Responses to Public Comments on the Office of Pesticide Program's Draft Science Policy Document:

Choosing a Percentile of Acute Dietary Exposure As a Threshold of Regulatory Concern

Health Effects Division
Office of Pesticide Programs
US Environmental Protection Agency
Washington, D.C. 20460

March 16, 2000

List of Acronyms

CEC	Critical Exposure Contribution
CSFII	Continuing Survey of Food Intake by Individuals
DRES	Dietary Risk Evaluation System
DEEM	Dietary Exposure Evaluation Model
FDA	Food and Drug Administration
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
IABNMRR	Interagency Board on Nutrition Monitoring and Related Research
LOAEL	Lowest Observed Adverse Effect Level
LOD	Limit of Detection
LOQ	Limit of Quantitation
MaxLIP	Maximum Likelihood Imputation Procedure
MOE	Margin of Exposure
MRL	Maximum Residue Limit
NHANES	National Health and Nutrition Examination Survey
NHEXAS	National Human Exposure Assessment Survey
NOAEL	No Observed Adverse Effect Level
NOEL	No Observed Effect Level
NAS	National Academy of Sciences
OPP	Office of Pesticide Programs
PAD	Population Adjusted Dose
PDP	Pesticide Data Program

PHI	Pre-harvest Interval
QA/QC	Quality Assurance/Quality Control
RfD	Reference Dose
SAP	FIFRA Scientific Advisory Panel

Table of Contents

I. Introduction

A. Background

On April 7, 1999 the US Environmental Protection Agency (USEPA), Office of Pesticide Programs (OPP) issued in the *Federal Register* a Notice of Availability (along with a request for comment) regarding a science policy paper on the threshold of regulatory concern for acute dietary risk assessments. This document, entitled *Choosing a Percentile of Acute Dietary Exposure as a Threshold of Regulatory Concern* (U.S. EPA, 1999a), discussed a set of issues dealing with the selection of an appropriate threshold on which to base regulatory decisions concerning pesticide registrations and reregistrations.

Many parties commented on the interim policy statement (submitted under docket OPP-00593). They included pesticide registrants, environmental and public interest groups, consultants, private citizens, the Canadian government, and numerous farm bureau federations. All comments and recommendations were reviewed by OPP and incorporated as appropriate, into the current 1999 revised science policy document. The comments ranged in specificity. Some commenters addressed the general policy and its rationale as well as all of the specific questions posed, while other reviewers provided detailed comments only on certain aspects of the policy, such as risk management issues, data quality and uncertainty, modeling issues, and suggested enhancements or modifications. A listing of the names and affiliations of the parties submitting comments is provided at the end of this document (See Section IV– List of Commenters)

B. Organization of this Document

This response package contains OPP's responses to the comments raised on this science policy paper. The document is organized by topic area, each of which contains a summary of the key elements of the 1999 interim science policy guidance, a synopsis of the public comments which were submitted, and the Agency's response. These responses include OPP discussion of the comments received on the seven questions posed by OPP in the science policy paper:

1. What are the appropriate statistical techniques for characterizing the uncertainty at the high end of the distribution of probabilistic exposure assessments? At what point does an exposure estimate become so uncertain that it would be inappropriate to use the estimate in regulatory decision making? How does uncertainty about one or more high-end values in a data set affect the reliability of the output of probabilistic models using that data set as an input?

2. Regarding the Agency's current methodology for performing Monte Carlo analyses, at what percentile of estimated exposure is it appropriate for the

Agency to establish its threshold of concern? 99.99, 99.9, 99, 95, or some other percentile? What are the reasons for recommending that percentile? How should the characteristics of the data sets used as input to the assessment (*e.g.*, the type of residue data, field trials vs. PDP monitoring data) affect the choice of a percentile exposure for OPP's threshold of concern?

3. If OPP chooses to set its threshold of concern lower than the 99.9th percentile, should any other steps, such as the application of an additional safety factor, be employed to assure that the statutory safety standard is satisfied?

4. Some advocate a "sliding regulatory scale" with more serious toxic effects regulated at higher thresholds; they contend that such an approach would explicitly acknowledge all aspects of the risk management decision and incorporate the nature of the toxic effects and the built-in conservatism on the hazard identification and dose response side of the equation. Instead of using only a single percentile for all toxicological effects (regardless of severity), should the Agency regulate pesticides at a variety of percentiles, depending upon the toxic effect observed? For example, would a lower threshold of regulation (perhaps the 98th percentile) be warranted for fully-reversible effects (such as mild anemia) or would a more stringent threshold (perhaps the 99.9th percentile or higher) be justified for severe, non-reversible effects (*e.g.*, birth defects)? Finally, should the Agency regulate pesticides at different percentiles according to the nature and size of the subpopulation groups (*i.e.*, use the 99.9th percentile for larger groups and another percentile for smaller groups)?

5. How should "outliers" be identified for food consumption data sets? For residue data sets? When an "outlier" is identified, how should the data point be handled in generating probabilistic exposure estimates?

6. If OPP conducts a Critical Exposure Contribution (CEC) analysis, and excludes one or more data points because they appear to drive the high-end estimates of exposure, should OPP perform an additional CEC analysis on any revised estimate of the exposure distribution?

7. Should OPP's probabilistic assessments attempt to reflect variability in human sensitivity to toxic effects, as suggested by the FIFRA SAP? If so, how should this be done?

To organize the responses to the comments received on these seven questions, OPP has combined them into several larger topic areas:

- ❖ Risk Management and Policy and the "Bright Line" Issue
- ❖ What Population Percentile Should Be Used
- ❖ Monitoring vs. Modeling
- ❖ Data Quality and Uncertainty
- ❖ Clarification of Issues and Ideas
- ❖ Suggestions for Future Directions
- ❖ Issues Beyond the Scope of the Document
- ❖ Incorporating Toxicology as a Probabilistic Distribution

A brief summary of the comments in each topic area is provided immediately prior to the detailed responses in the relevant section.

II. Response to Comments

A. Risk Management and Policy and the "Bright Line" Issue

✔Science Policy vs. Management vs. Assessment
✔Choosing a Single "bright line"

Overview. *A number of comments relate to the presentation of this document as a science policy issue, when in reality the commenters claim it to be a risk management policy. Several commenters pointed out that OPP should be mindful of the NAS Risk Assessment paradigm that calls for risk assessment and risk management to be distinct areas which should be considered separately. The commenters claim that the Agency is blurring this distinction by issuing the document as a science policy paper.*

1. Policy vs. Science

Comment. Several commenters were concerned that this document (which is henceforth called the *Percentile Policy* document) was presented as "science policy" and contended that it instead relates to a risk management

decision. One commenter stated that decisions to use one percentile or another are risk management issues and are not based on science but on perception. Another commenter stated that while the material describing the methodology for acute dietary risk assessment was useful, the question of whether to use the 99.9th percentile is a risk management decision and values for public health protection are more important than scientific considerations in considering this question.

Response. OPP is in basic agreement with the commenters. The original intent of the document was to propose a regulatory threshold (99.9th percentile) and provide background information on the reasons that this threshold was selected, give support for its reasonableness and validity, and ask for comments on why this threshold should or should not be considered a "baseline level" for pesticide risk management decisions. OPP recognizes that the document relates not only to "risk management," but also to "science policy" in that the decision pertaining to a regulatory threshold combines both scientific considerations and societal values and choices. Regardless of whether it is regarded as a risk management document, a science policy document, or a mixture of both, OPP believes that the ideas it contains were worthy of a wide airing and public discussion.

2. Statutory vs. Selected Regulatory Level

Comment. Another commenter stated that there was an implication that the 99.9th percentile policy is somehow directly linked to the "statutory" language of FPQA and that, instead, the 99.9 is a value selected by risk managers or policymakers at EPA.

Response. The statutory reasonable certainty of no harm standard informs OPP judgment on the overall risk management decision as well as on component parts of the decision. Accordingly, when OPP selects a population percentage it must weigh how that percentage fits with OPP's overall obligation of determining whether the pesticide meets the reasonable certainty of no harm standard. That said, it is a mistake to suggest that the selection of a population percentage for conducting an exposure assessment is a judgment about how much of the population the reasonable certainty of no harm standard directs OPP to protect. As explained in both the policy document and this Response to Comments, OPP's goal in an exposure assessment is make a reasonable high-end estimate of exposure for the general population and all major, identifiable population subgroups. To that end, OPP, at times, will vary the population percentile used in computing the estimated high-end exposure based on several factors, including the conservativeness of the residue values used in the assessment. Thus, when OPP picks the 95th percentile, the 99.9 percentile, or some other percentile, OPP is not making a judgment that only that portion of the population deserves protection.

