professional guidelines are based in part on research and in part on expert consensus. So you will see some inconsistencies between what the research I presented shows and what is here in the professional guidelines.

So now, for a summary of my presentation, first, ADHD is a highly prevalent, brain-based disorder, which is associated in most cases with lifelong impairment and functioning. Environmental factors may contribute to the expression severity course and co-morbid conditions that may develop. Long-term developmental outcomes for individuals with ADHD can include serious substance abuse, chronic criminality, depression, and suicide.

Stimulant medications and behavior therapy are currently the only established evidence-based treatments for ADHD. And combined behavioral and pharmacological treatment has the greatest impact on functional outcomes, is preferred by parents and

teachers, and is most likely to result in normalization of behavior.

DR. ACUFF: Thank you very much. Questions? We're going to start with Dr. Jones.

DR. JONES: Thank you. So you started out saying or noting a population prevalence of up to 10 percent. So in a disease that's not consistently persistent, and every controversy you noted would bias towards underestimation, would suggest that the true incidence is really high. Some might say improbable.

So how credible do you think that estimate is?

DR. CHRONIS-TUSCANO: The estimates have really ranged anywhere from 4 to 6 percent to 6 to 10 percent, and in the 1994 version of the DSM-IV, it was published as 4 to 6 percent. More recently, it has been increased to 6 to 10 percent.

The estimates are based on different methodologies or assessment practices, which may or may not include teacher ratings, and may include more or less comprehensive approaches to assessment. So

all of those factors, as you stated, can contribute to the prevalence. Well-conducted epidemiological studies have arrived at the 6 to 10 percent prevalence rate.

DR. ACUFF: Dr. Fenner-Crisp?

DR. FENNER-CRISP: Have the diagnostic criteria for the disorder been modified significantly since 1975?

DR. CHRONIS-TUSCANO: They have been modified, most recently in 1994. And actually, given the controversies, particularly with the developmental insensitivity, are expected to be modified in the DSM-V.

DR. FENNER-CRISP: That might affect selection of participants in clinical trials of those that are affected and potentially more susceptible or not, over time.

DR. ACUFF: Additional questions? Ms. Lefferts?

MS. LEFFERTS: I was interested in your comment that, if I understood you, for diagnosis, that the teacher evaluation is given more weight than

the parent evaluations. And as you heard already, something our committee's struggling with is when we have studies where the parental measure doesn't match up with the teacher measure, and what to make of that.

Could you comment on that?

DR. CHRONIS-TUSCANO: Sure. I want to begin by saying that it's not uncommon for parent and teacher ratings to not be very highly concordant. However, because of the way ADHD is defined, it's required that at least some symptoms be present in both settings, because if indeed ADHD is a brain-based disorder, it should manifest, at least to some extent, in more than one setting.

The reason that most clinicians place a bit more stock in the teacher ratings than the parent ratings is that many of the rating scales ask parents and teachers to indicate if any of the symptoms that I had listed out in my early slides are present just a little, pretty much, very much, not at all, just a little, pretty much, or very much, which requires substantial judgment.

Parents do indeed know their children better than anyone else, but they often have a smaller comparison group than teachers. So anyone who's observed a classroom full of children knows that it's not too difficult to point out children whose behavior appears aberrant compared to the other children in the class, or the other boys in the class, the other girls in the class.

So it is for that reason that most clinicians put more stock in teacher ratings. Although the teachers are observing a whole classroom rather than one individual child, their anchors for what's appropriate at a given age tend to be more solid.

MS. LEFFERTS: Just two other questions. A lot of the studies we're looking at go beyond ADHD, more generalized behavioral disturbances. Do you think that it's possible that there might be a spectrum where over here we've got a normal kid, over here we've got a clear ADHD kid, and maybe in the middle, we have some aberrant behaviors that might not reach the threshold of ADHD but seem to be a

little bit away from normal, sort of like with autism and Asperger's, et cetera?

DR. CHRONIS-TUSCANO: Right, being on a spectrum. And that was what I was referring to when I talked about the fact that one criticism of the DSM is that ADHD is viewed as either present or not, whereas most of these behaviors occur on a continuum, which can range from not being present at all to being present to a very extreme level.

So, really, the defining feature there is, is this child impaired? How are they functioning at home? Are they getting along with their parents? Do they get along with their classmates? How are they doing academically? And so for that reason, we place a lot of stock in functional impairment and measures of functional impairment, in addition to measures of symptoms. And as we see with the MTA study, you really get a very different answer in terms of how effective treatments are, when you look at various outcomes.

The other thing that I want to mention in response to your question is that oppositional

defiant disorder or conduct disorder co-occur in about 50 percent of children with ADHD. Behaviors, such as being actively defiant, being aggressive, saying no, those types of things, better characterize children with oppositional defiant disorder rather than ADHD, per se. So also temper tantrums, irritability, those types of things, may go along with ADHD, but those are actually symptoms of oppositional defiant disorder rather than ADHD, per se.

MS. LEFFERTS: Just a final question, if I may. You made the point that it's not like we can get a brain scan and say, yep, this kid's got this issue. It's not like getting a blood test, and yep, they're anemic. It's much more tricky. So I guess what I'm asking about is the instruments to measure, how much confidence we have in those instruments?

DR. CHRONIS-TUSCANO: Multiple instruments are certainly always advisable as opposed to one instrument alone. And because there is the issue of differential diagnosis, we as clinicians combine different sources of information, and in structured

diagnostic interviews, try to tease apart, is this truly a persistent pattern of behavior?

ADHD is defined as a persistent pattern of behavior. Most individuals, when they're depressed or anxious, may have trouble concentrating, may be a little bit restless or jittery, but if we're truly talking about ADHD, this is a problem that emerges early in development that is persistent across, in most cases, the lifespan.

So use of a simple rating scale is not considered to be diagnostic. However, the Connors is a well-validated instrument that's been used in many studies. But diagnostically, it's only one piece of evidence-based assessment.

DR. ACUFF: Dr. Voorhees?

DR. VOORHEES: I wanted to ask you about laboratory-based testing instruments and how useful those have proven to be, the CPT or other tests of those kind to diagnose ADHD. Did they show good, high correlation, not so good, or what?

DR. CHRONIS-TUSCANO: Well, the correlation may be good, and oftentimes, ADHD in non-ADHD groups

can be differentiated based on their CPT scores. But they're not used diagnostically because of the issues related to sensitivity and specificity. So you may see other diagnostic groups that show similar impairments on those measures, and for that reason, they're not used diagnostically. And they're not considered part of evidence-based assessment.

DR. ACUFF: Dr. Fernandez?

DR. FERNANDEZ: I was interested in the treatment for the ADHD disorder. You said that the approved treatment was medication, and in some cases, behavioral changes were also helpful. Now, what about diet? Have you ever tested diet, if it affects behavioral changes?

DR. CHRONIS-TUSCANO: At the time of these reviews in 1998 and 2008 -- and these are reviews that were initiated by the American Psychological Association in an effort to identify evidence-based, non-pharmacological treatments for various disorders -- the strength of the evidence for diets, or the studies that were included, didn't meet those criteria that I presented in my talk.

DR. FERNANDEZ: Thank you.

DR. ACUFF: Additional questions?

[No response.]

Okay. Thank you very much.

So we're going to break for lunch, and we'll reconvene at 1:50, 12:50.

(Whereupon, at 12:33 p.m., a lunch recess was taken.)

DR. ACUFF: Okay. I'd like to reconvene the committee. Our first speaker this afternoon is Dr. Jim Stevenson, and he'll be presenting information on the Southampton study.

So, Dr. Stevenson, whenever you're ready.

DR. STEVENSON: Mr. Chairman, thanks very much. Thank you to the FDA for inviting me to come over from England to make this presentation. I should make it clear that I'm talking in my own personal capacity as an academic at the University of Southampton, and although our research was funded by the Food Standards Agency, I'm not representing them. It's my own personal opinion.

We've heard a lot about ADHD, hyperactivity.

I just want to spend a minute clarifying what I see to be the relationship between the two, because the findings that I'm going to be emphasizing probably have a slightly different focus from much of the other research that's been discussed so far.

Hyperactivity, undefining, is a pattern of behavior shown by marked individual differences in the general population and comprises overactivity, impulsivity, and inattention. As you see immediately, I'm talking about individual differences. I'm talking here about interpretable and meaningful differences and individual differences in behavior. At the far extreme of this continuum are those children with ADHD, and we've had a very thorough presentation of the characteristics of that diagnostic group.

One thing that I will be coming back to at the end, which is I think very germane to the concerns that this committee is being faced with -- this last point is something I'm going to come back to at the end, which is that increase in hyperactivity is associated with later educational

difficulties and antisocial behavior, because I think the main issue here is what's the significance of the differences that may come out from a study such as ours; what's the real-world significance, functional significance for the children?

So here is a distribution, roughly normal distribution as I put it here. And I'll show you that the outcome measure that we use follows roughly normal distribution, and that children with ADHD are at the top end.

Now, one of the things I really wanted to emphasize is that this is not just a distributional characteristic. If you study the causation of variation in individual differences across the normal range and at the top end -- for example, if you do the genetic studies -- what you find is that factors that determine individual differences down here operate in a very similar way to those that determine the extreme condition.

So from an etiological point of view, and as a symptomatic point of view, we see ADHD as the high end of the continuum of individual variation. If you

want rather more biological evidence as well as genetics, there's a recent study that's shown it's in cortical thinning, so it's the same sort of property, that individual differences and variation in cortical thinning are related both to an extreme form of ADHD, but also to variations in behavior here.

I'm going to jump quite a way from dietary effects to begin with, because I wanted to put those into the context of what we know about other influences on hyperactivity. And there are twin studies. We've had a heritability above .8 quoted. I'm content to suggest it's about two-thirds, but it's a substantial influence that comes from genetic differences between children, both in terms of variation in this continuum and in terms of the determination of extreme group membership, ADHD.

As well as quantitative genetic studies, we've got molecular genetic studies, and again, we've heard a presentation that has referred to some of the findings on the dopamine system that have been linked.

But I also want to emphasize something about

other experiential and environmental influences on hyperactivity. It seems to me that this is a very common endpoint of any form of brain damage, or many forms of brain damage, so that if you look at prematurely born infants, for example, there's a substantial increased risk of ADHD in that group. We've studied children's hydrocephalus, for example. Again, a distinctive feature of children with that form of brain pathology is that the behavioral consequence of that is increased hyperactivity.

Also interestingly, and this is something which has become more recently investigated in some children's study in Romanian orphanages, but also some work in Britain that was done during the '70s, suggested that children that experienced adverse early events in particular institutional care, if they show a distinctive pattern or profile of behavior, it is in terms of elevated hyperactivity. And the Romanian adoption study has shown a similar pattern.

So there's a wide range of factors that act in concert to increase the degree of hyperactivity

shown by a child. Are food colors one of those factors?

I was unaware, really, of the scope of the meeting here, so I'm going to comment on one or two things that actually are going to be picked up later on. So I'll go through these things rather quickly.

The Feingold hypothesis, as you well know, was reviewed in 1983 with the conclusion from NIH there was no consistent evidence of effect. And David Schab, who's talking next, is going to present this meta-analysis.

So, again, I will not discuss this in any detail, but just to suggest that this revisiting of these early studies, using a rather better form of meta-analytic methodology, has produced a slightly different picture. They identified 15 separate double-blind placebo-controlled trials, looking at the effect of AFCs on subjects with a baseline diagnosis of hyperactivity, where it has been graded. So they're dealing with extreme cases. They're dealing with reviews of children that either have ADHD or elevated hyperactivity.

Again, I am not going to give you the detail of this -- I don't want to steal his thunder -- but the meta-analysis suggests that there was a significant effect of food additives. And the effect on the meta-analysis ranged from, say, .2 up to .28, depending on the quality of the study. Generally, the larger the study, the better quality of the study, the smaller the effect size. But even without exclusion criteria applied to the quality of the study, the effect size was still significant at about .2. But as I've already suggested, those are studies of children with a known condition, established condition, a diagnosis, extreme cases.

The work that we've been doing in Southampton has a different focus. It's been on children from the general population, and we've done two studies. Those of you who know a little bit about the geography of southern England may realize that the Isle of Wight is actually quite close to Southampton. It's only about a mile over the water from the mainland. And we did our first study on the island. It's actually a rather nice place to visit

and a good place to do a study.

This was the paper that was published in 2004, and I'm only going to briefly present the results because I'm convinced that the second study that we've conducted is a much more substantial and secure test of the question about whether additives add an influence to the hyperactive level of the general population.

The interesting thing I want to point out about this particular study was that we start off with a very large sample of children. We screened all children whose third birthdays came across a particular time period, living on the Isle of Wight. And we screened for high levels of hyperactivity, not easy with three-year-old children. These are three-year-old children, and we talked a little bit about the difficulties of measuring hyperactivity in young kids. But we use a modified version of a Weissbury-Peters (ph) instrument.

We also screen using skin-pricked testing for atopy. The question of mechanism has been raised previously this morning, and one of those possible

mechanisms might very well be an allergic one, mediated by IgE. And we wanted to test that possibility. So what we did was to screen the general population, and we're aiming to identify a 2x2 design, children high in hyperactivity, H+, children with atopy, children with hyperactivity and no atopy, and then lower hyperactivity and atopy, and then the other condition where they show neither.

The actual endpoints of this particular study was that these two factors that we had built into our design, neither of them was significant in being related to the extent to which a child reacted to the presence of additives in the diet. It wasn't more marked amongst children that were initially high-level hyperactivity. It wasn't more marked about those with atopy.

What we found from this study was that all measures at home, rated by parents, that food allergies were having a detrimental effect on activity and possibly inattention, too. But as with a previous clinical-based study by Shulte-Korne in Germany, we couldn't pick out these effects at

clinic.

Now remember, we're dealing with three-year-old children coming to a clinic, having a lot of attention given to them, being assessed by members of our team. And we suspect in a certain sense, these children are on their best behavior. We also think that the reliability of our clinic-based measures, with the children as young as this, is really quite difficult in a study where you're trying to pick up various weak signals. So on the basis of this report, the U.K. Food Standards Agency decided that these results needed replication and extension. And that's when we came back from the island onto the mainland.

Again, our interest was on taking samples from the general population. But now, in addition to studying three-year-old children, we also studied a group some five years older than that, eight- and nine-year-old children. This was a randomized, double-blind placebo-controlled cross-over trial; important to emphasize that in these studies, the children act as their own controls. They're exposed

to the different conditions, and you're looking at behavioral changes within a child.

Now, I think that's an optimal design, as long as certain design characteristics are in place, like you've got washout periods in place; there are no order effects, and we duly tested for those kinds of things. And we are content that the crossover trial methodology was appropriate.

I suppose another feature -- actually, I didn't emphasize it so much in the Isle of Wight study, but in the Southampton study, of the large number of children that we're studying, we're doing clinical research trials in general population samples in their own homes on quite large numbers of children. These are not easy studies to do, and are very demanding on resource. So I'll leave it at that.

What happened in the design of the study was that we looked at the children initially during a baseline, when they're on their usual diet. And then they were put on a modified diet for the rest of the time period. And actually, for the rest of the time

period, they were also receiving the placebo drink week by week. I'll come back to that. So we never have, in this study, a straight withdrawal period, and we can look at the effect of taking the additives out of the diet and seeing whether or how the behavior changes, because after baseline, on usual diet, they were on a placebo all the time. So we never have a placebo-free period.

So over the next six weeks, they were either on a placebo or an active mix, one of two. We measured hyperactivity using a range of outcome measures, parent or teacher ratings, observations of the child in the classroom, and on the CPT, which has again been mentioned earlier.

There was a critique written about our studies, which I only got on Friday evening. So I've really only had Monday to look at it. And one of the comments that was being made there was about the nature of the measurements that we were using, suggesting that the amount of observations we were using on children perhaps were not enough to pick up changes in behavior.

We used the classroom observation checklist, which was a coding, which was a thing that Howard Abikoff developed, and applied it and adapted it slightly for British context, where certainly with the younger children, the nature of the classroom activities is slightly different. And we consulted Howard about what we were doing, and he was content that what we did was satisfactory.

There was also a comment about the fact that we used a questionnaire that looked, initially, at six-month recall. Isn't that a long period of time to try and remember about children's behavior? Actually, that's one of the most well-used, reliable indicators of ADHD. And you saw this morning, if you're going to diagnose ADHD, you've got to look at it over a six-month period. So I'm content, the way in which we use these measures and the measures we've selected were appropriate.

We also -- and again, this is very important from our methodology. We aggregated these measures into a global hyperactivity aggregate, based upon parent/teacher ratings, observations in the

classroom, and the testing. Now, we specified that in our protocol. We specified that in our clinical trial registration. We specified that in our ethics approval. And there are a number of reasons why we specified it. One was to avoid outcome selection bias, which is universal in these kind of studies. It's been discussed this morning. You cherry-pick which measures you want. We weren't going to do that. We were going to define the measures we were going to take.

The second reason for using the GHA was one of avoiding multiple tests of significance by using an aggregate, rather than searching around for multiple tests to see where the significant differences lie, you avoid the problem of multiple comparison and type I error.

The third reason why we did this was to increase reliability. If you're trying to assess changes in hyperactive behavior in children, the more sources of information that you can get on doing this, the more reliable the measurements would be. That was the reason why we chose this measure, and

that was the reason why I'm going to be emphasizing this in the course of our discussion of the findings.

Here is that general global hyperactivity aggregate, just giving you an example from one of the samples. This is its distribution at baseline, and you can see that it's not perfectly normally distributed, but good enough, in terms of normal distribution, and appropriate for the kind of linear mixed modeling that we applied.

Here is the design of the study again in a bit more detail. Baseline week on normal diet, then the additives were withdrawn and a placebo given, placebo drink. And then each week hereafter, there was either an actual placebo and then a washout period, where again the placebo was continued to be given. We were very conscious of the need to try and avoid the placebo-driving thing, so that we tried to mask the placebo condition as far as we possibly could.

We tested the children on seven occasions. The testings were done during the course of the week, generally towards the middle end of the week, with

the children starting taking the drinks that we were giving them on a Saturday, and we were testing them towards the end of the subsequent week.

The randomization was 20 for each age group, into a mix of sequences like this, six sequences, 20 each. We actually slightly over-recruited, so our N was bigger than that in the end.

The mixes we used are here, and I know the focus of this meeting is on additives. So you might wonder why we put sodium benzoate. And indeed I think someone queried that this morning. The reason why sodium benzoate went in was that in the '80s and early '90s, a number of these challenge studies that you've already heard about, as well as giving colors, also gave benzoic acid or sodium benzoate as part of the challenge. And there was a history of interesting concern about the possibility of this being another food additive that might, alongside colors, affect behavior. That's why it was in.

I should say that the choice of the mixes for this study was determined by the Food Standards Agency. They wanted us, first of all, to replicate a

mix, a replicate, the challenge that we used on the Isle of Wight study. But they then also wanted to use a mix that had both a slightly different composition in terms of the colors that were included, but also included a slightly increased dosage to reflect for current exposures.

Having listened to the discussion this morning, I think you will see that the kinds of amounts of color that we were giving were the lower end of what was being described as the normal daily intake, so we were not using very small amounts, but equally we're not using implausibly large. And we had calibrated -– or the Food Standards Agency had calibrated mix B to look at a high average daily quantity in the U.K. for exposure to these particular additives.

I'm now going to present the results, and I'm going to talk about this slide in a little bit more detail because the other subsequent slides are of the same format.

The key measurements were in week 2, week 4, and week 6. That was when the children were having

either active or a placebo challenge that was following a washout period. And what these results on the left-hand side here represent are the averages. Actually, they're the estimated marginal means because they've been adjusted for various co-variants. But these are the average GHA, global hyperactivity aggregate, scores when the children were receiving mix A, mix B, or placebo. And remember, these are the same children.

On the right side -- and I think these are the results which I think I'll concentrate on -- these are the differences between active, mix A, against placebo and mix B against placebo. These are the mean differences and their 95 percent confidence intervals. And if it cuts its naught point, then it is not significant. And you can see here, for the three-year-olds -- and this is on an intention-to-treat basis, so these are all the children that were recruited into the study – the three-year-olds, mix A was having a significant effect. The mix B results are above zero. They're not quite significant. They cut that line.

Now, we replicated these analyses in three ways. We did it for the intention-to-treat basis for the entire sample. We did it for as-per-protocol sample, which were the children that took at least 85 percent of the drinks. And because of the use of the handling missing data, we also reanalyzed it on a complete case basis, just in case there was something in the way in which the missing data was being handled that was creating some kind of artifact.

So for both the three-year-olds and the eight-year-olds, I present you the results of each of these three analyses. For those children that complied to the protocol, again, we've got a picture of mix A being significant, mix B elevating the levels but not quite reaching significance. And when we use a complete case analysis, again, mix A was showing a significant effect. Mix B didn't quite -- it cut this point.

One thing to draw your attention to here is that mix A was the mixture of additives that we used on the Isle of Wight, again, with three-year-old children, so in a sense, this is replicating our

previous work. This is the mix that was being shown to have an effect, both in the Isle of Wight and in the Southampton studies.

The results in the older children presented in the same way, with the actual means up here and the differences here. You'll see that there is a difference between -- the three-year-olds, the results here are slightly below what they were at baseline. For the eight-year-olds, they're slightly above what they were on baseline, and this is driven by the CPT, where the children's performance over the course of the study, particularly between the initial testing and the first two weeks, showed a marked decline. They got fed up doing the CPT repeatedly, week after week, which if you've ever done the CPT, you would probably sympathize with them considerably.

What we have here for the whole sample of eight- and nine-year-olds is a significant effect, increasing the level of hyperactivity for mix B, but not for mix A.

If we go to the protocol analysis, you can see again we've got an even more significant effect

here, and the mix A not quite reaching significance. And when we use the complete case approach, then we now have both A and B being significant for the eight-year-olds.

Before I move onto this, one other point about the evaluation report was that they emphasized what they saw as an inconsistency within our results. I think the inconsistency only comes if you're driven by P values. If you want P values, then, sure, we haven't got everything being significant for all analyses. However, if you go for effect sizes, we have a quite consistent pattern, where in every case, the results indicate that the behavior is more adverse when on the additive compared to placebo, and the effect sizes -- well, they're not constant, but they are of a similar magnitude. So as far as the GHA is concerned, I'm quite content that we've got a good internal consistency in the pattern of results.

Coming back to this issue of the various components of the GHA, remember, it's parent-reported ratings. It was a CPT. It was the classroom observations. Again, the evaluation report

emphasized what it saw as inconsistencies in the data. I've already indicated that we saw the GHA as our primary outcome.

If the pattern of results had been around the other way, if we hadn't got significant results on our primary outcome measure, and then started to hunt around in the individual components of that for significant findings, we would have been crucified. People would say you're just sniffing around, trying to find significant results, scrambling around.

Actually, if you look at those individual results -- I've already emphasized the fact that I think they're less reliable than taking the aggregate. But if you look at those individual results, by far, I think the proportions are something like 18 out of 24 and 22 out of 24, in the direction of the additives having an adverse effect. It isn't as if we're flip-flopping around a nil result. These are all -– not all; say, 80 percent of them are producing effect sizes in the right direction. They're not all significant. But I think the pattern of results is, in my mind, the more

important thing that one should be looking at when you're looking at those individual indicators. I don't think that's special pleading. I think it's a very genuine feel, a very, very genuine view I have about the importance of recognizing the specifications of a primary outcome measure.

Now, to my great surprise, a week after we published our paper in the Lancet, the Food Standards Agency told me that all my raw data had to be made available to the European Food Standards Society, which I didn't know about, which was quite an interesting thing to be faced with. And what we did was to provide them with our raw data.

Again, one of the features that the evaluation report brought out was what they saw to be a kind of lack of robustness in our data and in our approach. Well, this data was given to the European Food Standards Agency. Steve Nissen and Kenwood, who are international experts in the use, in the analysis of cross-over trials, directed a reanalysis of our data. And what they concluded was that the broad -- the broad conclusions of their analysis were in

conformity with what we'd shown. There was one or two minor differences, something a bit more significant, something a bit less. But using an alternative statistical approach threw out exactly the same results, essentially, as we produced.

What the European Food Standards authority concluded was that we had shown a small and statistically significant effect on activity and attention in children selected from the general population, and that excluded children medicated from ADHD. They confirmed what I felt, that we had a robust, small effect of additives on behavior.

I'm going to come back to the question of whether that small effect is important or not, because that's something, obviously, which I think is important to this committee. Before I do that, though, I want to talk a little bit about mechanisms, because, again, that was alluded to this morning.

We didn't directly test mechanisms in this community-based clinical trial. But what we did do was to -- because our previous work on the Isle of Wight hadn't identified any social factor, family

factor, or indeed developmental factor in the history of the child that would predict whether the child was going to be a responder or not -- so the atopy didn't predict it; the hyperactivity level didn't predict it; the social class didn't predict it -- what we decided to do was to look at whether if we'd measured something about the genetics of the children, whether this would give us a handle on what might be moderating a child's response to the exposure to additives.

In order to do this, we had to identify what genes we were going to study. And as one of these dopamine genes, we included histamine n-methyltransferase gene because – and I'll come back to something that was said this morning. There was some evidence from my colleague, John Warner, who's an allergist, a pediatric allergist, that one of the mechanisms that might be mediating the impact of additives on children's health is histamine release. They've done some laboratory studies which had shown that. So for that reason, we brought in this histamine n-methyltransferase gene, and you'll see

there's a good job we did. We also had some other dopamine genes.

I'm not going to go through the full details of that paper; it's available to you. But what was striking was that it was this histamine gene, this histamine n-methyltransferase gene, that was moderating the impact of the additives. These are the children's mean scores under the A, B or placebo challenge. And you can see for this particular genotype, there's not much of a difference in their scores under those different challenges.

But for this particular genotype with this allele absent, you can see that there is a substantial difference between their scores on the mixes and on placebo. That was true for the three-year-olds, and I think interestingly and importantly, because it's often very difficult to get replication on these gene environment interactions, that we got a replication on this separate, independent sample of older children, where, again, here you can see that compared to the placebo, the scores are elevated here for this particular group, but not for that. We

didn't get the same sort of pattern results for the dopamine genes that we were also interested in.

So in terms of a mechanism, what we tentatively suggest is that histamine may mediate the effects, and the variation in genes influencing the action of histamine may explain some of the inconsistency between previous studies, because I don't think anybody has ever tried to look at genotype in this kind of way, in these kind of studies before.

I'm now going to come onto this issue of what I think are these effects important or not. And one of the issues here is to try and calibrate what you see as a small effect. But how small is small, and when does it become big? And so what I've done here in this slide is to go through the literature. It goes back to those factors that I suggested earlier, have been shown to elevate hyperactivity levels in children.

I just summarized. The effect size on a measure like our global hyperactivity index, produce an effect-size estimate.

Here's our study, with effect size just below .2. Here are the other additive studies, the high-quality one, which I said was just above .2 and the overall, which was getting off of .3. Those are small. They're the smallest effect sizes on this slide.

This is the effect of institutional care, in terms of comparing the remaining children that were in institutions with those that were adopted early. And you can see that experience elevates the level of hyperactivity by a percent of effect size, averaging at about .5, twice, probably over twice, the size of the effect of additives.

Birth weight, again something which has consistently been shown to be related to elevated levels of hyperactivity; a pattern of behavior, short of ADHD, but that pattern of hyperactivity. The impact of low birth weight compared to non-low birth weight, again, averages out at about .5. So the effect size of additives is lower, but it's not off the scale lower than those two very adverse experiences.

If we look at treatment effects -- and remember, I'm really only interested in shifting the scores in the general population in this particular analysis of our own studies -- you can see that the effect sizes that were being described, or the treatments that were being described this morning, are obviously much more substantial, .9, .7. But there you're dealing with changes in the hyperactivity score of children who are very elevated in hyperactivity already. You're trying to reduce their scores. What we're doing is pushing -- by putting the additives in, we're pushing the scores up. There's another way of looking at that.

Here, we have a population mean of zero. By definition, the way we calibrated our studies, our population mean for this general hyperactivity index, was zero. If you then put the additives into their diet, you're shifting the population mean by about .2.

If you look at the thing around the other way, if you take children who are ADHD, and their scores, say, are about two standard deviations above

the population mean, the Schab and Trinh analysis brings their score down by about .2, and the combined treatment in the MTA trial brought the score down by about .7.

But think about the differences in the size of those two populations. Here, you're dealing with -- we've disputed the prevalence figure this morning, but let's say 5 to 10 percent. Here, we're dealing with the total population. We are shifting that up by about a fifth of a standard deviation.

Does that shift mean very much? How important is that in terms of the outcomes for the children? We know, we've had it described, that ADHD is a lifelong condition with probably adverse effects on function through into adulthood. Do these differences in hyperactivity that I've been talking about, do these, in a sense, normal-range variations, do they matter?

A very interesting analysis was done by McGee of two Antipodean studies. I think one was in New Zealand and one was Australian, where in general population samples, epidemiologically ascertained

population samples, they measured hyperactivity score at ages between three to eight, and then looked at the odds ration of various adverse outcomes at ages 13 to 14. And as one might expect, those with the highest scores had the worst outcome.

But what you can see here also is that there is an elevated rate, and that even in the middle of this range here, these children with scores of 3 to 5 were significantly different from the reference category in their risk for delinquency and school difficulties. It wasn't so for substance abuse.

I'd like to read you that quote, "There was strong linear relationships between early hyperactivity and later adverse outcomes. Adjustment of other childhood variables suggested that early hyperactivity was associated with continuing school difficulties, problems with attention, and poor reading in adolescence." We're not talking about hyperactivity. We're talking about a linear relationship between the risk of those adverse outcomes and the level of hyperactivity being shown.