Actually, commenters from both industry and environmental groups seemed to understand this underlying principle even if they disagreed with how it was implemented. Thus, for example, the Environmental Working Group wrote that "[a] policy is inherently flawed if it protects less than 100 percent of the population. The Agency plainly recognizes this fact and attempts to justify its 99.9 percent proposal by invoking what it sees as the inherent conservatism of the risk assessment process." Similarly, a broad-based industry group [IWG] commented that "[t]he issue is not whether OPP should try to protect everyone from adverse effects from dietary exposure. Rather, the issue is *how* OPP should do that." (emphasis in original).

3. National Academy of Sciences Risk Assessment Paradigm

Comment. Several commenters brought up the National Academy of Sciences' (NAS) risk assessment paradigm that recommends that risk assessment be based on reliable scientific information and be separated from policy issues. This commenter stated that it is critical that separation between risk assessment and risk management be maintained:

> The risk assessment paradigm established by the National Academy of Sciences during the Reagan administration specified a clear separation between risk management and risk assessment, because risk managers had attempted to influence the results of assessments. If the separation between risk assessment and risk management is weakened, this moves the Agency closer to a situation where risk assessment scientists are constantly looking over their shoulders to conduct re-assessments to support pre-determined risk management decisions. If such a situation results, then the risk assessment process will have been corrupted and regulatory decisions will not be credible.

Another commenter discussed the NAS risk assessment paradigm more specifically: Under the NAS paradigm, risk assessment for the evaluation of safety of pesticides should be conducted based on available data without including safety factors or other risk management tools. Risk management should be conducted after the risk assessment is complete and, and that point, science policy and other societal and regulatory factors should be considered alongside the science-based risk assessment to make final regulatory decisions.

Response. OPP generally agrees with the comments and does use the NAS risk assessment paradigm as a model: risk management activities are conducted separate from, and following, risk assessment activities.

4. The Pragmatic Value of These Policy Papers

Comment. One commenter believed that the Agency's description of these policy papers would lead one to question their practical value; the FR notice for each paper describes it as a policy document and not a binding rule. The commenter is concerned with the phrase in the document "on a case-by-case basis, EPA will decide whether it is appropriate to depart from the guidance...." He stated that the phrase "case-by-case" can cover a "multitude of sins," and that "one is left with the impression of documents written in sand." The commenter stated that the FR notices commit the Agency to explain its departures from the policy documents and the Agency should hold to this commitment strictly, making clear the impact of each deviation on particular risk assessments.

Response. Any deviations will be fully explained by OPP's risk managers and will be supported by a full and open risk characterization performed by OPP's risk assessors. An inherent feature of a guidance policy is that it is not binding on either the Agency, the regulated industry, or members of the public. Decisions following the guidance cite it not as authority for the decision but as an explanation for the reasonableness of the decision. If EPA departs from the guidance it will separately have to provide an explanation for the reasonableness of its decision.

5. The Bright Line

Comment. Several individuals discussed the concept of a "bright line" at the 99.9th percentile. That is, they were concerned that OPP might apply this policy guidance inflexibly and would invoke a blanket policy that it is appropriate to regulate in all cases acute dietary risk at the 99.9th percentile, regardless of any mitigating factors or the quality of the supporting database. One commenter, however, specifically stated that he was encouraged by the statement that the 99.9 is not necessarily a "bright line" and considerations would be given to the "drivers" of risk assessment before making mitigation decisions. The commenter indicated that this point should be emphasized and built upon and that 99.9 as a "bright line" is highly inappropriate as a regulation point. In a similar vein, one commenter indicated his support for the use of probabilistic assessments for exposure assessment, but indicated that he did not support choosing a single exposure percentile at which to set a threshold of regulatory concern. Rather, he indicated that all relevant information should be considered, and each regulatory decision should be made on a case-by-case basis using all available information on potential risk, including exposure, hazard, magnitude and severity of potential effects, and data quality and certainty. He stated that allowing the use of probabilistic techniques and then choosing a single percentile at which to regulate exposure for all substances is contrary to using the best available science and risk assessment technique in that it inappropriately mixes the science of assessment (*i.e.*, probabilistic analysis) with a risk management decision. Risk management (here,

the selection of a threshold of concern), the commenter claims, should be conducted after the risk assessment is complete. An inflexible regulatory threshold cannot be selected for each and every substance before a risk assessment is complete and strengths and limitations of the assessment are carefully considered. The commenter continued:

> By using all of the available information, a probabilistic exposure analysis provides a wealth of information to a risk manager in a transparent manner, including the range of possible exposures, uncertainties, assumptions, and variability. *Requiring the risk manager to essentially ignore these data and use a single pre-determined threshold will defeat the purpose of providing the manager with all of this information*. The broader picture of exposures is lost and the risk manager cannot consider how the results compare with other points along the distribution....
>
> ...The guidance recognizes the role of the risk manager in determining the level of regulation... where it states: 'the conservativism [of a risk management decision] is determined by a risk manager when he or she determines the appropriate percentile of the model's output distribution (*e.g.*, 99.9^{th} percentile) to be used for regulation.' [We] believe that it is inappropriate that the guidance has nonetheless usurped the role of the risk manager by pre-determining a level at which the manager would regulate each and every substance irrespective of their unique analyses.

Continuing on this theme, another commenter stated that the greatest overall shortcoming in the current policy draft on the choice of a percentile for a regulatory threshold of concern is the failure to recognize the difficulties inherent when selecting any single "bright line" from a Monte Carlo analysis as a decision point for regulatory managers. Two fundamental difficulties, the commenter continued, undermine the utility of a probabilistic approach to acute dietary risk assessment. Firstly, the use of a discrete exposure endpoint negates the major strength of a probabilistic assessment, *i.e.*, the ability to evaluate the entire distribution of likely outcomes arising from consumption of pesticide residues. And secondly, the use of an extreme outlier in the output distribution adds unnecessary uncertainty to the risk assessment and clouds sound risk management decision making. Both of these difficulties arise from limited use of the rich information contained within the outcomes of a Monte Carlo analysis. In particular, the commenter argued that Monte Carlo results must be fully utilized in coming to a regulatory decision and that a single "bright line" for decision making cannot be established *a priori*. Instead, the selection of appropriate risk management decision points should consider the nature of the exposure distribution, the severity of the effect being assessed, and the robustness of the available residue and consumption data.

Response. OPP does not intend that the 99.9^{th} percentile be used as a

"bright line" for regulatory decision making. It is meant, instead, to represent a "baseline" or "benchmark" which is evaluated on a case-by-case basis. When exposures exceed a threshold of concern calculated using the 99.9th percentile, OPP would potentially be concerned about the level of exposures to the general population or specific subgroup of concern, but these potential concerns could be appropriately addressed by a full characterization of the issues including the inherent uncertainties and biases in the assessment. In some situations, a threshold based on a lower population percentile may be appropriate and could be determined on a case-specific basis. Any decision to depart from the 99.9th percentile would be fully and clearly explained by the risk manager. The specifics of this approach, and the criteria relevant to any decision to depart from the 99.9th percentile, are discussed in additional detail in the next section ("What Population Percentile Should Be Used") of this response to comments.

6. Policy or Rule

Overview. OPP requested comments on how this policy could be structured so as to provide meaningful guidance without at the same time imposing binding requirements on either the government or outside parties. Other than the comments on the "bright line" issue, OPP received few comments on this issue. Nonetheless, OPP believes it is appropriate to respond to two other sources of comment on this issue in this document. The first is a Petition from the American Farm Bureau Federation, the American Crop Protection Association, and other food and pesticide industry groups. Petition for Rulemaking to Develop Policies and Procedures for Implementing the Food Quality Protection Act of 1996 (May 22, 1998). That petition claimed that OPP's policy on use of the 99.9th percentile had been implemented as if it was a rule and urged OPP to promulgate that policy as such. The second is a lawsuit filed against EPA by the same parties making similar allegations and seeking similar relief. American Farm Bureau Federation, et al. v. EPA, Case No. 1:99CV01405 RCL (D.D.C.).

Petition. The AFBF/ACPA Petition requested that the Agency undertake rulemaking on a number of topics including aggregate exposure. Although not specifically mentioning the percentile of exposure as a topic for this rulemaking, the petition did suggest that the rulemaking address "how exposure from the specified routes will be assessed." Pet. at 27. Elsewhere the petition asserts that EPA's policy decision to use the 99.9th percentile for probabilistic acute assessments "clearly constitutes a 'legislative rule' that, for both legal and practical reasons, should have been issued through notice-and-comment rulemaking." Pet. at 20.