Okay. Let me conclude. I'm concentrating

on the McCann study because I think that was the one that was of greatest interest here, mixtures. And we can't from our work disaggregate the effect of the colors from that of sodium benzoate or the colors from each other. That was the charge that we had, was to look at the effect of these additives on children's behavior. And we suggest that the presence of these in the diet increase the average level in hyperactivity in both three-year-olds and eight-year-olds in the general population, not the same mix of both age groups, but increase the average level.

The interesting thing here is to observe that these effect sizes are somewhat similar to that found by Schab and Trinh in clinic samples, trying to bring the behavior back the other way, again, looking at the effect of colors in that particular case.

Although the results of the study suggest that some mixtures may affect the level of hyperactivity in children, from our data, you can't suggest that the removal of these additives in the diet will be a panacea for ADHD. There are just so

many other factors that are producing that high level of hyperactivity.

This was a joint effort between a large research team, and I just want to acknowledge the role that these various people played in the study. Thanks so much.

DR. ACUFF: Thank you very much, Dr. Stevenson. As the committee begins to accumulate questions, I'll ask one.

In both the Southampton or the 2007 McCann study, and the more recent 2010 study, was the placebo identical in both of those studies?

DR. STEVENSON: The 2010 American Journal of Psychiatry paper and the 2007 Lancet paper was actually the same – is the same study. It's a reanalysis of the same study.

DR. ACUFF: Can you tell us what the placebo was?

DR. STEVENSON: Do you want me to go and get it? I've got it on a paper on my desk.

DR. ACUFF: Yes. That'd be great, please.

[Pause.]

DR. STEVENSON: Yes. The placebo mix was fruit juices. There was a tropical fruit juice, a white grapefruit juice, red grapefruit juice, prune juice, black currant, beech root, which is a magic ingredient because that masked a lot of the color, cranberry, cherry, pear.

That was the mix. If you want the volumes, I can give you those.

DR. ACUFF: Okay. Thanks.

Dr. Burks?

DR. BURKS: Thank you for coming. It was an interesting presentation. If you go to slide 22, that is your outline of the McCann study, if you could, walk through that. The one question I had, on the bottom left is like the placebo, and then walk me through what that means.

DR. STEVENSON: This represents the sequences with which children had the challenges. So one sequence, in week 2, you got placebo, week 4, you got mix A, week 6, you got mix B. Remember, two, four, and six were the key weeks because they were the ones that followed a washout. And then PBA

represents a different sequence. That's placebo, B, then A; A, placebo, B, and so on. And there were 20 of each of those sequences.

DR. BURKS: So there's really only one placebo for two active treatments? Is that right?

DR. STEVENSON: Yes.

DR. BURKS: There's one placebo and two active treatments –

DR. STEVENSON: That's correct.

DR. BURKS: -- not two placebos and two actives?

DR. STEVENSON: No. And that placebo was the same.

DR. BURKS: Is there any order effect from that, that you have looked at that you could tell us about?

DR. STEVENSON: No. We checked for order effects because, obviously, it's very important when you're doing a cross-over trial to see whether that's the case, and there weren't order effects. There was a general effect of time. I think that was for the eight-year-olds, there was a general effect of time,

but it wasn't an effect of order.

The other thing -- and again, it was alluded to this morning -– was how long do these effects last for. And what we did was, for example, in week 4, when we'd already looked at whether the ingredient in week 4 was affecting behavior, did the exposure two weeks earlier have an effect? And it didn't. So the washout period was doing its job.

DR. BURKS: Then the tests that you did, say, at the end of week 6, on your active or placebo -– so on day 1, they start at whatever, A, B, or placebo, and then on day 7 is when you do your --

DR. STEVENSON: The testing time was spread. We started the exposure on a Saturday. We then went into the schools on Tuesdays to do the observations. We did the CPT towards the end of the week. And we got the parent/teacher ratings at the end of the week.

DR. BURKS: So some of those were after two or three days of either placebo A or B?

DR. STEVENSON: Yes.

DR. BURKS: Okay. Thanks.

DR. ACUFF: Dr. Winter?

DR. WINTER: I appreciate your effort to come out here and describe this. It certainly must have been quite an effort just to try to put this thing together and figure out what you wanted to test. It looks as though your conclusion is that the mixtures did increase the level of hyperactivity in the students that you studied.

Is that correct?

DR. STEVENSON: That's correct, yes.

DR. WINTER: Would you contend also that is an indication that these mixtures caused ADHD?

DR. STEVENSON: No. We really tried to be very careful about extrapolating from our results to ADHD. We did at one point, and the evaluation report picked us up on that, where we talked about the genotypes. That may be what we were identifying in the histamine gene. It may be relevant in thinking about future drug treatments for ADHD.

Now, I would defend doing that because I've already suggested is that from a genetic point of view, variation in the normal ranges of what we were

studying, the genes that are involved in that are the same as the genes that are implicated for extreme forms of ADHD. So I think that extension, that leap of imagination, if you like, was justified.

DR. WINTER: And one other question. We learned earlier from Dr. Chronis-Tuscano about the DSM characterization or criteria for ADHD. And that involved, as I understand it, at least two different settings, so school plus home, or other recreational activities or work, depending upon the populations you're looking at.

Do you agree that that's an appropriate way to identify something like ADHD?

DR. STEVENSON: Yes, I do. I think it's important for ADHD and also for individual differences in the normal range that I'm talking about here, because although in ADHD, there is this question of pervasiveness – you need to show it being pervasive -- a feature of hyperactivity is that it might be situationally specific; so one child might show it at home but not at school and vice versa.

So if you're trying to capture changes in

that, then using the kind of approach that we did, which was to measure them at school, at home, direct observation, and testing, we would argue, captures your maximum opportunity of picking up that effect. So, yes, I think it's very important.

DR. WINTER: Can I ask one more question real quickly on that?

You showed a comparison of different studies that showed the difference in responses, and I think you indicated that I think there's a Kreppner study and a Fabiano study that show that their response at the school was far greater than the response the parents noticed. Yet, in your study, and then the other study, the AM study, whatever that was, the parents' response was much greater.

Can you postulate as to what was going on there? What was different about those studies?

DR. STEVENSON: Let me get it up.

Those were different drugs. That was amphetamine, and that was --

DR. WINTER: Fabiano, to the left there.

DR. STEVENSON: Fabiano had a stronger

effect on teacher than it did on parent, but noticed that for the low birth weight, it was parent more than teacher. So it was alluded to this morning that you don't always necessarily get the same result, depending on who you ask for.

But the thing that I would caution against is I wouldn't want to over-interpret that as being a significant difference. It may be higher for one than the other, but that doesn't mean to say that they're different from each other, significantly so

DR. WINTER: Thank you.

DR. ACUFF: Mr. Waldrop?

MR. WALDROP: Thank you again for coming out and speaking here before the committee. We had some discussion this morning about a number or many of the studies being sort of on, not the general population but on ADHD-designated children. And your study showed a small, slight increase in terms of the general population.

Can you talk a little bit about -- or do you have any thoughts on sort of the implications of finding a small increase on a general population

versus increases that you might see in the ADHD population?

DR. STEVENSON: In terms of changing exposure to additives, I'm really quite struck by the fact that those two effect sizes are very similar to each other. So the effect on reducing ADHD in children is about the same as increasing it, if you put it into the diets of children in the general population.

It's going beyond the data, but I've generally taken the view, and I think I've emphasized it here, that I see ADHD as an end of a continuum. And that's not just in terms of symptom severity, but also in terms of causal factors; that the same cause of factors act on variation in the normal ranges as they do at the extreme end.

I think that kind of finding just confirms for me that that was an appropriate view to take; that the same causal factors are having a similar effect of extreme hyperactivity as they are in the general population. I wouldn't say that was true for everything, but at least it's consistent with that.

DR. ACUFF: Dr. Gray?

DR. GRAY: Thank you.

Just a quick question in something that was raised in the staff report here that I wanted to get your thoughts on, the question of the quality of the blinding. And the issue was raised about whether the blind was verified or not. And I was interested in what happened, whether it was verified, and if not, why not.

DR. STEVENSON: Thanks very much. I didn't have a chance to modify my slides to answer all the points that were made in that report, but that's absolutely key, I think; I think something which I strongly disagree with. But what we did was to test for blindness in quite a large sample of people not involved in the study, who were given the active mixes and placebos, and asked to just discriminate between them. And they couldn't do so. The critique was that we hadn't tested that blindness at home.

Now, I think there's a major logical difficulty in doing that, because if in the course of a trial, you ask parents what treatment do you think

your child is getting, that won't be just a reflection of what they see to be in the drinks themselves, but it would be reflecting the behavioral changes that they're observing in the children at the same time, so that there's a circularity there.

We, therefore, deliberately chose not to use that method to test for blindness. We thought it was much more appropriate to do that independently in the university rather than in the home setting.

DR. GRAY: Was that study published or looked at, then?

DR. STEVENSON: It's included in the Lancet report, I'm pretty sure.

DR. GRAY: Okay. I must have missed that. Thank you.

DR. STEVENSON: Let me just clarify that. If it's not in the Lancet report, it's in the facts report we wrote to the Food Standards Agency, which I think we made available to you.

DR. ACUFF: Dr. Voorhees?

DR. VOORHEES: Thank you.

So you had two mixtures for each age group.

Is that correct, mixture A and mixture B?

DR. STEVENSON: Yes.

DR. VOORHEES: And mixture B was the higher dose in each case. Now, the effect that you got was not really dose dependent.

DR. STEVENSON: No.

DR. VOORHEES: But was it your intent to try to find a dose-dependent outcome? Is that why you did two different dose levels? I mean, why did you include two different --

DR. STEVENSON: It very much was not to do a dose response. I would have been very interested to have done it, but it would have actually required a lot more investment that we could put into here.

The reason why we had these two mixes -- and you're right that mix B was both a higher dose, but it was also a different mix, so you can't compare like for like -- was that the Food Standards Agency wanted us to both have a more, kind of robust, a stronger effect that we were testing, because the effects that we had with the three-year-olds on the Isle of Wight were small. They wanted to maximize

the chance of getting an effect out, but they also wanted to use a mix that was more reflective of the kinds of additives and colors that children are currently exposed to.

For example, in the U.K. -- I don't know about the States -- tartrazine is very frequently used now. So that was in our mix A. They wanted it out from mix B and replaced with quinoline yellow, so that that mix was a more ecologically valid test of color exposure these days.

DR. VOORHEES: Why did you have a different level of the additives in the three-year-olds and the eight- and nine-year-olds?

DR. STEVENSON: Because we couldn't prescribe in relation to body weight. It would have just been impossible. But we wanted to make some adjustment for developmental changes so that we had a multiplication factor, which increased the dose for the eight-year-olds. I think it's roughly in proportion to body weight.

You will notice, again, the report query, why we didn't change the sodium benzoate levels; we

kept those constant across. And my recollection of that is that this was because of another complication, which was that we didn't want to have the concentrations of additives in the drinks we were giving, getting to a level where we were violating U.K. law about the concentration that you could have. It wasn't about the amount so much, but the concentration in the liquid.

So we kept that at a lower level, rather than pushing it up so that we weren't going to start to violate any kind of guidances about the concentrations of that exposure that children should have.

DR. VOORHEES: So the other thing I wanted to ask you is, so you had a preservative as part of your mixture. So how do you know whether the effect that you saw is not the result of the preservative, rather than the food colors?

DR. STEVENSON: No, I did. We can't differentiate the two.

DR. VOORHEES: Okay.

DR. ACUFF: Ms. Lefferts?

MS. LEFFERTS: Again, we very much appreciate your coming here and making this presentation. Your study has been criticized by some as saying the fact that effects weren't noted by both teachers and parents, but yet you used an aggregate index.

Am I understanding you correctly to say that you feel that is not a valid criticism because the aggregate index is a superior measure than either one by itself? That's one question.

DR. STEVENSON: Yes. Yes, it is. I think on general principle the more source of information that you might have about behavior, the more reliable the measure would be. I think you've had a debate in this country about the analysis of the MTA trial, where it was mentioned this morning that there was an emphasis on particular indicators for measuring effect.

When they analyzed their data along the similar lines that we used, of using these factor scores, but where they were putting together multiple indicators, they got a different pattern of results.

And indeed you've got a slightly stronger pattern of results. So I think methodologically this is a much more satisfactory way of approaching it. And I've also alluded to the fact that it avoids these problems of type I errors, where you're searching around 20 or 30 different indicators to pick up a significant effect. We prespecified what was going to be our primary outcome measure and analyzed that.

MS. LEFFERTS: And another criticism that has been made is the fact that mix A and mix B, in one group, we're seeing a statistically significant effect in one mix, and the other group, a different mix. It is a little puzzling for me to understand that, although I do note your point that even though the results didn't always reach statistical significance, there's a pattern there.

But could you talk a little bit about that criticism and why it might have been that the mixes had these different results in the different age groups?

DR. STEVENSON: I don't know enough about the biology of what's being looked at here in terms

of these different color mixes, to know whether one is going to be more potent than the other. You might think that mix B would have a bigger effect because of the higher dose, but it's got a different mix, so you can't tie it down.

Can I just flash through these things again? If I'm asked, what's the striking feature of our results, it's where these are, rather than where it's cutting the -- it's effect size. They're pretty consistent. I'm not claiming no significant results, but I'm saying the pattern of results seems to me to be much more consistent with the notion that both mixes are having an adverse effect on both age groups.

DR. ACUFF: Dr. Freeland-Graves?

DR. FREELAND-GRAVES: I like your presentation. It was excellent. I'm just a little confused about the difference in the effects on the two groups. They seem to go the opposite way. You have an effect size, but one's going down and one's going up in this.

Could you explain that better?

DR. STEVENSON: Yes. Sorry. The effect size over here, which is on the right-hand side of the graph, is always in the direction that the scores on the active mixes are higher than they are in the placebo mixes. They're all in the direction -- they're above zero, indicating an effect size which is the standard deviation differences between the scores on mix A and placebo, mix B and placebo, at a roundabout sort of point, .18 there.

Over here, these are the actual means. These are differences in means. These are the actual means. And what happened with the older children was that as we went from the first couple of weeks of the trial, their performance on the CPT plummeted, which meant that they overall were getting worse scores, which is why their scores in a sense have gone up a bit.

But over and above that, the interest is in the differences between their scores on these different weeks, and those are the ones that are significantly different; where there was a slight difference is on the three-year-olds, where again,

the effect sizes here show that the scores are worse on the active compared to placebo. But here, the scores are slightly below zero. They got slightly better when we introduced them to the trial and took the additives out of their diet.

Yes? The oldest children got slightly worse doing the trial, and that has elevated their scores. The other children got slightly better. But the importance in the analysis is the difference between those means, and that's what's presented over on the right-hand side.