The petition lists various policy and legal reasons for issuing rules regarding FQPA implementation. The policy reasons include: (1) a rule provides

greater transparency because the notice-and-comment process will provide formal notification of EPA's views; (2) rulemaking will give all parties a chance to participate in the development of policy not just those invited to Agency advisory committees; (3) in a rulemaking EPA must respond to public comments on the public record and must provide a concise statement of the basis and purpose for the rule; (4) a rule provides certainty and stability because rules are subject to judicial review and legal issues can be resolved once and for all; (5) the advisory committee process and SAP review of policies has not adequately provided for public participation; and (6) rulemaking on individual tolerances has not been an adequate substitute for generic rulemakings. The legal reasons listed in the petition include: (1) that FQPA policies 'impose obligations' and have 'significant effects on private interests' and thus are, in fact, legislative rules requiring notice-and-comment procedures; (2) the FQPA "requires EPA to use notice-and-comment rulemaking to establish general requirements or procedures for implementing the key provisions of the FQPA." Pet. at 15

Legal Challenge. In the course of the AFBF/ACPA lawsuit, these industry groups have cited portions of the Percentile Policy that they consider to impose binding requirements. AFBF/ACPA wrote:

The 99.9th Percentile Policy is also written in binding language. It states on page 1, for example, that it "has broad applicability to many pesticides and potentially significant impact on the assessment of these pesticides." It goes on to say that EPA "*has decided* to express its risk management judgment for acute dietary risks in quantitative scientific form, as a 'threshold of concern'" -- "such that the 99.9th percentile of estimated daily exposure, using probabilistic exposure estimation techniques, *must be equal to or less than the Population Adjusted Dose* (PAD)."

....

Notice and comment are also required because the two Science Policies make significant changes in prior EPA practice and policies. It is axiomatic that an agency's change in existing policy constitutes a legislative rule requiring notice and comment. [cites omitted] The 99.9th Percentile Policy without doubt represents a significant change from EPA's prior policy.

. . . .

EPA's assertion that "the plain language of the policies makes clear that EPA does not intend to bind itself" is demonstrably false. The policies themselves contain no such indication. The general disclaimer EPA cites is found only in the Federal Register notices, not the policy papers themselves.

Response. After considering the petition and the material in AFBF/ACPA legal papers, OPP has decided to issue the 99.9th Percentile Policy as a nonbinding policy guidance not as a binding rule. Accordingly, EPA denies the AFBF/ACPA petition to the extent it sought rulemaking regarding this policy.

The reasons for issuing this document as a policy guidance are set forth in the policy itself. In the policy OPP explained:

> Because of the need to balance a variety of factors in selection of a population percentile for calculating a threshold of concern, OPP is issuing its views regarding population percentiles as a non-binding policy guidance rather than as a binding rule. Complex risk assessment and risk management issues such as those involved in this policy seldom can be reduced to meaningful rule-style commands. Rather, the scientist and risk manager need to have flexibility in considering a variety of factors and outcomes. This policy is intended to focus the analysis on factors deemed most critical without barring consideration of other factors which may be found to be relevant. As a policy, this guidance does not – in fact, as a legal matter, cannot – draw bright lines or preclude reconsideration of basic principles. EPA would retain the option to depart from the policy. Further, affected parties remain free to challenge the specific application of the policy or the underpinnings of the policy itself.

Percentile Policy at 29. This position is consistent with the manner in which the Agency generally approaches complex risk assessment issues. Thus, EPA's views on major risk assessment topics have been issued as policy guidances not binding rules. See e.g., Guidelines for Carcinogen Risk Assessment, 51 FR 33992 (September 24, 1986); Guidelines for Reproductive Toxicity Risk Assessment, 61 FR 56274 (October 31, 1996); Guidelines for Exposure Assessment, 57 FR 22888 (May 29, 1992); Proposed Guidelines for Carcinogen Risk Assessment, 61 FR 17960 (April 23, 1996). In their petition, AFBF/ACPA cited to one EPA proposed rule that included "models and assumptions for estimating public exposure" concerning certain air emission standards. See 59 Fed. Reg. 15504 (April 1, 1994). However, OPP would note that when that rule was finalized, the portions addressing risk assessment were omitted. 61 Fed. Reg. 68384 (December 27, 1998).

EPA found none of the arguments set forth in the AFBF/ACPA Petition to be persuasive. Each of those arguments are addressed in turn below.

> Transparency. AFBF/ACPA argued that a rule would provide greater transparency because there would be formal notification of all parties concerning the rulemaking. However, this formal notification concern was met by the procedure EPA followed in developing this policy. EPA

published notice of the draft policy in the Federal Register. 64 FR 16962 (April 7, 1999). That notice provided a concise summary of the policy and requested public comment on the policy. Further, EPA put a full copy of the policy on its Internet Web site and generally made copies available to the public.

Public Participation. AFBF/ACPA argued that a rulemaking would allow all affected parties to participate not just advisory committee members. That concern, however, has also been met by EPA's public comment process.

Response to Comments. AFBF/ACPA expressed a concern that without a requirement to respond to comments and to provide a statement of the basis and purpose for the policy, OPP would not in fact produce such documents. OPP, however, believes that its policy document clearly articulates the basis and purpose of the policy and that this Response to Comments document has adequately addressed all significant comments.

Judicial Review. AFBF/ACPA argued that a rule provides certainty and stability because unlike a policy document it would be subject to judicial review. Generally, policy statements are not reviewed as ripe for review until they have been applied to a concrete regulatory action. Similarly, generic rules are often found unripe on the same grounds. On occasion, courts will review a generic rule in the absence of a concrete application of the rule where a challenge to the rule presents purely legal questions and there would be hardship to the challenger in delaying review. As to the *Percentile Policy*, however, few, if any of the comments on the policy raised purely legal questions. Rather, most of the comments addressed the factual underpinnings of the policy and its application in specific circumstances. Thus, even if this policy was promulgated as a rule, OPP does not expect there would be many issues that could be resolved by immediate judicial review. Thus, this consideration does not appear to strongly support issuance of the policy as a rule.

Advisory Committee Process and SAP Review. AFBF/ACPA claimed that Agency attempts to get outside input into its policies through various advisory committees and the FIFRA SAP have been inadequate. OPP believes the advisory committee process and SAP review have provided important input. However, to the extent these processes have provided only a limited forum for public participation, the notice-and-comment process for the policy has addressed any such concern.

Individual Tolerance Rulemakings. AFBF/ACPA argued that OPP has not

opened its policies up for comment in rulemakings addressing individual tolerances. AFBF/ACPA also imply that application of OPP policies in the context of such tolerance actions is not subject to judicial review. Pet. at 24. Although EPA has not specifically requested comments on its policies in tolerance actions, such comments would certainly be appropriate to the extent the policy formed part of the basis for OPP's decision. Moreover, AFBF/ACPA is clearly incorrect if they are suggesting that the lack of an explicit request for comment on policies underlying a specific tolerance decision somehow insulates the policy's application from administrative and judicial review. In any event, OPP has held a separate notice-and-comment period on the Percentile Policy.

Similarly, EPA found none of the legal reasons contained in the AFBF/ACPA Petition to have merit.

Policies Impose Obligations. AFBF/ACPA argued that FQPA policies generally and the *Percentile Policy* specifically impose obligations and have significant effects on regulated parties and thus these policies are binding rules and must be promulgated following Administrative Procedure Act (APA) requirements. OPP has attempted to make clear that the *Percentile Policy* does not impose binding obligations on either regulated parties or the government both in the policy document and in this response to comments. Further, OPP does not believe that the policy itself has significant effects on regulated parties in that it imposes any rights or obligations. Rather, the considerations in the policy when taken into account in an individual action may affect the ultimate decision in that action.

FQPA Requirement for Rulemaking. AFBF/ACPA claimed that section 408(e)(1)(C) requires that general procedures for implementing section 408 must be promulgated as rules. The language of section 408(e)(1)(C), however, is clearly permissive – "EPA *may* issue a regulation . . . " (emphasis added). This language authorizes EPA to establish rules for "general procedures and requirements to implement this section;" it does not mandate such rules.

EPA has modified the language in the policy to some extent in response to the AFBF/ACPA statements in their court papers concerning particular language in the policy. However, EPA was not convinced by the AFBF/ACPA's broader arguments that the policy is, in fact, a binding rule. AFBF/ACPA first argued that the following language showed the policy was a rule: "[the policy] has broad applicability to many pesticides and potentially significant impact on the assessment of these pesticides." EPA does not agree that a statement that the

policy will be useful in many risk assessments ("has broad applicability") shows an intent that the policy be binding; thus this language remains unaltered. However, the language concerning a "potentially significant impact" on risk assessments is subject to the misinterpretation that EPA intends the policy to have a significant impact on the rights and obligations of affected parties. Thus, this language is deleted. Clearly, selection of a population percentile in a risk assessment is an important part of the risk assessment and, in some circumstances, the population percentile that is selected can significantly affect the estimated risk. Because the population percentile of exposure is an important part of risk assessment is why the Agency has made its views public and sought comment on those views. EPA does not read the cases cited by AFBF/ACPA as holding that an administrative agency can only express its views regarding an important issue through a substantive legislative rule. Rather, the critical inquiry is whether the agency action binds private parties (e.g. imposes rights or obligations) or binds the agency. As explained above, this is not the situation here. See, e.g., Troy Corp. v. Browner, 120 F.3d 277, 287 (D.C. Cir. 1997) ("EPA's exposure policy was exempt from the notice and comment requirements of section 553. The EPA's exposure policy merely informed the public that the agency would exercise its discretion by considering exposure only for low toxicity chemicals. The EPA did not thereby curtail this discretion; it did nothing more than clarify its own position. The policy does not impose rights or obligations or bind the agency to a particular result.")