DR. FREELAND-GRAVES: So the older children became significantly better? Did you just say that?

DR. STEVENSON: In the course of the trial, their scores on the attention measure, the CPT, got worse. They got bored. They got fed up. They didn't like doing it. So, in general, their scores got worse. But around that worsening, their scores when they were given the active mix were higher than they were on placebo. And all these three means are being subjected to that slight elevation above the baseline point.

Does that answer your query?

DR. BURKS: Can you go to the three-year-old on that?

DR. STEVENSON: Again, their scores on the active mixes are worse than they are on placebo, but in this case, they were slightly better than they were at baseline.

The other thing to mention, which I haven't emphasized so far, is that if you look at the standard errors, the standard errors for the younger children, they're wider than they are for the older. There was more variability in the younger children's responses than they were in the older kids.

DR. ACUFF: Dr. Jones?

DR. JONES: Tim Jones. So I guess, actually, this slide is a good example. You talked about relative differences when the GHA changed, and you talked about long-term effects. But in absolute rather than relative terms for a non-psychologist, I realize it's an aggregate score, but how different would a child look if their mean score was .15 difference? I mean, is it really subtle or is it

pretty obvious?

DR. STEVENSON: In terms of how they would behave as you were looking at them? You would notice it. Yes, you would notice it. And, indeed, one of the reviews that was made of our work suggested that these were, in a sense, clinically significant changes. They were noticeable.

What I've argued is that they are of some long-term significance, from that analysis there. This is the change, and it represents another way of calibrating it. If ADHD kids are two standard deviations away, we're shifting them about one-tenth of the distance to ADHD by putting the colors in.

DR. ACUFF: Dr. Fernandez?

DR. FERNANDEZ: That was one of my questions, so I just wanted an explanation of what the values meant, the clinical significance rather than the statistical significance, but I think you covered it. But I had another question.

When you were explaining to us the polymorphisms, in your study about the polymorphisms, you said that the histamine degradation in

polymorphisms, you found some interaction. Right? So can you tell us how much the polymorphism is present in the general population that might respond to that challenge?

DR. STEVENSON: Yes. I tried -- again, there was a discussion about the proportion of vulnerable children that might be in the population. And I went back and looked at my files over lunchtime. And actually, in our study, this notion of responder/non-responder really is on a continuum. It's roughly normally distributed. Some kids respond a lot to colors. Some children respond very little. There isn't a discontinuity in that distribution. There isn't a lump. There isn't, suddenly a group of children that are responding a lot.

So we don't really know what those vulnerable children are, but the genotyping starts perhaps to give you some handle on that. And if you look in the American Journal of Psychiatry paper, you'll see that we give the frequencies with which these genes were found -- these polymorphisms were found in our population. And they're not uncommon,

30 percent, 40 percent, 50 percent.

Now, equally, I'm not saying if you've got that particular genotype, you will be a responder. That's not the analysis that we did. We simply showed that the genotype moderated the effect. So it wasn't that if you had the genotype, you were a responder, but you were more likely to respond.

DR. FERNANDEZ: So of the various polymorphisms that you looked at, the only one where you found some kind of relationship was this one, the histamine?

DR. STEVENSON: Yes. There were two different polymorphisms in the histamine n-methyltransferase gene. And they came out for both the three-year-olds and the eight-year-olds. The dopamine ones didn't. I think the comp team may have -- I haven't gotten the details in my head, but I think the comp team may have done one moderation, but it seemed to be a one-off result.

DR. FERNANDEZ: So you think based on these results, you can say that maybe a mechanism is related to histamine release? Is that a hypothesis

you could come up with?

DR. STEVENSON: Yes. It's what we put in our paper. The hypothesis is that the colors are having an impact, a pharmacological impact, null IgE mediated histamine release as being the factor that triggers the biological consequences.

The point about that is that in terms of the distribution in the brain, histamine is very much represented in the brain areas, particularly the pre-frontal cortex that were being described as being implicated in ADHD earlier. So there's a biologically plausible story that's coming together. I mean, this is going beyond this. This is just one finding in one paper. But it certainly seems to be interesting and worth following up.

DR. FERNANDEZ: Yes, and just a final question. So the children in your 2007 study, in 2010, are the same?

DR. STEVENSON: Yes.

DR. FERNANDEZ: Okay. Thank you.

DR. ACUFF: Dr. Voorhees?

DR. VOORHEES: I just have a quick question.

In this slide, you talk about your concern about the .2 shift to the right. Would you be equally concerned about a .2 shift to the left? If you had run a study on food additives and seen a .2 shift in activity to the left, is that of equal concern to you?

DR. STEVENSON: An interesting question, because hypoactivity, underactivity, is not a very-well studied phenomenon. I've already suggested that there are individual differences in variation here. It may be. It may be that an extreme end, if we were going right down here, hyperactivity. And indeed hypoactivity has been characterized in certain children with various kinds of medical conditions, for example, that they would be very, very underactive. But generally, the significance of changes of around down there is really poorly understood.

Let's go here. What we have here is an increase in hyperactivity score. I kind of guess the mean -- I haven't got it on here, but the mean point on this distribution would probably be about here

somewhere. So there isn't a lot of gain for being made low down at that end, but there's clearly a risk of becoming high at this end. But as they say, it's a strong linear relationship. They weren't suggesting it was curvy linear, that it was becoming irrelevant at that bottom end.

DR. ACUFF: Dr. Vugia?

DR. VUGIA: Thank you, Dr. Stevenson. Can you go back to the slide showing the differences between placebo, and the mix A, mix B, one of those? That's fine, one of these.

Would you consider looking -- since the comparison on the right of mixture A versus placebo and B versus placebo, in most of these comparisons, whether among the three-year-olds or the eight- to nine-year-olds, above elevated in just matter of degree, and they follow the same trend.

Another way of looking at it is that it put more suspect -- and I understand already your contention that there are five components, four colors, and a preservative in each of these mixtures. But the fact that there are two colors that are

changed between the two mixtures, and the fact that these move up to either close to or being significantly different from placebo, would you even consider the possibility that it's the three remaining constant components of the other two colors, sunset yellow, carmoisine and sodium benzoate that are actually contributing to most of the changes that you're seeing, and that the other two colors, the changes just makes very little changes in what you're seeing here between the two graphs that you're showing?

DR. STEVENSON: I mean, it's certainly possible. I've already indicated I think there's too many variations, both in the composition and in the dosages to be confident about interpreting too widely the differences between the different treatments.

Just as a corollary of that, notice that in each case, we compare active against placebo. We're never comparing active against active. We deliberately did that as part of our protocol because we realized we were underpowered to try and detect the differences between the two active mixes. So it

was always against placebo that we were comparing.

But certainly, one can argue a lot of different ways about what might explain that variation. My view is that it's actually not the variation that's a lot of interest. The main thing is the actual pattern of results, which indicates that for both mixes, for both age groups, the children were worse off when they were exposed to the additives.

DR. VUGIA: Thank you.

DR. ACUFF: Dr. Gray?

DR. GRAY: I'd like to follow up on that very briefly, just because another way to think about that is if you look at the way that the mix A and mix B results across both age groups and moves, the effect sizes are quite similar, as you pointed out in every case, in spite of the fact that we have both a different mix and we have very different levels in there. And, in fact, perhaps the simplest explanation is that it's sodium benzoate, if it's anything, because that is the thing that is at a constant level, and it gives you a very similar

effect size in each one of these experiments.

Is that, again, a reasonable explanation; in fact, the colors aren't involved at all?

DR. STEVENSON: It's certainly possible. And indeed I can't remember where we wrote it, but somewhere we've written that it's an absolute priority to do a study of sodium benzoate by itself.

DR. ACUFF: Any further questions?

[No response.]

DR. ACUFF: Thank you very much, Dr. Stevenson.

AUDIENCE MEMBER: May I ask one from the back?

DR. ACUFF: Sorry. We don't allow questions from the back.

MS. JELETIC: The chairman has to recognize everybody.

DR. ACUFF: Our next speaker is Dr. David Schab, and he'll be speaking on meta-analysis of double-blind placebo-controlled study trials.

Dr. Schab, when you're ready.

DR. SCHAB: Thank you so much for having me

here today and giving me the opportunity to participate with such a range of experts here to talk about some of these important questions. The major important question that my colleague and I tried to answer in this meta-analysis is right there. You can recognize the question by the question mark.

Do artificial food colors promote hyperactivity in children with hyperactive syndromes? I am going to tell you a little bit about this. Some of you have, presumably maybe, looked through this already, and maybe hope you'll pick something up new or contribute in some other way with your questions and so on.

While I go through it, I'm also going to try to respond to some of the FDA interim report's questions and concerns about the meta-analysis. Let me just say what those are now because hopefully I'll summarize those concerns successfully. And by highlighting them now, I can probably address them more smoothly as we go along.

There were two, it seemed, two main concerns in the staff report, two main concerns that limited

the reviewer's confidence in the results of the trial. I hope to allay some of those concerns, but I'm certain I'll also probably trigger others. These concerns were about the difference in responses that parents, on the one hand, and teachers and clinicians on the other hand, noted in some of the outcomes. So that was one concern.

Another concern -- which is more a concern about how we conducted the actual study as opposed to a question of the underlying studies themselves, how we conducted the meta-analysis -- was if and how and why we applied these validity criteria to the studies.

So our main hypothesis was just what you see there, that artificial food colorings would be associated with symptoms of hyperactivity in children with hyperactive syndromes. There were these other three -- I called them subhypotheses in the paper. They were meant to be suggestive, and raise questions, and provoke questions, and research, and so on, and so forth, although now I see that these subhypotheses actually touch on a lot of important

issues.

I'll focus on the first one about the issue of parents on the one hand, and teachers and clinicians on the other hand, differing in the size of the response that they detected in these underlying trials. And we'll come back to some of the others, particularly also I think the second one about screening methods, because questions were raised this morning about what does it mean if -- is it sort of circular if we find that some kids who are thought to be responsive are then found to be responsive. It's a good question.

So let me move on here. Let me just say what they are. So those sort of three subhypotheses, again, where the parents and the teachers and clinicians differ. The second one was whether screening methods of any kind could identify responders. And the third was whether the rigor of the diagnosis, underlying diagnosis, or even just the presence of a diagnosis at all would predict anything about outcome.

Again, these are just meant to be

suggestive. And in a meta-analysis with only so many trials, you can't ask too many questions unless you adjust your statistical techniques.

So let's see here; so inclusion criteria. There was so much research done in those kind of heady, excited days in the '70s when people got really excited about the Feingold diet, and there were lots and lots of studies that couldn't really answer the kinds of questions that they hoped to answer. So I tried to make this, our inclusion criteria, in certain ways narrow and in other ways, broad.

The first -- I guess I should probably make that the second. Really, the most important criteria is that these trials had to isolate the effects of artificial food colorings. So there were lots and lots of studies in those days, on the Feingold diet, on challenges with groups of substances like Dr. Stevenson's group. Those kinds of studies would not fit into this meta-analysis.

So they also had to be double-blind, placebo-controlled. And for reasons I'll be happy to

talk about a little bit later, I had said to myself that they had to either be randomized, the trials, or they had to employ any of these kind of reversal designs that are often used in small-subject studies that I had come to learn about through reading about the trials of education interventions and so on.

The trials had to include children who were less than 18, and I had broad criteria for what kind of diagnosis would be acceptable for a kid to be enrolled in any of these trials. And for better or worse, those were quite wide. I maybe shouldn't say any rating scale there. The outcome measures had to be behavioral outcome measures, and we'll come to talk about that in a minute, and what that means. But if you have a handout, you should put a little carrot in there and write behavioral rating scales.

What that really meant, in fact, was that the scales had to -- that most of these were going to be some version of one of the Connors scales that were so widely employed at that time. I was going to exclude trials where the outcome measure or the only outcome measure was continuous performance task, or I

guess a test of visual attention would be another kind of continuous performance task, paired associate tests, any of these kinds of neuropsych tests. Maybe I regret having done that now, but that was what I had at the time as a medical student thought was important.

So let me go on here. So what did we do? Some of it -- a lot of it was very standard. I extracted data from -- first of all, I did a big surge. I think there was nine or ten databases. I used lots and lots of search terms, so on and so forth. In some cases, I wrote to authors, or called them, or tried to track them down if there was some evidence of an existing trial or insufficient data.

Then I graded those addresses, well, the second of these two concerns in the FDA report. And I'll talk more about this a little bit later, but I graded these trials according to some validity criteria. We'll come back to that in a little bit.

Then I had to -- remember, there was that second subhypothesis, if you will, in there about the question of whether -- or I guess it was the third,

but the presence of a diagnosis predicts anything. So in order to answer that question, I had to also collect studies that enrolled non-hyperactive kids. Dr. Weiss, his study, for example, enrolled non-hyperactive kids. And then I had to segregate those into two piles, the studies of children who had a hyperactive diagnosis and kids who did not. And then there were a few trials where there was a mixture of kids, some of whom were hyperactive and some of whom were not.

There's only two things I could do in those circumstances. Some of the publications included enough information that I could separate out the kids into the hyperactive column and into the non-hyperactive column, and some of the trials did not include enough information to do that, in which case, all the data I got, I crudely put into this kind of secondary analysis I did of trials of non-hyperactive or heterogeneously diagnosed kids.

Just while we're talking about diagnosis, I'm thinking of going back a few slides, about the question of underlying diagnosis. Again, for better

or worse, I had very broad criteria for what the authors of these studies called -- or what I accepted. I was willing to take anybody who was described as having certainly ADHD or ADD, but also hyperkinetic reaction of childhood, hyperkinesis. I don't think anybody actually did this, but if one of the trials had said minimal brain dysfunction, that would have been fine, too. These are all old terms that I guess we think more or less describe what we call today, ADHD.

So we talked about the segregation. I picked a random-effects model, really, for two reasons. The main one -- I think this question was brought up earlier today, I think by Dr. Burks, which was the question of, are we measuring a single effect when we're testing kids with different dyes, and different mixtures of dyes, and so on, and so forth. And the random-effects model is a kind of statistical model that says that imagines, for the interpreter of the data, that there are a number of different effects that are being agglomerated in one place.

Also, but not always, the random-effects

model can make for more conservative outcomes or more conservative confidence intervals, but that's not always true.

Then this is, I think, moderately important, which was, at that time when I did this, there was almost nothing published on how to do cross-over trials, or how to do meta-analyses of cross-over trials. And I called around and asked different statisticians, and meta-analysts, and people that had written textbooks, and so on, on meta-analysis. And I finally found this group at the Cochrane collaboration, who had not yet at that time, but ultimately did publish a meta-analysis on opioids for breathlessness. And they had worked out the statistics for measuring the variance for standardized mean difference in meta-analyses. So this is some boring statistical stuff, but it could be important.

We had to do some of the things that have to be done. In meta-analyses, what do you do when there's a missing -– for example, that R, that correlation coefficient that measures correlation

between individuals in different parts and individuals in different arms of a study. We had to impute that, and then there was also treatment of non-continuous data. We used some standard techniques for that. Then there are some tests for homogeneity, again, about the sort of combinability of the data. We did some sensitivity tests and also some tests for publication bias.

So this chart here, don't try to absorb it; I'll just go through it column by column. The first column is just a list of the trials. And the second column lists what I call these kind of diagnostic criteria, that these letters A through E are not really judgments on the trial by any means. They're just sort of my, at the time, sort of subjective sense -- I might redo it at this point, but my subjective sense of what would be a reasonable way to order the rigor of diagnosis.

So A meant that there was some mention of DSM-III, or the criteria met DSM-III or DSM-IV criteria, or DSM-II criteria with a Connors scale attached to it. What would be the lowest one on

there? I guess an E right there, Swanson and Kinsbourne's. So those kids had a clinical evaluation, and that was a B, and E meant that they were responsive to meds. I'm just listing those as examples of what all that means.

Then the third column is the subjects, that the number of subjects varied widely. So there was in the very middle there, the Mattes and Rachel Gittelman-Klein.

There's a study there on one subject, and another one worth mentioning might be there at the very bottom, the trial that has 54 subjects in it. You see that, 34 out of 54? So that would be an example of a trial that contributed patients to both the primary analysis and the secondary analysis. There were 34 patients -- 34 subjects that made it into this analysis, and I guess another 20 that made it into the secondary analysis.

The length of time varied widely. In the intervention, again, I haven't listed doses, and those varied very widely, but the M is mixture and the T is tartrazine. And so most of these studies

use some sort of mixture of dyes. A couple of them use tartrazine. There's one that used tartrazine or Yellow number 6, sunset yellow, and then the outcome. You see most of them are some kind of Connors scale, and I really privileged that over anything in this meta-analysis. So that's important, and we'll talk about that in a moment.

Down at the bottom, you can see there are a couple trials that use this row, behavioral research inventory. As far as I remember, at the time, when I looked that up, there was very little information about it, but subsequently there have been some studies, and so on, and maybe even factor analyses. I'm not sure of that particular scale.

Oh, one other thing I want to mention, that I'm going to get to again, is about the validity issues. So this is a mixture of subjective and objective application of these criteria, which is how it has to be. There were, I think, four or five criteria. Well, we'll get to them a little bit later. But, for example, that study by Rose at the very, very top wasn't randomized. And that got my

lowest grade, a C. And the study by Rapp, that study -- if I remember, the blinding was described poorly or it didn't sound convincing. And so that got a C as well. Most of them got a B, and we'll come back to that. But that usually meant that I wasn't particularly convinced about some feature.

This is a secondary analysis. Maybe let's come back to that if there's time, and questions, and so on. Again, these are either non-hyperactive kids or trials that have a mixture of kids. For example, that very, very top line and the very, very bottom line represent trials that included data from -– those are trials in which subjects were contributed to both the secondary analysis and the primary analysis, 18 of 30; here, the Connors one on the first line, 12, in the primary analysis, 12 patients.

So let's look here. So just to orient you, this is a slide showing the kind of overall effect, the kind of crude effect of combining all of these. So overall effect size down there is a little more than a quarter of a standard deviation. Everything to the right of that dotted line indicates kind of

the harmfulness of dyes. And everything to the left of that dotted lines means that, to some extent, kids got better on dyes.

Now, I want to point something out, which illustrates something about the way the study was conducted and also something about the way the study should have been conducted, and wasn't. It was actually a mistake I discovered this morning.

If you look two-thirds of the way down, there's a trial by Swanson and Kinsbourne, 1980. So if you read the Swanson and Kinsbourne study, it looks like there was a big effect. And then as you read further, based on this neuropsych test that they do, a paired associate test of some kind, some kind of lab test of executive function -- but because I was privileging –- I was not -– I excluded those because it was not a behavioral outcome. I might have done it differently. But I'm going to show you in a few moments a mistake I made related to that.

But this is just to show you that the choice of the outcome measure, as you might expect, makes a difference. They say in their text there that on the

Connors questionnaire, they found no effect. And that's how we came up with that standardized mean difference that circulates on the dotted line, despite the very positive, I think very positive, effect they found on that lab test. So I just wanted to point that out.

Let's go on here. So in this slide here, I've segregated outcome measures according to who made the outcome measure. And that first stratum up there is providers, healthcare providers, clinicians. The next stratum is of parents, and you can see that parents have quite a large effect size that they demonstrate there; and down at the bottom, the teachers.

Now, I discovered while I was preparing for this, this morning at about 5:00, there's a mistake here. The Swanson and Kinsbourne study on the fourth line down from the top, it looks like the way I had read the text in the study had been to give me the impression, based on my misreading, that the clinicians had administered that Connors scale. That's wrong. The next two or three sentences after

they say there was no effect found on the Connors scale, they say this is not uncommon, that there is a difference between what parents and clinicians sometimes find. So that's a mistake.

So as you look at this slide and think about it, imagine moving that effect from -- well, first, before you do that, just absorb this thing that comes up over and over again in the research, whether by meta-analysis or Dr. Stevenson's studies. It's all throughout the literature, this issue of parents finding a bigger effect. That holds here, and it still holds, but holds to a lesser effect if you just make a kind of -– if you just sort of eyeball or make a casual adjustment here, which would be to move the Swanson and Kinsbourne material from that top stratum into the middle, presumably.

I'm not running the numbers right here, but presumably, the effect of that would be to probably raise the providers' summary effect size a good bit, but I think it would probably lower the parents' effect size just a little, because it would be 1 of 14 different studies, whereas in the case of the

parents up there, or the healthcare providers, would be 1 of 4 studies.

So there would still be, I'm quite certain, this difference between what parents tend to discover and what everybody else seems to find. But it might be a little bit smaller. These differences here were not statistically significant, and they would be, certainly, not statistically significant after that adjustment.

And I apologize very, very much to the FDA interim report, but it involves an incredible amount of work, and the review of this paper was incredibly thorough and reflective and thoughtful. So to the degree that this mistake may have adjusted the writer's thoughts, I apologize.

So let me move on here. This is sort of a boring slide, but it's a lot of information on it. So the very first line is just a statistical representation of something we already saw in the graph, the summary effect size; in that second column, standardized mean difference. And then there's the stratification by raters. So you have to

imagine this adjustment of some kind.

Then the next sort of middle part of the page is an attempt to try to address this question that was one of the subhypothesis about whether stratification by inclusion criteria could demonstrate whether there was some easier way to figure out if kids were responsive than to subject them to an extremely difficult, complicated, expensive, blinded cross-over trial.

So in some of these studies, prior responsiveness of any kind was not even asked. It wasn't a question. It was take all comers. And some of them, either there was an open trial in the beginning before the blinded phase or there was just a parents' report. And in the third case here, there were trials that had -– the requirement to enter the randomized blinded phase required responsiveness to a Feingold diet.

Among these primary trials, the differences between those three strata were not significant. But it is worth reflecting that there was such a large effect in that middle stratum of prior responsiveness

by open trial or parental report; that that criteria would lead to such a large effect size is I think worth reflecting on.

So then, moving on again -- I'm sorry for the dullness of this chart -- we move on here to the sensitivity analysis. So the first thing I did is I said, well, I don't trust those two trials, that the one that wasn't randomized and also the study by Rapp, which had a very large effect size. And we can go back and look at the picture if anybody's curious, and I think it's in your handouts. The study by Rapp had a very large effect size, but there was something about the blinding that was -- the blinding sounded bad or unconvincing.

So I excluded those two, so I was left with 13 rather than 15 trials, and got a smaller effect size. However, the statistical test that would tell us whether it made sense to lump these remaining 13 trials together said, no, that doesn't work. So I just put that aside.

Then I said, well, another concern I have is that these two small trials that have these strong

effects, let's see what happens if we exclude those. And also in the random-effects model, smaller trials get weighted a little bit more heavily than the smallest of their size would suggest. And in a fixed-effects trial, the number of participants in a trial is a better predictor of how heavily they're weighted.

So I tried eliminating those. We got, again, a smaller effect size than that smaller number back up -- than that .283 at the top. We got .255. And then I thought, let's lump those two concerns together, and we ended up with a value of .210, about a fifth of a standard deviation.

Then there was one other question I just asked myself, which was that because I had to impute a value of the correlation coefficient in order to calculate the variance around the effect sizes in a group of the studies, I pretended that there was no more correlation than you'd find among two individuals who would be entered into a parallel-arm study. And we got, actually I guess to my surprise, a bigger effect there.

Finally, there are these secondary trials. You can see, again, it says that there are eight of them. And we found -- I am not really using the word "statistically significant" correctly, so I hope the statisticians and epidemiologists and biometricians in here excuse me, but we got a statistically non-significant effect among those trials in general.

However, when these trials were segregated by responsiveness or by responsiveness criterion, studies in which prior responsiveness by open trial parental report was required, found a positive effect. Again, there are lots of questions. There are too many questions being asked here for all the data, so don't just take all these things with a grain of salt.

This is a representation of those eight trials that were the secondary analysis trials of non-hyperactive or mixed kids. And, again, there was this bigger effect size among kids who were screened.

So then we did a sensitivity analysis we've already talked about. We also did a funnel plot, which we can look at in a moment. We calculated the

fail-safe end, which is one of multiple ways of evaluating a publication bias like the funnel plot. And this number here is the number of trials it would take to lower that standardized mean difference down to an effect size of .15. It doesn't say anything about what the confidence interval would be around those trials and so on, but for that standardized mean difference, effect size score, where it would land.

I already told you of exclusion of the smallest trials and inclusion of trials with the lowest validity score. We already talked about that. And now we'll talk about separate slides, the validity issue and the parent ratings issue.

Again, validity is the degree to which non-random error or systemic error gets introduced into an outcome. I use the Cochrane collaboration's approach or what they were publishing as their approach at the time. They listed what they thought was a reasonable way to describe studies. And I believe the particular measures that I graded them on were roughly corresponding to theirs; were they

randomized? Was it actually random?

What was allocation concealment like? That's often a place where trials –- or maybe sometimes it's done very well, but often in clinical trials, the allocation concealment, which is the process of getting subjects after it's been determined whether they're going to be in one arm or another of a trial, getting them into that arm before the intervention is made, getting them into that arm without anybody knowing what they're going to be getting. And it's only at that point that blinding becomes an issue. So blinding, again, would be the third. It was the third one. Attrition and whether attrition was similar in both ends, in both arms, is important, and then also the issue of selective reporting, which comes up all the time.

There was a question in the FDA staff report last week about -- a question about whether the use of the validity rating somehow influence how studies were weighted in the main meta-analysis. It didn't influence them at all. The trials were weighted by the inverse of the variance and nothing else.

However, in the sensitivity analysis, like I said, those trials that got my C grade were excluded. And once those questionable studies were excluded, the results did hold up. Though the results were smaller, they may be rested on a somewhat firmer foundation.

Then the next question that came up was this one about parents versus -- it sounds like a battle -- parents versus clinicians and teachers, and why their outcome numbers seemed so different. So first, let me re-bring in the effect of the mistake I made, which would have diminished it a little bit, but since these differences were not statistically significant in the first place, the problem would still persist, just like it did before.

I missed the talk about ADHD here this morning, so some of the things I may say here perhaps have already been addressed. But these are just possible reasons -- and I don't expect people to buy these, but they're just possible reasons why these differences come up, I think, in these trials that made it into this study, and also into some of the

other studies. One would be the issue of assessment, what time the assessment was done, particularly if the child was medicated.

Actually, before I get into that, let me just mention that the few trials -- it was unfortunately few, but the few trials that did test the blind in their study and reported doing so said that the blind held up. Those studies and clinical trials in general, unfortunately, frequently don't report the testing of the blind, if they even do it in the first place.

So the timely assessment, if the child's medicated, about half the trials said that children who are medicated were excluded. I don't remember if I even recorded it, how many of the trials actually, though, did include kids who were on medication. Then would be the issue of differential sensitivity or differing attunement to the different effects, whether parents and teachers are interested in or noticed different kinds of effects. And I understand -- certainly, Dr. Stevenson touched on this, and it sounds like maybe the presenter this

morning from the University of Maryland also did.

But this would, for example, perhaps explain why, in the example of the Rowe behavioral rating scale, that parents noticed such large effects. That scale emphasizes, as far as I understand, irritability over inattention. And, again, this issue of symptom cluster -- I'm going to show you a slide in a second, which may be superfluous. You may have already talked about enough of this, this issue of differential sensitivity symptoms on teachers and clinicians, and you should please stop me if it's too much.

Then there are issues about the nature of ADHD. Also, again, there was Dr. Weiss's scale in which parents actually picked what symptoms were important. So that would be another example of -- he found that -– well, we've talked enough about that.

But there's the nature of ADHD. And first of all, there's just simply diurnal variation in concentration levels. For example, the continuous performance test, like the test of visual attention,

it says don't -- the instruction manual says don't administer this test after 1:00 because there's so much inter-subject variation in concentration in the afternoon.

Not enough of these trials said how soon after administration of dyes the parents were rating them, and like has been said, we don't even know enough about the pharmacodynamics of dyes in the first place and when is the optimal time to assess effect.

Then there's this issue of is ADHD worse in the complex settings of home life compared to the laboratory? It's been touched on already today. It certainly does seem -- at least clinically, I can say that the more kinds of stimuli and the more number of demands that a child with –- or an adult -- when I say adults, I mean adult psychiatrist -- is under, the more likely the adult is going to have difficulty attending to tasks, and sticking to strategies, and so on.

Then finally, this seems like more or less the same point, that ADHD is probably better in

simple settings. In one of Dr. Biederman's papers, he says that in a structured setting, some of the differences between parents and teachers tend to disappear. Now, this slide, again, addresses this question -- and it may be superfluous because it sounds like this issue was addressed. Don't hesitate to interrupt me.

So this slide I will hope you see may address some of the doubt you have about parents and teachers having differing attunement to symptoms and concern with different symptoms. These three trials, which are listed in six columns, are found in a meta-analysis by Joseph Biederman's group. And the question is -- the name of the meta-analysis is How Informed Are Parents' Reports? And they're specifically wondering how informed are parents' reports in trials of long-acting stimulants compared to placebo? And their inclusion criteria required that there had to have been reports by both teachers and parents.

So there's the MTA study. There's the study by Wolraich, and there's the study of Dr. Biederman.

The first two interventions, the MTA study was thrice-daily methylphenidate. And in the Wolraich study, Concerta was used, and in Dr. Biederman's group study, some formulation of a long-acting amphetamine compound was used.

Let me skip to the fourth column. The fourth column is the ratio of effect sizes that parents and teachers notice on different symptom domains or during different times of the day. And, again, don't draw too many big conclusions from this, but this slide suggests, for example, that when we're talking -– and I'm going to pick the most extreme example just to make the point. Again, you don't have to buy any of this.