Second, AFBF/ACPA cites the language stating that exposure "must be equal to or less than the Population Adjusted Dose (PAD)." EPA would agree that the use of the word "must" in this sentence conveys the impression that this aspect of the policy is binding. However, that impression was dispelled in the language immediately following this sentence in the draft policy. Those following sentences made clear that, if a risk assessment using the 99.9th percentile exceeded the PAD, OPP would conduct a further analysis to evaluate the acceptability of the risk. In the revised policy, the word "must" has been deleted from this sentence and the entire paragraph where this sentence appeared has been reworked to clarify that the policy does not establish binding rules regarding the effect of the outcome of a particular risk assessment but rather describes the path OPP generally will follow in evaluating whether a specific exposure exceeds the safety standard. The policy leaves OPP wide latitude in varying the choice of population percentile where the considerations warrant.

EPA does not agree with AFBF/ACPA that the use of the 99.9th percentile represents a significant change in policy. It has been a longstanding policy for the OPP to assess exposure in a manner designed not to underestimate exposure. As OPP explains in the Percentile *Policy*, it believes that use of the 99.9th percentile with probabilistic risk assessments (a relatively new form of risk assessment) is the

way to remain consistent with that longstanding approach. In any event, EPA disagrees with AFBF/ACPA's assertion that when significant changes are made in prior policies, those changes can only be made through a substantive, legislative rule. The essence of a policy statement is that it is not binding on the agency; thus, the agency remains free to act in variance with the policy so long as it explains its change in course. See Syncor Int'l Corp. v. Shalala, 127 F.3d 90, 94 (D.C. Cir. 1997) ("The agency retains the discretion and the authority to change its position--even abruptly--in any specific case because a change in its policy does not effect the legal norm.").

Finally, AFBF/ACPA cited the lack of a general disclaimer in the policy stating that the policy was not intended to be binding on OPP, regulated parties, or the public. OPP believes that such a disclaimer is important and has included one prominently in the policy.

B. What Population Percentile Should be Used

✔ Coordination with Other Agencies and Federal Programs
✔95th DRES vs. 99.9 DEEM -- Why is the Agency Becoming More Stringent?
✔ Suggestions on What Percentile to Use
✔ Flexible Approach and Criteria for Consideration

Overview. *A variety of comments were received which concerned the appropriate point (or percentile) to use in calculating the threshold of concern. Some of these comments urged the Agency to use a population percentile consistent with other federal programs while other comments addressed the perceived change from the 95th percentile (formerly used by OPP when exposure estimated a calculated the Dietary Risk Evaluation System (DRES) software) to 99.9th percentile (when using probabilistic techniques with the Dietary Exposure Evaluation Model (DEEM) software). In addition, some commenters made a number of suggestions and recommendations with respect to where they believed the appropriate point of regulation should be, addressing the issue of an appropriate regulatory threshold and other related concerns. Overall, certain commenters addressing these issues believed that the appropriate point of regulation should lie between the 90th and 95th percentile, with others indicating that the point of regulation should be at the 99.9th percentile or higher.*

1. Comparing DRES 95th Percentile to DEEM 99.9th Percentile

Comment. A number of comments were provided about the DRES-type 95th percentile exposure estimate when residue inputs are treated deterministically and compared to DEEM's 99.9th percentile estimate used when residue inputs are treated probabilistically. Many commenters stated that this was an invalid

comparison and was in some ways deceptive and misleading. Several commenters stated that, contrary to our contention, OPP was taking a more stringent approach by using the 99.9th percentile of a Monte Carlo (or probabilistic) analysis as opposed to the 95th of a DRES-type deterministic assessment.

Several commenters specifically objected to the following table in the document containing a comparison between a previous DRES analysis (performed at the 95th percentile) and a more recent DEEM analysis for a widely-used agricultural pesticide.

Comparison of DRES 95th Percentile Exposure And %aRfD Estimates with Monte Carlo 99.9th Percentile Exposure and %aRfD Estimates for One-Widely Used Agricultural Pesticide

Population Subgroup	Exposure (mg/kg bw/day)		%aRfD[a]	
	DRES 95th Percentile Estimate	Monte Carlo 99.9th Percentile Estimate	DRES 95th Percentile Estimate	Monte Carlo 99.9th Percentile Estimate
U.S. Population	0.005	0.000542	300	32
Infants	0.008	0.000804	480	48
Children 1-6	0.008	0.000905	480	54
Females 13+	0.0036	0.000468	216	28
Males 13+	0.0038	--[b]	228	--[b]

[a]The %aRfD represents the portion of the acute risk cup which is occupied. The %aRfD is obtained by dividing the estimated exposure at any given percentile (e.g., 95th or 99.9th percentile) by the aRfD. It should be remembered that the aRfD may be modified to reflect the decision with regard to the FQPA 10x Safety Factor. Comparison of the estimated exposure to the resulting Population Adjusted Dose (PAD) is then done to determine the acceptability of that exposure.
[b] not calculated

They stated that they did not believe that it was appropriate to compare the DRES and DEEM analysis in that the DRES analysis was based on 1977-78 food consumption data and utilized relatively severe assumptions while the DEEM analysis was based on 1989-92 food consumption data and takes into account many factors such as food processing, percent crop treated, and probabilistic analysis. Specifically, concern was expressed that the comparison table of DRES and DEEM appeared as though these were side-by-side comparisons using the same data when they are not.

One commenter disagreed with the Agency's assertion that the OPP analysis supports the DEEM 99.9th percentile as a less conservative replacement of

the DRES 95th percentile results and stated that he cannot accept this justification for use of the 99.9th percentile. He argues that use of the 99.9th percentile "negates the use of Monte Carlo approaches leading to refined understanding of acute dietary risk by trying to shoehorn the results of advanced higher tier distributional analysis into the old worn shoe of the overly conservative and lower tier DRES approach." He concluded that the Agency "appears to simply be assessing a penalty for use of improved science to arrive at an understanding of exposure." The single example presented by the Agency in the *Percentile Policy* document, he argued, is inadequate to prove the point made, and the Agency's conclusions are couched in terms such as "tends" and "almost invariably" that have limited scientific relevance unless supported by actual data. Similarly, another commenter expressed concern about OPP's statement made in Section I.C.5 of the *Percentile Policy* document that OPP is not taking a more stringent approach by using the 99.9th percentile of a Monte Carlo rather than the 95th percentile of a DRES-type analysis. He believed this statement was not true and that OPP's use of the 99.9th percentile under Monte Carlo is indeed more stringent. The commenter stated that just because exposures using the 99.9th percentile under Monte Carlo methods are generally lower than exposure at 95th percentile under DRES does not mean they are less protective, but means instead that more realistic and accurate data are usually being used in a Monte Carlo assessment. The commenter stated that OPP should not be using the 99.9th percentile under Monte Carlo methods just because 95th percentile under DRES is showing much lower risk estimates in a Monte-Carlo analysis and that this kind of reasoning "flies in the face of reasons for advancing technology and science in any area."

Response. The comment is correct that OPP's DRES calculations used the 1977-78 food consumption survey data rather than the 1989-92 data. OPP agrees that a comparison in which the exposure calculations use the same food consumption data set would provide a more accurate comparison. OPP's revised version of the document illustrates the difference by using the DEEM software (and 1989-92 consumption data) at the 95th percentile to generate a revised table in manner similar to the way a DRES analysis would have been performed so that a more valid "head-to-head" comparison can be seen. The following table reflecting this comparison appears in the revised document[1]:

[1] This table represents a comparison which uses a *different chemical* than was used in the original document and this was done for convenience reasons only. The data from the previous comparison was taken directly from DRES acute system outputs which could not be re-run since the mainframe software is no longer supported. Instead, this new table was generated from the DEEM software currently in use and reflects a direct "head-to-head" comparison using Tier 1 vs. Tier 3 techniques.

Comparison of DEEM 95th Percentile Exposure and %aRfD Estimates From a Tier 1 Analysis to Monte Carlo 99.9th Percentile Exposure and %aRfD Estimates from a Tier 3 Analysis for One-Widely Used Agricultural Pesticide (expressed on a per capita basis using 1989-91 CSFII Data)				
Population Subgroup	**Exposure (mg/kg bw/day)**		**%aRfD[a]**	
	DEEM 95th Percentile Estimate (Tier 1)	**DEEM Monte Carlo 99.9th Percentile Estimate (Tier 3)**	**DEEM 95th Percentile Estimate (Tier 1)**	**DEEM Monte Carlo 99.9th Percentile Estimate (Tier 3)**
U.S. Population	0.0192	0.0013	770	50
Infants	0.0375	0.0007	1500	38
Children 1-6	0.0402	0.0017	1610	67
Females 20+/np/nn[b]	0.0126	0.0011	510	45
Males 20+	0.0119	0.0014	480	55

[a]The %aRfD represents the portion of the acute "risk cup" which is occupied. The %aRfD is obtained by dividing the estimated exposure at any given percentile (e.g., 95th or 99.9th percentile) by the aRfD. It should be remembered that the aRfD may be modified to reflect the decision with regard to the FQPA 10x Safety Factor. This modification results in an acute Population Adjusted Dose (aPAD). Comparison of the estimated exposure to the resulting aPAD is then done to determine the acceptability of that exposure.