You go down to the third line. When it comes to inattention, in Wolraich's study, parents and teachers more or less agree on the effectiveness of this long-acting form of methylphenidate. But when it comes to the hyperactivity impulsivity group of symptoms, parents notice a much, much stronger effect of methylphenidate than the teachers do.

You see, actually, the opposite in the case

of this long-acting amphetamine compound; that in this case, the symptom domains are not divvied up, but the time of day is divvied up. But the important point I wanted to make is that, in this case, parents noticed half the effect that teachers do of the drug. So there seems to be some sort of ongoing variation of this question that's consuming everybody today, which is a really good one. It's not specific to the food dyes. It pervades the entirety of the literature on ADHD.

So in conclusion, it seemed from our primary analysis that administration of food dyes to kids with underlying h broadly-defined hyperactive syndromes does seem associated with further hyperactivity; again, broadly defined. And there's a question which I think, Dr. Stevenson talked about elegantly, which is the question of what is it? Even taking our more conservative number, the .210, which I represented to be the standard effect size of 13 trials that met our higher validity criteria, what is the meaning of an effect size of about a fifth of a standard deviation? And Dr. Stevenson made some nice

comparisons. So we could talk about that, and what that number means, and so on.

Then there were these conclusions of these subhypothesis. We talked about the issue of parents noticing a greater effect than teachers and clinicians. And we also talked a little bit about the issue about screening. So again, maybe I'm warning you too much, but don't make too much of comparisons, all these multiple questions that are being asked. But there was this suggestive finding that among hyperactive kids, screening didn't seem to make much of a difference, but that among non-hyperactive or heterogeneous populations, screening would make a difference. And I didn't get into this very much, but I could not find that either rigor or presence of diagnosis was particularly predictive or not predictive of outcome.

So that's it. Questions, and ideas, and so on

DR. ACUFF: Dr. Burks?

DR. BURKS: So David, that was really nice. What I understood when you were talking, you said you

did this as a medical student. Is that right?

DR. SCHAB: Yes.

DR. BURKS: It's an amazing amount of work to be able to do during that time. What got you interested in the area during medical school?

DR. SCHAB: Yes. I had taken a year off between my third and fourth year of medical school, and I was trying to decide -– I was at the School of Public Health at UC Berkeley for a year, you know, one of these kind of one-year masters in public health programs. And I was trying to find something that would kind of bridge -- help me make a decision about whether to go into pediatrics or psychiatry or whatever. And also, I was interested in food, and I was kind of excited about food and diet, the effects of diet, and so on, and so forth.

I don't remember exactly how I came on the whole issue of food dyes and the Feingold diet, but I do remember thinking, "Oh, this is cool." This is kind of a -- if you want to call it a discovery, a finding, or a suggestion that arose in San Francisco, because Dr. Feingold was an allergist at -- I can't

remember -- Kaiser San Francisco, I think.

So that was sort of the basic. That's kind of how I came to it. And I wanted to learn how to do meta-analysis. Thanks for asking.

DR. ACUFF: Dr. Voorhees?

DR. VOORHEES: I just wanted to ask you one question. In the figure that showed the average effect size, I'm just not clear. How did you arrive at the average effect size? Does you treat each study as an individual data point, or is it a weighted average, weighted by size of the sample?

DR. SCHAB: Sure. They're weighted not quite by the size. They're weighted by the inverse of a variance, which is mostly determined by the size.

Is that good?

DR. VOORHEES: Yes. That's what I wanted.

DR. ACUFF: Dr. Freeland-Graves?

DR. FREELAND-GRAVES: In your handout, you have a last slide that kind of summarizes.

DR. SCHAB: Oh, yeah, yeah, that's good. Thanks for bringing that up.

So this is -– yeah, great. I'm glad -- I didn't really talk about strengths and weaknesses of a study. Let me include this in a bigger discussion of that.

So I think on some of the strains for -- I think we did a very thorough search. We chose statistics that were appropriate to cross-over trials. There had been another meta-analysis of a kind of a broader question of what happens if you put kids on -- it was basically of all studies that had anything to do in any possible way with the Feingold diet by Kavale and Forness. I can't remember when that was published. I don't know exactly what statistical methods they used, but it didn't look to me like they had taken into account the statistical advantages of cross-over trials. And I thought our sensitivity analysis was useful and helpful.

Then, right, so some of the weaknesses. Let's start with this, which would be –- and I guess we could categorize weaknesses into weaknesses of the conduct of the study and then weaknesses of the underlying trials. A publication bias would be a

problem with the underlying studies.

So maybe I'll just walk up here. Oh, I lose the sound. So, again, the idea -- and this is a funnel plot. Ideally, just to orient you, the horizontal axis is the number of subjects evaluated by the raters. If all the studies that ever existed were published, what you would ideally like to see is kind of a perfect -- kind of isosceles triangle, demonstrating that large studies, which would tend to have less variance or less variability around them, that large studies would tend to hover around the same effect size, and that smaller studies would have a lot more variability about where they show up because, just by the nature of sampling, smaller studies have more variance and noise.

Oh, by the way, because of that error that I discovered this morning -- this is a mild effect here, but I guess one of those -- I guess one of those right here, this (inaudible – away from microphone).

It does look -- regardless of exactly where you -– kind of how you exactly -- a triangle fitting

on that page, it does look a little bit like there are small trials with zero or negative effects potentially missing from the lower left-hand corner.

(Inaudible – away from microphone) -- alternatively, you could draw kind of a triangle like this. But it does look like something's missing from the lower left-hand corner there.

Sorry if I hadn't touched on that.

Are there questions about the funnel plot? Because it can be a little disorienting to have a slide like this just thrown up if you're not ready for it.

DR. ACUFF: Dr. Winter, did you have questions?

DR. WINTER: So as I understand it, you published your review in 2004, your meta-analysis in 2004?

DR. SCHAB: Right. I was a resident. And my colleague -- originally it was my masters paper. And then I redid the search, and she helped me transform a masters paper into something that could actually be published. She had just published a

meta-analysis herself. Yes.

DR. WINTER: That may help explain, because it looks as though the vast majority, over 80 percent of the studies you cite were published 23 year or more prior to that.

DR. SCHAB: Right.

DR. WINTER: Then I think the most recent paper even cited in your work was 10 years old at the time of publication. So I'm just curious as to the significance and what that means. Is there other good data out there that's being excluded?

DR. SCHAB: I think the people here who did the search for the interim staff report could probably answer that better. There had been a meta-analysis, like I said, a number of years earlier. By my search, they had missed two studies that should have been included, and then there were two subsequent ones to their study.

I have not done the same kind of exhaustive review, and I would say, also, despite the incredible merits of that super-exhaustive FDA report, we looked at, I think, 10 databases or something. And I think

only two are specified in there. It may also be that years later now, there's more overlap between these databases. But I remember at the time, the European database, N-base and PubMed didn't have quite so much overlap as they may have now. I don't know.

So I have not redone this search. I don't envy anybody who does. It's exhausting to go into every bibliography and then look for things. But gosh, if somebody would do that, that would be great.

DR. ACUFF: Thank you very much. I appreciate your time.

DR. SCHAB: Great. Thank you so much. Thanks for having me.

DR. ACUFF: We will take a 15-minute break. So expect to be back in at 4:15 to finish up. Thank you.

(Whereupon, a recess was taken.)

DR. ACUFF: Our next speaker is Dr. Eugene Arnold. He's representing children and adults with attention deficit hyperactivity disorder, and that's what he'll be speaking about.

Dr. Arnold?

DR. ARNOLD: I'm Gene Arnold and I approved this message.

[Laughter.]

DR. ARNOLD: Being in Washington, maybe I should start with the political statement. But I've been well trained by the American Academy of Child and Adolescent Psychiatry to put disclosures at the beginning. The good news is, I don't have any financial interest in this issue. But the bad news is some complications.

First of all, I haven't done any original research on dyes, even though I've done a lot of research over the last 42 years on ADHD and its various treatments. So I'm just an educated consumer of the wonderful research that's been done by other people who presented here. All I know is what I read in the papers.

The second complication is that I was selected by FDA to represent CHADD, without consulting with CHADD. Fortunately, I have a pretty good relationship with the CHADD leadership, so I think we were able to work things about. But then, I

was told that rather than reviewing the scientific data, I'm supposed to present ADHD 101, which I think was already very well-presented by Andrea this morning.

So in situations like this, I resort to the medieval moral philosophy axiom, "In doubt, there is freedom. In dubitas, libertas." So I'm just going to tell you things that I think might be useful to you as I go along here.

So the plan is that I'll first present the CHADD position, then I'll speak for myself. And what I'll try to do is clarify relevant features I think of ADHD, as they relate to the issue of dyes and causation and things like that. And then I'll try to show that relevance to the dyes. Finally, I'll end with some interim conclusions of my own.

Now, I'm going to do what I excoriate my mentees for doing, namely read you a slide, partly because the print is smaller on this than the others, and partly because it's the only slide representing the CHADD position. And I feel that I need to make sure nothing is missed.

So first, there does not appear to be compelling scientific evidence that food dyes cause ADHD. Now, that relates to the diagnosis of ADHD, not ADHD symptoms in general, as they may occur in the population outside of diagnosable ADHD. There may be a small subset of children with a hypersensitivity to certain foods or food additives that result in an increase in activity level and/or inattention. These children may benefit from an illumination of these foods or food additives from their diets.

CHADD has no stand on the inclusion of food dyes in food consumed by the general population. We do not support or oppose the use of food dyes. CHADD looks to the FDA to protect the health and wellbeing of all of our children, including those with ADHD. And in doing so, we encourage the FDA to make decisions that are well-grounded in the available scientific evidence.

My summary of this would be that CHADD's mission is confined to those with diagnosed ADHD, to advocate for and improve their lives, and is not

wanting to get into the more broader area of public health. However, I have no compunctions about that myself.

So to clarify some things or flesh out some things that Andrea talked about this morning in regard to ADHD, it's a descriptive diagnosis. It doesn't say anything at all about cause. In fact, many different causes can feed into this as a final common pathway. But even the phenotype is not homogeneous. There can be different manifestations of it, taking different shapes that may not be easily recognizes as being in the same syndrome. There may be someone with a sluggish cognitive tempo who's inattentive and has absolutely no hyperactivity or impulsivity. There may be somebody else who's very hyperactive and impulsive, but has only minimal impairments of attention.

The reason that we can't separate these out into separate disorders is that there are many, many people that are in between there, so you can't tell where to make the cut to say this is a different disorder. There seems to be so many gradations of

it, that it's most convenient to lump it into the same disorder.

Now, Andrea had already talked about the five diagnostic criteria. I want to flesh out some issues surrounding those, concerning particularly, number 1, 3, and 4. First of all, the first criteria, symptom count and severity, implies that ADHD, like most mental disorders, and like many medical disorders, is dimensional. It's not a qualitative pathology; it's a quantitative pathology, like hypertension.

Everybody has some blood pressure. If we didn't, we'd be dead. But too much of it is a problem. And so the cardiologists have spent a lot of time trying to decide how much is too much, and that has changed over the years. And, in fact, I think they've very recently made some further changes, and I can't even remember the details of it. But I do know that they set one diastolic measure, and I think they've started to include systolic now, which in the past, was ignored, for saying, up to this level, it's normal. Then they have an area in

between that they say is a problem, but is not really hypertension yet. And then above a certain amount, it's hypertension.

It's the same sort of thing with ADHD. We have cut points that have been set. And I don't know whether you want to say that psychiatrists are like cardiologists or cardiologists are like psychiatrists, but we do the same thing. We kind of fiddle with and tweak, where do we make the cut point? Where is the threshold? So, for instance, is six the right symptom count, or would five be enough, or seven? Where do you set it?

Now, the really important thing is that no matter where we make that cut, no matter where we set that threshold, there are going to be people just below it and just above it. If you say the cut point for hypertension is 80 diastolic, there are going to be some people with 78 and some with 82. And those with 82, maybe you can get them to come down below that by relaxing, getting away from stress, cutting down on their salt intake, et cetera, and those at 78 might be pushed over the threshold by excess salt,

dietary ingredient like dyes, or stress, which seems to make everything worse, or obesity. So you don't need diagnosable hypertension to be harmed by stress, excess salt, and obesity because the closer you get to whatever the arbitrary cut point is, the more risk you're at.

So that's the dimensional issue in this. And that's related, of course, to the idea that there may be effects of food dyes outside of diagnosable ADHD.

The chronicity and pervasiveness criteria affect the assessment. It has to be a consistent pattern of behavior in more than one setting, and over time. So we can't really take one little fragment of observation and make conclusions from that, at least not in diagnosing. And one common mistake that used to be made by many physicians was they would see a youngster for 15 minutes and conclude that he's not hyperkinetic, the old term, or not hyperactive, or doesn't have ADHD because he sat there in the office, paid perfect attention to the physician, and didn't fidget or was restless, didn't

climb the bookshelves, et cetera. And they didn't believe the parents' report of what he was like at home or the teachers' report of what he was like in school.

So we know better than that now, that we have to believe the caregivers, that they observe over time in an ecologically valid setting, and that's where the diagnosis is made, by history, not by a small observational window.

So we have to depend on caregiver ratings or at least caregiver information. And there are FDA-approved indications based on only caregiver information, only on parent information. And sometime, one case -- not in ADHD but in another disorder. I know of one where a drug was approved by the FDA based on parent ratings, direct ratings, not even filtered through a clinician.

So although they're very subjective, the parent ratings and teacher ratings are the most valid. And I might mention, incidentally, that for the pre-post change for that effect size, parent ratings tend to be the largest because they have the

largest placebo effect. But if you take a placebo-controlled effect size, teacher ratings tend to be larger.

For example, in the MTA study which was cited by a previous presentation, the placebo-controlled effect size in the initial titration was .7 for parent ratings and .8 for teacher ratings. But in any event, we have to depend on those, and this is relevant to the issue of people wanting to pick apart the global hyperactivity rating used by the Southampton studies and saying, well, it was only in the parent ratings that you found this. The causes, of course, are multi-factorial. I think the point has been well-made already by previous presenters that this is not a disorder that you can pick one cause and say that's it.

Now, the cause that's best documented, and often misinterpreted, are the genetics and epigenetics, and as presented before, up to 80 percent heritable. This has been documented in twin studies, family studies, other genetic and molecular genetic studies, and so forth. So I think it's

pretty well-established that genetics are a major contributor as a cause of ADHD.

However, this is often misinterpreted as genetic determinism. There's only 20 percent due to the environment, so there's not much you can do about it. However, genes, as pointed out by somebody else, are only expressed through interaction with the environment.

So let's take an example, phenylketonuria, well-documented, 100 percent heritable disorder, inborn error of metabolism, can't metabolize phenylalanine. They poison themselves on the byproducts of phenylalanine. And what do we do about it? Well, we give babies a diaper test, and if they have this, if they've got this genetic effect, we take phenylalanine out of their diet and, presto, they have fairly normal lives, not perfect, but within the broad range of normal.