[b] Females 20+, not pregnant, not nursing

However, OPP's principle point remains the same: generally, probabilistic analyses which consider and incorporate all available sources of information result in lower estimates of exposure (and resulting risk) at the 99.9th percentile than (un-refined) non-probabilistic techniques do at a lower (95th) percentile. Thus, using the 99.9th percentile to calculate the threshold of concern when highly-refined probabilistic techniques are used does not represent a raising of the bar, but rather a recognition by OPP that when more realistic probabilistic methods are used to generate more realistic estimates of exposure, it is necessary that these facts be considered in deciding what population percentile of exposure should be used.

Further, OPP is not assessing a "penalty" for use of improved science for understanding exposure. What OPP is saying is that if a data submitter wishes to take advantage of more realistic exposure scenarios (*e.g.*, to include use monitoring data or entire range of field trial data, incorporation of percent of crop treated information, and use of residue reduction factors resulting from food processing, among others) which produce more realistic estimates of exposure distributions, he or she will also be held to a percentile of the population's exposure that better represents the full range of exposures (*i.e.*, 99.9).

We remind the commenter that in comparing and contrasting risk assessment and risk management approaches using probabilistic estimation techniques, it is important to consider both the percentile of the distribution as well as the method by which the distribution is calculated. Different data and assumptions yield very different estimates of exposure, and understanding the underlying methodology is as equally important as considering the percentile of the distribution. As indicated in the *Percentile Policy* document, exposure estimates at the 99.9th percentile using the full gamut of probabilistic techniques (*i.e.*, Tiers 3 and 4) are almost invariably lower than corresponding exposure estimates using the more limited Tier 1 and Tier 2 (DRES-type) approaches.[2] The reason for this is readily apparent to those who are familiar with the Tier 1 and Tier 2 evaluation methodologies: a Tier 1 or Tier 2 "95th percentile" assessment does not really produce a 95th percentile exposure estimate because there are so many conservative assumptions front-loaded into the estimate. In the vast majority of cases, our DRES-type "95th percentile" estimate exceeds the actual 100th percentile – *i.e.*, it is likely higher than any individual actually receives. In an attempt to produce better (*i.e.*, more realistic) estimates of exposure upon which to base Agency risk-management decisions, OPP has tried to take advantage of advanced probabilistic techniques and stripped these default assumptions from the analysis. The result, almost invariably, is a lower (but more realistic) estimate of exposure that is more appropriate for use in risk management decisions. By virtue of the fact that exposure estimates at the 99.9th percentile using probabilistic techniques are lower than those produced at the 95th percentile using deterministic techniques that do not take advantage of all available information, OPP believes that the threshold of concern has not, in fact, been raised, but rather that OPP has appropriately adjusted the percentile of exposure considered in recognition of the use of real-world input values.

2. Using the 90th Percentile for Acute Analyses

Comment. Several commenters indicated that OPP should use an allegedly less extreme upper bound for acute dietary exposure such as the 90th percentile as a threshold of concern for regulatory purposes, citing (in part) the need for a consistent federal policy with respect to regulation of risks. One commenter indicated that FDA regulates the safety of food additives and OPP should seek a threshold of concern which is consistent with that used by FDA. Another commenter indicated that EPA regulates other media (air and water) to protect consumers and the FDA regulates food additives under the same statute and criteria (i.e. "reasonable certainty of no harm") that OPP uses to regulate

[2] To date, out of the dozens of acute probabilistic dietary exposure assessments performed, only one exception to this prediction has occurred. This exception involved a pesticide used on single commodity with a very high percent of crop treated.

pesticides. In all of these programs, the commenter stated that a conservative, statistically valid upper 90th percentile of exposure is used as the threshold for regulatory purposes. To apply consistent criteria across programs, the regulatory endpoint used for pesticide regulations should be closer to the upper 90th percentile consumer, certainly not the extreme 99.9th percentile as in the proposed policy.

Response. In comparing and contrasting risk assessment and risk management approaches using probabilistic estimation techniques, it is important to consider both the percentile of the distribution, as well as the method by which the distribution is calculated. OPP's experience has shown that there are a variety of data and assumptions which may be used in estimating the distribution of exposure to pesticide residues in food consumed on a single day and different data and assumptions yield very different estimates of exposure. Thus, understanding the underlying methodology is as important as consideration of the percentile of the distribution. The Agency's draft policy recommends using either the 95th percentile or the 99.9th percentile, depending on the method by which the exposure distribution is calculated. As explained in further detail below, EPA's screening methodology (with which EPA uses 95th percentile) best resembles the methodologies used by other government bodies, for which they use values around the 90th or 97.5th percentiles. Moreover, the screening methodology used by OPP tends to produce exposure estimates at the 95th percentile that are significantly higher than the estimates at the 99.9th percentile of exposure using the refined data methodology (that is, exposure estimates at the 99.9th percentile using probabilistic techniques would be lower than that generated by other government agencies if they were to use non-probabilistic techniques at a lower (*e.g.*, 95th) percentile threshold). Therefore, OPP believes its policy choice in this area is not significantly more conservative than the policies of other government agencies.

Before issuing its proposed paper for public comment, EPA consulted with FDA and USDA, the only other federal agencies that regulate chemical residues in food. Neither reported that it was routinely using probabilistic methodology for acute exposure assessment. Upon further investigation, it appears that the FDA Office of Premarket Approval uses the 90th percentile of *consumption* of foods when evaluating direct food additives. EPA understands, however, that the 90th percentile FDA *consumption* value is only used by FDA for chronic exposures. Thus, it is not directly comparable to the acute exposures proposed to be regulated by EPA at the 99.9th percentile. In fact, when performing the chronic assessments for pesticides that are most comparable to the food additive assessments for which FDA uses the 90th percentile of consumption for eaters-only, OPP uses the mean consumption levels (considering both eaters and non-eaters) which is typically lower than the FDA's 90th percentile consumption.

FDA does not routinely apply probabilistic methods for its acute exposure

scenarios using a formal documented procedure. In addition, EPA knows of no other federal, state, or foreign agency that applies probabilistic risk assessment methodology specifically to acute dietary exposures to pesticides or any other substances. EPA is aware that the Nuclear Regulatory Commission, the U.S. Department of Energy and EPA's Office of Radiation utilize probabilistic risk estimation methodologies concerning different scenarios possibly leading to radiation exposures, but none of those techniques is even remotely similar to EPA's acute pesticide dietary exposure/risk assessment approach.

OPP has also investigated the regulatory threshold used by foreign governments and international bodies. The United Kingdom's and Codex's[3] decision to use the 97.5th percentile is not comparable to EPA's proposal to use the 99.9th percentile of exposure. The Codex 97.5th percentile refers to *consumption*, not exposure. Codex calculates the exposure by multiplying the estimated high-end *consumption* of the single commodity in question (*i.e.*, consumption at the 97.5th percentile) by the MRL (conceptually comparable to the U.S. tolerance). Neither information on the percent of the crop which is treated nor actual monitoring data are taken into account in this assessment. The Codex methodology also calculates exposures on a crop-by-crop basis and fails to account for the fact that other foods consumed by an individual in a day may contain residues. For many of the same reasons that the DRES method overestimates exposure, the exposure estimated under the Codex proposed methodology at the 97.5th consumption percentile will greatly exceed the exposure calculated by the Agency's probabilistic method at the 99.9th exposure percentile in the vast majority of cases (particularly when USDA's Pesticide Data Program data or market basket survey information are available). This is true despite the fact that Codex only considers exposure through one crop at a time, ignoring exposures to the pesticide from all other commodities eaten that day.

In sum, we believe that the selection of the 99.9th percentile when probabilistic risk assessment is used is not inconsistent with the policies and practices of other federal agencies or international bodies. In fact, in cases where comparable situations are present, OPP believes that its more realistic estimates of exposure at the 99.9th percentile would be lower than estimates produced by other organizations' different methods at supposedly lower percentiles.

3. No Significant Gain in Moving from a Point Value to a Distribution

Comment. One commenter thought that the move towards

[3] The Codex Alimentarius is an international organization established under the auspices of the United Nations Joint Food Standards Program of the Food and Agriculture Organization and the World Health Organization. A Codex MRL (Maximum Residue Limit) is the equivalent of a U. S. tolerance.

data-based decision making will be helpful, but expressed concern about this change when the 99.9th percentile will be used. He stated that the tails of the distribution are very sensitive to the shape of the curve and the extent and quality of the data. Since the distribution is unknown and the data are poor, the commenter indicated that a number of assumptions would be needed to calculate the 99.9th percentile. And since the results are going to be very sensitive to the assumptions, there is probably no significant gain in moving from a point value to a distribution. The commenter concluded that "Monte Carlo is better than no Monte Carlo, in principle," but stated that applying this approach with the current state of knowledge using the extremes of the distribution can be worse than not using it at all.