So is it 100 percent heritable or 100 percent genetic -- 100 percent environmental? Well, I think it's both. I think the denominator we have to think of is 200 percent, not 100 percent,

because if something is 80 percent heritable, that means that the environmental contribution is somewhere between 20 percent and 100 percent. It's at least 20 percent, but it might be as much as 100 percent.

So that's cause of hope. It's a little discouraging to scientists to think we've come up with all these findings and they don't mean as much as we thought they did, but it's very encouraging, from the viewpoint of a clinician, that there's hope that we can do things to help these people.

I need to point out that one of the possible explanations for this is that the genetic, the heritable component, is actually an inherited vulnerability to some environmental toxin or other factor. And we can see that very well with PKU. It's an inherited vulnerability to phenylalanine. There can be other chemicals, though, also, even in the diet, that might be in this.

Now, I'm going to have to resort to reading this slide also, and I apologize for the fact that it's so fine, but I didn't know how else to get it

all on the screen. This is the hydraulic parfait model. And what this amounts to, is there are different layers here of etiology that add up to enough of a problem that impairment spills out, and then you have symptoms of the disorder, and if you have enough symptoms, they push up, with pressure, the piston up that triggers a diagnosis.

Now, the things, the etiologies, that we have in here, genetics of course is at the bottom. It's the basis, the genetic predisposition. Then you can have brain damage, infectious sequelae, hypersensitivity, which could be either allergies or just a plain sensitivity like a toxic chemical irritating, stress, acute and chronic, toxins and other heavy metals, endocrine abnormalities like thyroid, nutritional deficiencies. And then comorbidity tends to add to the burden with all of this, whether it's anxiety, oppositional defiant disorder or a conduct disorder, depression, mood disorders. And that tends to increase the impairment and symptoms.

Let's take a look at some of these things in

a little bit more detail. Benjamin Pasamanick in the '50s, '60s and 70s, talked about reproductive casualty, by which he meant the prenatal, perinatal and immediately post-natal things, brain insults, that can result in what was then called minimal brain damage or minimal brain dysfunction, which by and large, maps pretty well onto what we now call ADHD. It went through many permutations in the name, hyperkinetic syndrome, hyperactive child syndrome, hyperkinetic reaction, and then attention deficit disorder, and finally, attention deficit hyperactivity disorder.

This has changed over time, as all of these etiologies have. We now have less problem with kernicterus because of the blue lights and rhoGAM, but we may have more problem with other things because of the increased rate of fetal salvage of infants who previously would have died, but who have some minimal brain dysfunction as residue of that physical stress. Infections and parasites have contributed their share. Von Economo's encephalitis was a big cause in the early part of the 20th

century, intrauterine rubella.

Some of these things now, with the advent of antibiotics and immunizations, are no longer as big a cause, but then, of course, new pathogens keep developing. There's an interesting interaction between ADHD and child abuse because battered children often have brain damage from head trauma, the shaken infant even, but even older children who may be struck or hit or otherwise abused. And on the other hand, a child with ADHD, particularly the hyperactive impulsive type, may invite abuse.

In fact, such children may be inspiration for the bumper stickers you may have seen, one of them, "Insanity is heredity. You get it from your children." And another one, "Grandchildren are God's reward for not killing your own children." As a grandfather of 10, I can vouch for the fact that it's much easier to tell your children how to raise your grandchildren than it was to raise your children in the first place.

So we've had an increased prevalence of ADHD over the years, along with an increase of some of the

possible causes that have been imputed by various people. The prevalence rate went from 3 to 5 percent, somewhat less than 5 percent back when I was first introduced to ADHD, back when it was amenable brain damage before 1970. And that rate was still maintained in the literature as late as the 1994 DSM. It's now estimated, at least in epidemiological studies, to be 10 to 12 percent. However, I need to qualify that, that some of those epidemiological studies do not use all the five criteria that Andrea presented, but rather, depended just on cut points on rating scales of symptoms.

However, I think the general agreement is that the prevalence rate is up in actuality. Part of it, of course, is more liberal diagnosis and increased recognition, but there does seem to be more of a problem for some of the things I've already talked about, and also some additional issues.

If, as an educator, I were trying to design an ideal educational setting for a child with ADHD, I would want a small class. I would want individualized instruction. I would want older kids,

who tend to have a settling influence on them, and who can mentor them and tutor them, younger kids available so the child with ADHD can feel some success in being able to teach somebody else -- teaching is a good way of learning -- and breaks in the schedule, like chores, cleaning the blackboard, carrying in wood for the stove, carrying out the ashes.

What am I describing? The one-room schoolhouse. A hundred years ago, most of the kids in the U.S. were educated in a one-room schoolhouse. One other advantage of that setting was close cooperation between home and school. If you got a licking at school, don't tell your parents because you'd get one at home, too. So my father attended one of those. My mother taught in one, not at the same time or the same school, in case you're wondering. But the problem was being naturally treated.

Or you could look at it another way. If you take away that support system, then you bring out the problem. You bring out the disorder. Parents

working outside the home; take away the one-to-one adult mentoring that is so valuable for all kids, but particularly for those with ADHD. They can often perform very well in a one-to-one relationship, particularly with an older person. In fact, they can be quite charming and interesting in that sort. I have a grandson with ADHD, and by himself, he's delightful. I really enjoy having him around and doing things with him.

So it was originally called the working-mother problem, but it's actually the working-parent problem. It just so happened that the mother was the last one to leave, which had the greatest effect because having none is a whole lot worse than having only one instead of two. So that could be another issue, the social breakdown with divorces.

Color TV is an interesting thing. You can make all kinds of chronological associations between things that have happened in the environment, in the technology, and society, and the rise of ADHD as a problem, and some people have actually posited the x-rays, from the soft x-rays from color TV. Another

issue is the couch-potato problem, lack of cerebellar exercise.

But then there's also the new chemicals in the environment that we didn't have 50 or 100 years ago. Now, lead, of course, has been around for a long time, but I mention it because it's sort of the prototype of environmental contaminant. And I think this has already been alluded to.

Interestingly, the subclinical levels result in a correlation with ADHD symptoms, IQ, and other things, reported by Needleman, that was of the same magnitude -- after corrected for social factors, which you don't need to correct for in a cross-over study like we've heard about. But after correcting for social factors, the correlation was about .17. And of course we've heard .18 for the food dyes. I just thought it was kind of an interesting coincidence.

There's accumulating evidence that insecticide residues in the food supply and in the environment can be associated with ADHD. I cited two studies here, and particularly the organophosphate

insecticides. One was picked up in 24-hour urines in children, twice the risk in those with high metabolites and insecticide residues, compared to those with low rates, low levels. And then if you take the serum of mothers while they're pregnant, and check their insecticide level, and then check their children at age 5, you find a correlation there.

Of course, other chemicals from industry, consumer products, construction, also show up, PCBs and polyfluoroalkyl levels. These were studies done by the National Institute of Environmental Health Sciences and other investigations in peer-reviewed journals.

Then of course we have the artificial food dyes, which I won't go through all of that, but I do note this graph, which looks like the national debt over time, is actually the rate of the consumption. And I want to point out -- it's kind of interesting that we have a colorless slide to represent the dyes here.

[Laughter.]

DR. ARNOLD: But the consumption from 1950

to 2010 almost quintupled, and it doubled from 1990 to 2010. And this level of intake is about the same as the very largest dose used in the Southampton studies.

Nutritional issues can also be a problem. Maybe it's not things like insecticides and dyes that are in the food supply that shouldn't be. Maybe there are things that aren't there that should be. There's been documentation that there are fewer minerals in vegetables, fruits, and so forth since intensive farming with fertilizer that only puts in the three big ones, nitrogen, phosphorus, and potassium. Deficiencies of iron, zinc, and magnesium have been reported in ADHD compared to controls.

Non-anemic iron deficiency has become more and more prevalent with our high-carb fast-food diet. There's not much iron in fries and chips, sugar, pastry. And there are several studies in the literature indicating that the iron levels, subclinical, not anemic, correlate with ADHD symptoms, and this should not be unexpected because iron is a co-factor for production of dopamine and

norepinephrine, which everybody agrees are deficient in ADHD. Again, some may be genetically more vulnerable to these nutritional deficiencies per poor absorption or food preferences or other things, aberrant metabolism even once it's absorbed.

Another big change in the diet has been the polyunsaturated fatty acid ratio of Omega-3 to Omega-6, which has changed from about 5 to 1 about a century ago, to 50 to 1 now. These two series compete for desaturase enzymes. The Omega-3 is particularly important for neurological function, normal brain development, deficiencies in babies, results in impaired visual attention, and there have been differences in polyunsaturated fatty acid profiles reported for ADHD compared to controls. And clinical trials suggest a medium benefit after three or four months of supplementation with Omega-3.

There's also an interaction of these nutritional issues with inter-uterine stress. We now have more babies exposed to nicotine and alcohol as some women take pride in proving they can drink and smoke as much as any man. The association of

maternal smoking, prenatally and postnatally, with ADHD is significant in the child. Of course, there could be a genetic link that mothers with ADHD tend to drink and smoke more, and they have children who have that genetic diathesis. That hasn't been teased apart yet.

The point I wanted to make is that exposing the youngster to these stresses in utero can result in an epigenetic phenomenon of the thrifty phenotype, which then, when it's exposed to an environment with normal nutrition, tends to get bigger and fatter. And on average, the MTA data showed that kids with ADHD are taller and heavier with higher BMIs compared to the local normative comparison group.

Here's an interesting interaction of a food dye with a nutrient. And I happen to know about this one because I have a special interest in zinc with ADHD, but I don't know how many other nutrients they might interact with, and I don't even know whether anybody's studied it. It certainly should be. But there were two experiments done in the U.K. by Ward, one a replication of the other. So it was very

impressive.

In the first one, there were only 10 hyperactive versus 10 controls, and he found lower serum, urine, hair, and nail zinc in the ones with ADHD compared to the normal controls who didn't have ADHD. And then he challenged them with tartrazine. These were kids with ADHD whose parents said that they had been responsive, reactive to food dyes. And he found that the serum in saliva zinc went down non-significantly. The urine zinc went up significantly. And the behavior deterioration correlated with the change in zinc.

He replicated this in 23 hyperactive kids with parent-reported reaction to dyes, and compared age and sex-matched controls again. They had lower serum zinc than the controls, and when he challenged them with 50 milligrams of dye, the serum zinc again went down and urine zinc went up, more so than in the controls. And, again, these were associated with behavioral changes. Of course one could argue that somehow the parents' expectations influenced the children to excrete more zinc. A hard-nosed critic

could resort to that, I think.

In fact, there are enough flaws in all of the studies that someone who has their mind made up to discredit the studies can easily find chinks in all of them to criticize. And I think that many of the criticisms are valid, and I'll just mention a few of them here.

One of them is that the diagnosis is often not made by DSM criteria. Now, that's not a criticism of the Southampton studies because they deliberately set out to study kids in the general population and did not require diagnosing for ADHD. Blinding is often partial. They start out with open challenges, then those who respond, then they go to placebo controlled, more or less.

For example, the zinc study that I just cited, the challenge there was not blind. So we could use better blinding. And this is not a criticism of the Southampton study. Again, it's very well-blinded. In fact, in addition to the controls that Jim Stevenson described, if you look in the first article, the Bateman (ph) article, they

actually did ask the parents to guess the order in which the challenges were given. It's a binary outcome. It's like flipping a coin, and they guessed right half the time and guessed wrong half the time, which is just what you would expect by chance on a binary outcome. So, to me, their blinding is excellent, outstanding.

But one thing I would criticize that study for is mixing the dyes, which we've already talked about, so you can't tell how much of each. But I'm going to have a little more to say about that in a minute, and also, mixing in the preservatives. And the reason this is important, to distinguish the preservative from the dye, is that the dyes have no economic or food safety value. They don't preserve foods, they have no nutritional value, and they don't prevent spoilage, whereas the preservatives do have economic and safety values.

So we need a different level of evidence. The little rule of thumb that I use is sex versus rude. Sex is safe, easy, cheap, and sensible. If it's sex, you don't need as much evidence for an

intervention to take action. If it's rude, risky, unrealistic, difficult, or expensive, then you better have a lot better of an evidence base to make a decision.

Now, the dyes, as I said, were mixed with other components, and I'm not going to go through this because I think it's been covered already by Shula. This is just to show that although there are many different dietary components involved in these elimination diets, and suspected, and to some extent, found to be a problem, colors and preservatives are most often found in that.

This is a study focused on food coloring. And I'm pointing out here the denominator problem. The two who were proven to be reactors by a double-blind cross-over -- should we say that's 25 percent of the eight who had the double-blind cross-over; or should we say that it's 15 percent of the 14 who claimed to have hyperactive behavior in response to an open challenge; or do we say that it's 2 percent of the 55 who were selected; or should we go to 220 for the denominator and say it's 1 percent?

Here's another illustration of the denominator problem, but the reason I'm showing this is to point out that 2 out of 20 normal, non-hyperactive controls clearly reacted to the challenge. That's 1 percent of the controls, kind of a premonition of the Southampton studies on the general population.

Now, coming to the Southampton studies, there were some puzzling results by age, which I think actually provides some insights, some hidden answers to some of the questions that have been asked earlier. Recall that the preschool group, going just by the significance levels, the preschool group reacted to mix A the same as in the first study. Now, that was a replication, very impressive. And it seems to establish pretty well that that effect occurred with that particular mix of food dyes and sodium benzoate. But mix B did not quite show the significant effect, and it had a substitution of quinoline yellow and allura AC instead of tartrazine and ponceau 4R. So that raises the question. Could we have here an inkling of a more specific effect of

particular dyes?

Then the older children reacted to mix B, not mix A, and I don't see this as any impugnment of the results because I think it offers a valuable insight. What was the difference between mix B and mix A? Well, it's not only the difference in the dyes used, but also dose.

Mix B was twice the dose of mix A, and when the data were reanalyzed, taking those with complete case data and good compliance with results, mix A did then become significant. So it seems like this may be showing a dosage effect.

Also from the Southampton study, we had the significant histamine gene results. Three out of the six genes that they checked for moderated the effect of food dyes, which to me is a very powerful argument that there's something biological going on there.

This is a different presentation of some data that Jim Stevenson showed. Again, these are the kids. This is a different gene from what he showed. And in this case, it's whether the C allele is present or absent. These have the C allele. And

this green triangle is the placebo result. And it looks like the C allele is protective against the deleterious effect of the food dye, whereas for those who had no C allele, this is the placebo. And you can see the much higher ratings of ADHD symptoms, the hyperactivity symptoms as he called them, on these two mixes, with mix B being stronger than mix A, as was shown in the general data.

Now, one of the important questions for me, when I saw this, was what percent of the population does this represent? And I think Jim misspoke himself. He said 30 percent. I think he was thinking of those who had this. This was actually 60 percent of his quasi-epidemiologic sample. So I think this applies to a good proportion of kids out there.