Response. OPP agrees with the commenter that the results may be sensitive to the assumptions (*e.g.*, percent crop treated, residue concentrations in treated crops which have less than Limit of Detection residues, representativeness of sampling procedures, etc.), but believes that as long as: (1) assumptions are well-explained, reasonable, and transparent; (2) sensitivity analyses are performed to determine if any assumptions are "driving" the risk or control the resulting risk estimate; and (3) the resulting risk estimate is properly characterized and incorporates the results of the sensitivity analysis, then the risk estimates are an adequate basis for a regulatory decision. With respect to the commenter's belief that applying this approach using the extremes can be worse than not using it at all, OPP notes that this has not yet occurred. Virtually all exposure and risk estimates generated to date using Monte Carlo techniques (*i.e.*, Tier 3 and Tier 4 analyses) using the 99.9th percentile have resulted in lower estimates than the deterministic techniques (at the 95th percentile) which do not incorporate probabilistic techniques (*i.e.*, Tiers 1 and 2).

4. The 99.9th Percentile Represents Poor Policy

Comment. Another commenter indicated that while EPA is correct in selecting a threshold of regulation above the mean or median exposure and acknowledged that the specific cut-off value is a matter of policy, the selection of the 99.9th percentile represents poor policy as it is not reasonable in relation to the certainty of no harm. He indicated that protecting a binge eater from the health effects of minute doses of a pesticide is not reasonable and that "someone consuming, for example, a stalk of bananas or a flat of tomatoes will have more problems from the acute toxicity of the food constituents and the food itself, than from the pesticide residues on the food." The 90th percentile of acute exposure, he continued, is an appropriate cut-off point to establish a regulatory threshold of concern for pesticides, independently of the basis for the exposure estimate (whether a deterministic or probabilistic model or observational data). Protecting bizarre eating behavior does not meet the definition of reasonable in the new

FQPA standard.

Response. OPP recognizes that binge eating can occur (and is not uncommon among children who may preferentially consume one food or class of food for days at a time) and OPP considers this to be an activity that can, on occasion, occur and should be protected. Nevertheless, OPP believes that it is important that this phenomenon be properly characterized during the risk assessment process and appropriately considered during the risk management process. Therefore, both the *Percentile Policy* document and the revised guidance explicitly consider the concern that extreme consumption values (perhaps the "binge eaters") can potentially drive a specific risk assessment. OPP has indicated that the current software can identify all the individual consumption events which lead to high pesticide exposures. If the "tails" of the exposure distribution consist mainly of unusual, unrepresentative, or suspect reported consumption values, this will be fully described in the risk assessment for the risk managers in OPP. In any case, all such conditions and resulting decisions will be fully and clearly explained in the risk assessment document and can be reviewed and commented upon by the regulated industry, public interest groups, and the public at-large. Given the careful scrutiny USDA has given the data, OPP does not feel it is appropriate to conclude, *a priori*, that specific consumption values should be discounted or removed from the data set prior to a full analysis of the data and appropriate consideration of the implications of such removal in the context of the risk assessment. In any case, OPP notes that, for many of the risk assessments completed to date, "binge eaters" do not appear to be driving the risk assessments at the 99.9th percentile.

5. Alternative Approach

Comment. One commenter recommended that OPP take an alternative approach that would avoid relying on the extremes of food consumption data. Specifically, EPA's concern should, according to the commenter, address the "maximum amounts of residues" on "reasonable amounts of food consumption."

Response. EPA finds the comment unclear, both with respect to the "maximum amount of residue" and "reasonable amount of food consumption." EPA disagrees with such an approach because it would mean ignoring reliable data and producing an exposure estimate that is less realistic and representative than the estimates produced using probabilistic methods. In fact, the proposal seems to suggest that the Agency should move *away* from probabilistic assessments which attempt to capture the full distribution of exposures across the entire population.

OPP believes that with USDA's CSFII consumption survey data, the best information is available regarding what people eat. The USDA data are available

to address actual reported food consumption and actual diets from thousands of interviewed individuals. It is not a system that has to "make guesses" about what people eat and when they eat it. Any uncertainties which may exist in the high-end consumption levels reported in the CSFII are not so significant (and can be dealt with) so as to warrant such a revamping of the current system.

The current methodology using CSFII and the best residue data available will more accurately reflect real exposures occurring to the population than the methodology recommended by the commenter. In addition, FDA's monitoring program and USDA's PDP program data are available to address real residue levels which occur following harvest or (preferably) immediately prior to distribution to supermarkets and grocery stores. These two programs have analyzed thousands of samples for a variety of commodities for different pesticides. The commenter seems to be suggesting that the Agency should revamp its exposure evaluation methodologies which for the most part do (or at least can, if the proper information is made available) reflect actual exposure levels.

6. Use a Cost-Benefit Approach to Regulating Risk

Comment. One commenter encouraged the Agency to take a cost-benefit approach to regulating risk and consider a lower percentile for regulation. Citing a shoe manufacturing analogy, the argument was made that economic and practical considerations preclude society from attempting to cover all contingencies and for this reason shoe manufacturers choose to target a smaller percentage of the population (*e.g.*, the commenter suggested 95%) by making sizes available to only a limited portion of the population. Given that costs tend to increase dramatically as a larger and larger portion of the population is attempted to be fit, the commenter asked if more and more resources should be expended to attempt to cover individuals in the tails of the distribution (*e.g.*, he states >95th percentile). For this reason, the commenter indicated his belief that it would be appropriate to use the 95th percentile.

Response. The Agency considers the analogy inappropriate. OPP believes that public health agencies have a responsibility to regulate for health effects at a standard higher than private industry does in selecting what percentage of the population to serve and the comparison made by the commenter for this reason is not entirely valid. It is EPA's goal that pesticide residues in the American food supply be safe. Regulating safety of the food supply in the same manner that decisions are made about the range of shoe sizes to produce is not appropriate.

The laws under which EPA regulates the safety of pesticide residues in food provides that OPP must assure "reasonable certainty of no harm." Consistent with this statutory standard, EPA is not allowed to balance risks and benefits.

Thus, the Agency must decide, using its expertise in risk assessment and its judgment about safety, what maximum levels of exposure are appropriate and in accordance with this statutory standard. For reasons explained in the *Percentile Policy* document and elaborated in these response to comments, the Agency chose generally to use the 99.9th percentile in calculating a threshold of concern in connection with probabilistic risk assessments.

7. Use the More "Resilient" 95th Percentile

Comment. Another commenter recommended that the regulatory threshold of concern utilize the "more resilient" 95th percentile. This comparison would utilize all the necessary safety factors (*e.g.*, inter- and intra-species and FQPA, as necessary). As an added check, the commenter recommended that exposure at the 99.9th percentile be compared directly with the NOAEL (i.e, with no safety factors) to ensure to the vast majority of individuals are not experiencing exposures above the laboratory-derived NOAEL. The commenter recommended that those individuals who are above the 95th percentile but below the 99.9th be examined separately, looking for clusters of commonality or typical consumption patterns leading to high exposure estimates and that this information should be used to target risk reduction for high exposure behaviors. The Agency could then focus risk reduction programs, including public education toward the particularly risky behaviors.

Response. OPP does not believe that the NOAEL is the appropriate point to use in risk assessment. Current policy is that a 10X factor be applied for potential inter-species variation and 10X for potential intra-species variation to arrive at a "safe" dose (*i.e.*, the RfD). FQPA also calls for the use of an additional factor of 10 to account for the completeness of the toxicity and exposure databases and the potential that children may be more sensitive than adults. FQPA also allows the use of a different factor if OPP concludes that a different factor would be protective. The commenter has not advanced any persuasive reasons as to why OPP should depart from this long established approach in this circumstance by comparing the estimated exposure directly to a NOAEL.

The commenter also recommended that OPP examine those individuals which are above the 95th percentile but below the 99.9th, looking for "clusters of commonality" or typical consumption patterns leading to high exposure estimates and that this information should be used to target risk reduction programs, including "public education toward the particularly risky behaviors." OPP does in fact look for risk "drivers," but at this time only does so for those individuals whose exposure is estimated to be above the threshold of regulatory concern to determine if reported consumption levels are reasonable and to target risk mitigation activities. It is unclear to the Agency what the commenter meant by

targeting risk reduction programs including "public education toward the particularly risky behaviors." It is OPP's ultimate goal that all food should be safe to eat and OPP believes it would be inappropriate public health and food safety policy for government risk reduction activities to be limited, for example, to specific warnings to the public that might include such things as "avoid eating green beans and citrus fruits on the same day." In addition, the statutory standard requires that there be "reasonable certainty of no harm" - - not that there be "reasonable certainty of no harm only if certain combinations of fruits and vegetables are avoided."