Another bit of physiological evidence is the brain topographical mapping. Here, the interpretation of EEGs, finding differences between the provoking food and not, there was blind interpretation, but the challenge itself was not blind. So it is conceivable that parents'

expectation could have influenced the children's EEG.

Other evidence; people are bigger and better than rats, but still we can learn something from our four-footed friends, effects on serotonin activity, corticoid activity, liver function test. And then there's one human study which fits in very well with the histamine gene result of the Southampton study, mast cell degranulation with tartrazine, with release of histamine.

So we're always safe in reviewing any problem and concluding that more research is needed. Most people won't argue with that. But in this case, I think we have some very specific issues that could be addressed. One is the recruitment of the population to be studied, the recruitment of the sample.

So is it a specialty clinic dealing in allergies or sensitivities, or is it a general mental health clinic, a pediatric clinic? Where are you getting it? Are you advertising? Are you just going out in the schools, as the Southampton study? All of those are important issues to address the denominator

problem.

Another is a careful diagnosis by DSM criteria, in all the studies, would be very useful. Some of the questions that were asked here today could have been answered better and more clearly if the Southampton designation of ADHD had been a little better done.

We need to unbundle the things that are being tested, particularly unbundling dyes separate from the preservatives for the reasons I mentioned before, but also specific dyes, as well as mixes. And we have to keep in mind that sometimes, just as there can be drug-drug interactions, and there can be drug-dye interactions, and nutrient-dye interactions, there can be dye-dye interactions, so that it is important to study the mixes also, but it would be nice to unbundle a little bit and get some sort of notion of which is the most dangerous.

Then, of course, the age effects. Most of the studies so far that show an impressive effect have been in younger children, preschoolers and young, school-age children. What is the effect in

adolescents and adults? Do they outgrow this problem or is it still a problem? So the dose effect is a huge issue. As Paracelsus said, "The dose alone makes the poison."

How much is too much? And also, how much is ingested by the highest-ingesting children? I was interested to hear up to 400 milligrams a day by the 10 percent who ingest the most. I hadn't heard that before and I was amazed. But we need more data like that. We need careful, double-blinding, so that the investigator does not telegraph the expectation to the parents, teachers, child, or whatever.

We need to examine the interaction with nutrients, the interaction with drugs, and the effect on the whole classroom when all of the children are given a challenge, as well as the individual children, because not only do we have interactions of drugs, dyes, nutrients, but there's interactions of kids who escalate each other and interfere with their learning.

So the interim working conclusions from this. I agree that food dyes are not the main cause

of ADHD. I think that's been well-demonstrated. But they may contribute significantly to some cases, and in some cases, may push someone over the diagnostic threshold in this dimensional disorder. There are several threads of biological mechanisms that support the idea that there's something going on biologically here. And one of the things that I think has been pretty well-demonstrated is it's not an immune-mediated reaction.

We also have excellent evidence now that the deleterious effect is not confined to ADHD. It's a general effect, and in fact it's more of a public health problem than an ADHD problem. The small deleterious effect, regardless of diagnosis, was replicated and a possible mechanism identified. The magnitude of the reported effect is reminiscent of subclinical lead poisoning. The per-capita consumption has quadrupled in the last 50 years, and there may be a possible effect on classroom climate in the general population, from most children in the classroom deteriorating slightly.

DR. ACUFF: Thank you very much, Dr. Arnold.

Questions from the committee? Ms. Lefferts?

MS. LEFFERTS: Could you go back to that table that looked at consumption increasing from 1950? What was that? That was per capita?

DR. ARNOLD: Yes. These are FDA figures on I think it's the available dye, based on the poundage that was certified. I'm not positive. Somebody from the FDA can correct me if that's wrong. And this increase is from about 12 milligrams per capita, per day -- maybe that's 13 -- in 1950 to 60 some now.

MS. LEFFERTS: I'm just trying to look at the numbers and grasp these numbers in relation to the studies, and thinking about the information that was presented this morning by -- I think it was Dr. Edelkind and Dr. Jacobson, that one cupcake can have just about that high bar, just one cupcake. So, obviously, this is per capita and not what a child might be eating as a high consumer.

DR. ARNOLD: Yes. If you graphed a curve on this, it would be going like that. And of course, this is 2011 now.

MS. LEFFERTS: Right. But I guess what I'm

saying is that per-capita estimates are masking what children who are at the higher end of exposure might be exposed to, given that earlier data we saw. And that could be very significant, and also significant in terms of looking at these studies, which tend to be lower doses than that, and therefore could be missing something?

DR. ARNOLD: An excellent point. And I'm reminded of the old joke about the statistician who drowned wading across a river of average depth, 3 feet. We can take the average here, and that doesn't mean that that's all kids eat. And in fact, the foods that are targeted to children tend to be more brightly colored and have a greater density of the dye in them.

MS. LEFFERTS: Just one other question I had. You were talking about the need for careful diagnosis by the DSM, but weren't you also agreeing with what we heard from Dr. Chronis-Tuscano and Dr. Stevenson about the continuum of symptoms that we're seeing with a normal distribution of hyperactivity, and ADHD is way over here, so that we

don't -- in our zeal to get a good diagnosis, we don't want to also miss the more general population that may be having symptoms that are in the same ballpark.

DR. ARNOLD: Yes. I agree completely with you, that there are two issues here. One is, what's contributing to ADHD? The other is the general public health problem of what are the dyes doing to all the kids in the country. And those are both very important issues. And my only reason for wanting the DSM diagnosis is to be able to help clarify that first problem. It wasn't to say that those who don't meet the diagnostic criteria are not important, that we shouldn't worry about those.

DR. ACUFF: Dr. Jones?

DR. JONES: Tim Jones. This may be related, but I guess the Southampton study aside, a very substantial majority of the studies we were looking at involved a very select population of subjects. So could you maybe talk a little bit about your very last point, that you think this is more a public health problem than an ADHD problem?

DR. ARNOLD: Yes. That is based largely on the Southampton studies, which were presented earlier this morning, which changed my whole thinking about it. In the past, I knew that it was a small contributor to ADHD. Looking at all the studies on balance, and also clinical experience, it seemed that there's a small subgroup of those with ADHD who had a significant benefit from an elimination diet. And it's possible that that subgroup are those who don't respond well to medication or other treatments, so it's important to keep that in mind.

But the first study, the Isle of Wight study that Stevenson presented, started changing my thinking, and then I was really impressed when it was replicated. I thought, gee, here's something that affects all the kids in the country, or most of them anyhow, and we really need to think about this and take this seriously. It's not just a small contributor to ADHD.

DR. ACUFF: Additional questions? Dr. Freeland-Graves?

DR. FREELAND-GRAVES: I was interested in

your work on this study on zinc and the ADHD. Do using supplements reduce behavioral problems?

DR. ARNOLD: This is kind of a diversion from the topic. Now, summing up all of the work on zinc and ADHD, there have been quite a few studies that demonstrated that kids with ADHD have lower zinc by various measures than other kids. For example, in serum, nails, hair, urine, et cetera, maybe they don't absorb it as well or whatever. That includes at least one study showing that the level of zinc nutrition correlates with placebo-corrected response to dextroamphetamine, one of the medications commonly used, one of the stimulants that you heard about earlier. So then the next obvious question is, if you supplement with zinc, can you help that?

There were three studies done in the Mid-East, two in Turkey, one in Iran, which demonstrated a significant effect, both as monotherapy, a comparative placebo, and as adjunct to methylphenidate, Ritalin, one of the types of stimulant that's very useful in ADHD. So then I naturally thought that we should check this in an

American sample because of differences in diet and so forth.

The only thing we found in an American sample was that it reduced the dose needed of amphetamine for an optimal response. It reduced the dose by about a third if the kids were given a zinc supplement compared to placebo. It didn't improve the response, but it just lowered the amount of drug needed to get that same response.

This is very reminiscent of what behavioral therapy can do. It's been demonstrated several times over that you can get an optimal response with a lower dose of drug by supplementing with behavioral treatment. Zinc supplement is probably a little easier and cheaper to use than behavioral treatment, but this needs to be replicated before it would be widely applied.

DR. ACUFF: Dr. Fernandez?

DR. FERNANDEZ: I guess I have a question about the Southampton study. So you're saying that the use of dyes should apply to the general public. But in addition to the Southampton study, is there

any other study that shows that these dyes can affect the general population?

DR. ARNOLD: I don't know of any other study that was specifically done to test that hypothesis. However, I did point out that in one of the early studies, 2 of 20 control children showed a definite bad reaction to a dye challenge, and wondered if that may be relevant to that. In fact, looking back at that, you sort of wonder why somebody didn't think of doing what Stevenson's group did later -- why they didn't think of it sooner, that maybe it applied more to the general population.

DR. FERNANDEZ: So the other question that I have is, at this point, do you think it's possible to separate the effects of the dyes from the effect of the preservatives in the Southampton study?

DR. ARNOLD: No. I don't think it's possible, completely. However, there is one thing that we might look to, to help with that. All four of the mixes included the same dose of the preservative, 45 milligrams of sodium benzoate. However, when you look at the results, there seems to

be a dosage effect, at least in the eight- and nine-year-olds, a slight dosage effect which cannot be ascribed to the preservative, but might reflect the dye. However, that was not significant. So it's only a hint. It raises a question that needs a study design to answer it.

DR. FERNANDEZ: Thank you.

DR. ACUFF: Dr. Winter?

DR. WINTER: I think this came up a little bit in the last discussion. I think Dr. Gray mentioned this as well. There's a lot of different ways to look at that, and obviously nobody has the crystal ball because the study was done the way it was done with the different mixtures.

But I believe Dr. Stevenson said there was no effort to separate mix A from mix B. And, in fact, when you look at those results, they look very, very similar. So another way to look at that would be, what is the common factor in each of those mixtures, which was the sodium benzoate, and not the other colors. So there are a lot of different ways to look at it, so I think we have to be careful

there.

DR. ARNOLD: Yes. I think you're right. Without a study designed to look at that issue, we really don't know. We can only guess, and one guess is as good as another.

DR. ACUFF: Ms. Lefferts?

MS. LEFFERTS: The question about the zinc, iron, and magnesium deficiencies that you mentioned reminded me that there's been some consumer concerns about some of the contaminants in food dyes. For example, I was noticing that a number of known neurotoxic contaminants, like mercury, lead, arsenic, are contaminants in food dyes.

Where is it? For example, Blue 4 can have up to 20 parts per million lead. 3 parts per million arsenic, 1 part per million mercury. And some of these can interact, I know -- can compete, and influence it. Deficiencies in some of these nutrients can affect the toxicity of some of these compounds.

Has that at all been a concern from your perspective? Is that an issue?

DR. ARNOLD: I think it's an issue that needs to be looked at. Now, remember that the dose is a huge issue. And so the amount of these heavy metals in the dyes has been judged not to be of significant risk. However, the problem is that -- if you look at what they get in the dye, at the expected ingestion dose, that's not a problem. But if it's 10 times that much, if they're one of these outliers that take 400 milligrams, does it become a problem then?

Further; does it become a problem if you add that to a youngster who may have a little bit of heavy metal burden from some other source, like gentrification or living near an area that had been heavily lead polluted by highway traffic before it was taken out of gasoline, things like that? And you have the accumulation problem, of course, too, which is an issue for the dye, even disregarding the heavy metal. I think there was some data presented here before that if you take away all of the additives for a period of time, then the youngster can tolerate a challenge and not show the effects. But if they do

it for several weeks, then it kind of accumulates. And we saw the blue colon or whatever. So it obviously does accumulate in the body somehow. And I don't know that anybody's studied that effect. Maybe they have and I just don't know about it.

DR. ACUFF: Mr. Waldrop?

MR. WALDROP: Has anyone studied the effect of preservatives in food on ADHD, at all?

DR. ARNOLD: I don't know. In other words, a study where they used only the preservative by itself? I can't think of a study off-hand, but I know where you're going with it. If we had the results of that study, it would help to answer the other question that we were struggling with a minute ago. Good point. Maybe we should do a list search on that.

DR. ACUFF: Thank you, Dr. Arnold.

DR. ARNOLD: Thanks for your attention.

DR. ACUFF: We have one additional question for Dr. Weiss, if he's still here.

Ms. Lefferts?

MS. LEFFERTS: Thank you, Dr. Weiss. I

thought we needed to ask you to respond to the FDA criticisms of your study. In particular, they thought it was unclear how effective the blinding was in your study. Could you address that point? And if you want to, address any other point that they raised

DR. WEISS: We were very concerned about that because the blend of food dyes we used, which was bottled by 7-Up in San Francisco, left a red stain on filter paper. "Oh, my God, we said. We've got to blind it." So we tried beet extract, which worked fine, but made the drink taste like borscht.

[Laughter.]

DR. WEISS: So we ended up with caramel and cranberry coloring, and found that nobody on the staff could distinguish the two stains.

MS. LEFFERTS: I didn't know if you wanted to address any other points by the FDA.

DR. WEISS: I like to step back and think about this question in a much broader context. As Dr. Arnold pointed out, an effect size, say, of .2 really is the basis on which we have established, say on the basis of IQ rather than hyperactivity

symptoms, the current level of concern of 10 micrograms per deciliter.

But think again about blood pressure. The Congress asked the EPA about the economic benefits and health benefits of the Clean Air Act. This is in 1990. And they calculated that by removing lead from gasoline, which had two effects, reduced population of blood pressure and also changed lifetime earnings in those -- or based on IQ. The economic benefits they calculated amounted to $2 trillion, half for a lowering of blood pressure, hypertension, and second by earned income over a lifetime for IQ.

Now, a 2-millimeter drop in the population of blood pressure has enormous health implications because it would be correlated with the number of cardiovascular incidents. So a very small change in a population means a very big health and economic impact for many diseases, and of course, for the economic benefits.

Dr. Arnold mentioned something else that occurred to me. And that is, it looks as though, and it's our impression, that younger children are more

sensitive to the dyes than older ones. I think for Dr. Voorhees and myself, that raises the question, what about prenatal and early postnatal exposure?

Now, suppose I went before my IRB at Rochester and said, I would like to do a study feeding pregnant women food dyes to see what the effects are in the offspring. Now, they would look at me and say, "Are you crazy?" Could you recruit anyone into such a study, provided the IRB let you, with your statement of informed consent? And who would volunteer for such a study given what we now know about the impact of food dyes on the brain and behavior?

Now, you asked the question about the brain. Is it neurotoxic? If there is an effect on behavior, somehow, it's acting through mechanisms in the brain.

DR. ACUFF: Thank you, Dr. Weiss.

DR. WEISS: Sure. Thank you.

DR. ACUFF: We are adjourned for the day, and then we will re-adjourn tomorrow morning at 8:30.

(Whereupon, at 5:24 p.m., the meeting was adjourned.)

CERTIFICATE OF COURT REPORTER

I, Janet Evans-Watkins, do hereby certify that this transcript was prepared to the best of my ability.

I am neither counsel for, nor party to this action nor am I interested in the outcome of this action.

_____

Janet Evans-Watkins