8. The 99.9th Percentile Is Appropriate or Should be Raised

Comment. One commenter supported the 99.9th percentile as an appropriate point to regulate:

> With regard to the percentile value for regulation, numerical reality is relatively straightforward, and is described in the draft policy paper in Section III.
>
> The size of the exposed population potentially exceeding the PAD [Population Adjusted Dose, considered to be acceptable] at the 99th or 95th percentile would be 10 and 50 times larger, respectively, than the number at the 99.9th percentile.
>
> Even for population subgroups, 0.1% represents large numbers of individuals; for example, this portion represents approximately 23,000 children age six and under...

The commenter continued by stating that there are additional considerations to reinforce the message that Agency acute dietary risk assessment are not overly conservative; for example, while the paper addresses only acute dietary exposure, the Agency is required to conduct aggregate exposure assessment from all reasonable pathways. In addition, the Agency's acute dietary risk assessments have only been conducted on individual pesticides, and have not been extended to cumulative organophosphate assessment. Such considerations, the commenter stated, argue that the Agency should neither weaken its current policy of using the 99.9th percentile nor should it use a weaker benchmark for cumulative risk from the organophosphates.

Another commenter strongly objected to OPP's stated goal of assuring that in the case of acute dietary risk assessment, exposures to each pesticide chemical are regulated down to the level at which the individual at the 99.9th percentile level of the risk distribution meets his or her personal PAD or RfD for that chemical.

The commenter stated that this goal totally ignores the cumulative exposure mandate of FQPA and that using the 99.9th percentile in calculating the threshold of concern for any of the major OP's to the 99.9th level, even if met, would leave some 25,000 children over their RfD on a daily basis from exposures to that single OP. Given the likelihood of dietary exposures to three to eight OP's in any given day, this approach, the commenter contended, will "fall far short of the FQPA's mandate." The commenter stated:

> Given the gaps in EPA's knowledge of residues in food and water and even more spotty data on other exposure pathways, we are certain that there will be a substantially greater number of children over their RfD on any given day, and many by a wide margin, if the goal of regulation remains just reducing exposures to the 99.9th level one chemical at a time. This policy must be rejected.
>
> The Agency should strive to assure, as an initial step toward the FQPA's risk reduction mandate, that dietary exposures are reduced such that 100 percent of children eating day episodes result in exposures well within the allowable risk cup for any individual chemical. While further risk reducing steps may later be needed to meet the cumulative risk reduction goal, the above initial goal will clearly focus attention on the high-risk foods and encourage growers and the industry to take far more seriously the need for change in pest management systems.
>
> Instead of applying the 99.9th percentile goal to levels in the distribution of risks to one chemical at a time, the Agency should instead just apply the 99.9th percentile goal to the distribution of risk estimates produced as a result of cumulative acute dietary risk assessments. If the 99.9th percentile goal were applied in this way, the Agency would be able to argue forcefully that it had relied on the best data and risk assessment science available to assure that the 99.9th percentile of the eating day episodes for all infants and children result in total exposures below the level of concern. Applying the 99.9th percentile goal in this fashion is the most defensible approach statistically in the case of the OPs. The cumulative OP risk assessment will no doubt draw on a large Monte Carlo run, entailing millions of simulated child eating days, drawing on a very large residue database, especially after PDP and other composite data is decomposited. The enormity of this dataset, and the richness of the food consumption and residue data underlying it, will produce a much more realistic and reliable distribution of residues than ever before possible.

Similarly, another commenter indicated that, although the *Percentile Policy* document contained "generally sound principles that the Agency can rely on to set a regulatory floor," "...the document contains a disturbing bias towards traditional notions of risk assessment and a cavalier insistence that infants and children are more than adequately protected by current regulations." In addition, the

commenter stated:

> A policy is inherently flawed if it protects less than 100% of the population. The Agency plainly recognizes this fact and attempts to justify its 99.9 percent proposal by invoking what it sees as the inherent conservatisms of the risk assessment process. Unfortunately, the Agency has portrayed its risk model as being much more conservative and certain than it is. While it is theoretically conceivable that allowing 0.1 percent of the population to exceed the RfD on any given day could meet the reasonable certainty of no harm standard in the FQPA, this possibility should only be considered after a pesticide was in full compliance with all standards of the Act.

The commenter continued, indicating that the "risk assessment methods currently used by the Agency do not come close to the requirements of the law" in that they consider neither aggregate exposure to a given pesticide via various pathways nor cumulative exposure to multiple pesticides with a common mechanism of toxicity. The commenter concluded that "risk assessment methods that do not meet legal requirements can hardly be considered conservative" and that "until such time as Agency risk assessments meet these statutory requirements, we recommend that the Agency adopt a higher than 99.9 percentile regulatory threshold for individual acutely toxic pesticides."

Response. OPP believes that the methods used to estimate the distribution of exposures and evaluate the resulting risks are adequately conservative such that using the 99.9th percentile ensures that there is "reasonable certainty of no harm" particularly since the aPAD used as a toxicological benchmark is usually between 100 and 1000 times lower than that dose which caused no observable adverse effect in laboratory animals.

OPP recognizes the commenters concerns, but, at this time, the current 99.9 policy applies to daily exposures to a (single) given chemical through the acute food pathway only. It is considered to be a "first step" toward regulation of exposures on an aggregate, and then cumulative, basis. OPP believes that different types of risk assessments will be needed for aggregate and cumulative evaluations and these assessments will also be associated with regulatory thresholds of concern which are analogous to the threshold for acute risks from food and the range of 1×10^{-6} for carcinogenic risks. Although OPP is moving toward this direction of regulating on the basis of probabilistic aggregate and cumulative exposure assessments, a decision has not yet been made as to the appropriate threshold of concern for these types of assessments.

9. Sliding Regulatory Scale

Comment. Several commenters responded to Question #4 regarding a

sliding regulatory scale. Briefly, the question asked if a "sliding regulatory scale" with more serious toxic effects regulated at higher thresholds might be an appropriate basis for regulation; some contend that such an approach would explicitly acknowledge all aspects of the risk management decision and incorporate the nature of the toxic effects and the built-in conservatism on the hazard identification and dose response side of the equation. Instead of using only a single percentile for all toxicological effects (regardless of severity), OPP asked if it should assess pesticides at a variety of percentiles, depending upon the toxic effect observed or assess pesticides at different percentiles according to the nature and size of the subpopulation groups.

Several commenters addressed this question by stating that EPA should address the issue of differences in the nature of endpoints in the process of setting the RfD and PAD rather than changing the exposure percentile to reflect the nature of the toxicity. Several stated that this method would more appropriately account for severity of effect during the establishment of the safe dose, and that applying different thresholds of concern for effects with different severities contradicts the notion of a safe dose that is inherent in the FQPA standard. In other words, a sliding scale is one possible way to address the issue of minor effects, but the scale should be based on changing the PAD or a RfD used in the assessment rather than simply changing an arbitrary threshold value since changing the threshold percentile implies that somehow the exposure changes as the toxicological endpoint changes. Another commenter stated that the concept of a “sliding regulatory scale” for toxic endpoints appeared unnecessary and redundant since these special considerations are already accounted for during the determination of the reference dose (*i.e.*, the appropriate safety factors are used to calculate the reference dose and it is therefore not necessary to allow for a differential dietary risk assessment employing a percentile that is a direct function of the toxic endpoint).

Response. Apart from the case when children and infants display special sensitivity and an FQPA factor can incorporate the nature of the effect, OPP does not give the nature of the toxic effect special consideration in endpoint selection except when the effect is unusually serious. The nature of the toxic effect (or its severity) does not influence either of the two standard factors for intra- and inter-species variability.[4] The commenters are suggesting that the process through which the RfD is derived should be altered so as to incorporate the nature and the severity of the effect. Currently, the RfD is established at a safe level which

[4] Occasionally, an extra factor may be added if the endpoint is extreme.

assures "reasonable certainty of no harm."[5] The *Percentile Policy* document deals with the threshold of regulatory concern in terms of *exposure*. The selection of the appropriate NOAEL and derivation of the RfD is a *toxicological* issue that is outside the scope of the document.

10. Sliding Scale and Risk Management

Comment. A number of comments on the "sliding regulatory scale" supported incorporating the nature of the effect in any risk management decision. Rather than adjusting the RfD to account for differences in effect, however, these commenters suggested instead that these differences in severity of toxic effect be one component, of many, in the risk management decision. A number of criteria for consideration in a "case-by-case" approach were suggested and a flexible approach was encouraged based on both the nature and severity of the toxic effect and the overall exposure and risk situation.

Several commenters addressed this question by indicating a preference for adopting a range of percentiles (*e.g.*, 95th to 99.9th), with the use of the highest percentiles reserved for cases where the highest percentile values are not driven by implausibly high consumption or residue values derived from inappropriately small sample size for the particular combination of subpopulation and commodity. One commenter believed that the Agency should reserve for itself a reasonable degree of discretion that would allow OPP to avoid criticism that would inevitably result if it chose a lower percentile such as the 95th percentile or a very high percentile such as the 99.5th or 99.9th. This flexible approach would also be responsive to repeated cautions by the FIFRA SAP about problems with use of high percentile estimates. The commenter stated, too, that a flexible approach is inherently sensible: it does not make sense to regulate with the same rigor against potentially fatal effects, minor effects, and non-adverse effects, nor to come to the same conclusion in all cases regardless of differences in the potential for significant overestimation errors in the underlying exposure data.

Another commenter seemed to indicate a preference for a lower threshold of concern for minor and reversible effects. He stated that the toxic effect and the richness of data supporting a regulatory understanding of toxicity and exposure potential should be factored into any decision regarding the outcomes of an acute dietary exposure assessment and that considerations of reversibility and severity of effect should strongly influence regulatory endpoint setting and risk management decision making in the case of acute dietary risk assessment. Specifically, the commenter argued that OP insecticides are currently evaluated from a toxicity

[5] Normally, this is established at a level between 100 and 1000 times lower than that dose which caused no observable adverse effects in laboratory test animal.

perspective on the basis of a reversible biomonitoring exposure endpoint (*e.g.*, plasma cholinesterase inhibition) rather than on the basis of a toxicological endpoint (red cell or brain cholinesterase inhibition) that is associated with an adverse effect. In cases such as this, the established toxicological endpoint and its meaning should lead to acceptance of a lower threshold of concern with greater certainty that adequate margins of safety are being maintained.

One commenter provided comments on the choice of the appropriate, reliable percentile of the model's output to be used for regulation of risk. He suggested that selection of the appropriate percentiles should be made on a case-by-case basis using good scientific practice. Consistent with EPA's risk characterization policy, a risk manager should have available an understanding of all the key factors contributing to a risk, including:

- the level of confidence in the input data
- the reality of the exposure scenarios
- the sample size
- the size of the population
- the toxicity of the substance
- the nature of the toxic effects
- the application of default assumptions in the assessment
- and other factors of the particular case

The commenter urged the Agency to convene a panel, drawing on expertise outside the Agency, which would completely re-write the guidance to incorporate a scientific, case-by-case approach to selection of an appropriate, reliable percentile of a model's output to be used for regulation of risk. The commenter welcomed the opportunity to work with EPA and others in the scientific community to develop this revised guidance and encouraged OPP to continue to seek public comment and expert peer review in developing this guidance.

One individual responding to the question on regulating risk on a sliding regulatory scale indicated that he was unsure of the form that this proposed alternative would take. The commenter indicated that he could conceivably support such an approach since it is closest to the preferable case-by-case approach regulating more serious toxic effects at higher percentiles of exposure (*i.e.*, on a sliding scale) and allowing risk management on a case-by-case basis looking at the available data. However, the commenter stated, while toxic effects should be taken into consideration in regulating, they should not be assigned arbitrary values, as this would inappropriately imply scientific precision, which does not exist with value judgements. Assigning a more serious effect an arbitrary number is not a science issue, but a social policy issue based on perceived social

values. The commenter cited a previous OPP document entitled "Acute Dietary Exposure Assessment Office Policy" (U.S EPA, 1996) which directs that margins of exposure are to be calculated for a range of exposure percentile levels, and that the selection of an MOE that triggers a risk concern should be tied to the nature of the adverse effect under consideration and the type of study from which the No Observed Effect Level (NOEL) is taken. Effects that are reversible, the OPP document states, may be regulated less stringently than those that are irreversible and life-threatening, and dose-response information is also a consideration. The commenter cited this as a preferable route (rather than assigning arbitrary values at which to regulate) and supports this previous OPP position that all of the information should be considered on a case-by-case basis, and that the choice of what population percentage to use is based on a full consideration of risk, not solely on consideration of hazard or exposure.

Another commenter objected to regulating on a sliding scale which would incorporate severity of effect, stating that this would be difficult to do and would set back attainment of FQPA goals:

> This is a set of ideas perhaps 50 years ahead of its time. After a full and complete set of endocrine system, immune system, and developmental toxicity tests have been developed and verified, and carried out on all pesticides used on food, it might be useful to revisit this suggestion. Then, many years of expert advisory panels, at least one NAS review, and many consensus building activities among stakeholders will be required to forge agreement on how to rank health impacts on a relative scale, a necessary step to implement this idea. As sound as this suggestion might seem conceptually, the considerable technical and political challenges inherent in implementing it would set back attainment of FQPA goals by at least 20 years, and for this reason alone, the suggestion should be rejected.

Response. OPP has carefully considered the comments on whether a formal "sliding regulatory scale" should be adopted which would explicitly lead to the regulation of pesticides with less serious, reversible effects evaluated at a lower threshold. OPP believes that it is not appropriate at this time to identify various specific percentiles for each of the many disparate toxicological effects because there are many other factors which could potentially be incorporated into a decision to use a different threshold in a regulatory decision.

OPP does believe, however, that nature, severity, and reversibility of the effects caused by a pesticide are important types of information to include in the risk assessment for consideration by the risk manager and his or her selection of an appropriate regulatory threshold. We agree with the comments that OPP should retain some discretion to choose a different percentile as a threshold of regulatory concern. We also agree that we should consider a number of criteria which should

be fully characterized in the risk assessment. In this manner, the risk manager can evaluate how supportable the 99.9th percentile exposure estimate is and evaluate whether or not it is appropriate to deviate (up or down) from the 99.9th percentile. An adequate characterization could include, for example, the exposure estimate's perceived degree of conservatism considering in particular the identity of the risk "drivers," the reliability and characteristics of the input data, the size of the affected populations, the results of a sensitivity analysis, etc. Specifically, a full and adequate characterization of the risk estimates might include a review of the following (in approximate order of relative importance):

- whether a high-end consumption value acts actually acts as a "driver" in the risk assessment. *(in many cases, high-end consumption values may not be actual "drivers" (*i.e., *significant contributors) in the risk assessment and thus may not be the primary reason behind high estimated exposures at the tails of the distribution)*

- how extreme the upper tails of the consumption curve are. *(for example, is the 95th percentile consumption value greater than four times the mean consumption?; is the 99th percentile value greater than six times the mean consumption?)*

- how far the high-end consumption value is from where it would be expected to be given the pattern (or distribution) of reported consumption values in the lower percentiles. *(*e.g., *if a distribution can be reasonably established for the reported consumption values in the lower percentiles (*e.g., *70th through 95th percentiles), how extreme would the high-end value be in an appropriate Q-Q or other statistical plot)*

- the size of the affected subpopulation (and the statistical weights applied) and how likely exposure estimates for the subpopulation would be subject to undue effects of reported high-end consumption values. *(a high-end value would be expected to have more influence on the upper-end exposure estimates in a small subpopulation than it would in a large subpopulation)*

- from a dietary standpoint, how likely the high-end value is to be a valid reported consumption event. *(for example, although they may be equally extreme from a probabilistic standpoint, consumption of three gingko fruits in a day might be considered more reasonable than consumption of 10 apples)*

- ❖ the nature of the inputs both in the overall assessment and (particularly) for the drivers. *(this would include, for example, whether input residue data included field trials vs. PDP data vs. market basket survey data; the use of default vs. actual processing factors; extent to which single-serving values are measured vs. established by decompositing[6] , nature of percent crop treated data, etc.)*

- ❖ comparison of exposure and consumption estimates using the 1989-91 data and 1994-96 data. *(if both the 1989-91 and 1994-96 CSFII data sets produce similar estimates of exposure and contain similar extremes of consumption, it is more likely that the high-end reported consumption is indeed an actual value)*

In sum, OPP believes that the risk assessor should adequately characterize the nature of the assessment (including any biases and uncertainties) and to perform a sensitivity analysis, where appropriate, such that the reasonableness of the upper-end percentile estimates (including the 99.9th) can be properly gauged. Any risk assessment performed by OPP should characterize the effect of any high-end points (on the consumption) on the regulatory percentiles of possible regulatory interest. Likewise, it is important for the risk manager, in turn, to consider the entire set of data and information available in deciding if the 99.9th percentile is an appropriate demarcation point for use in risk assessment. In particular, any risk management decisions should consider the effect of any high-end data values (consumption or residue) or other relevant factors and, when appropriate, be flexible with respect to the population percentile used. Nevertheless, based on the several dozen risk assessments and sensitivity analyses we have performed to data using probabilistic techniques, we do not expect this review to warrant a departure from the 99.9th percentile in the vast majority of cases.

C. Monitoring vs. Modeling

Overview. *Rather than relying on modeling exposures through food, a more direct way of assessing exposures is developing monitoring programs to determine if exposures are at acceptable levels. This section addresses the comments received on this aspect of the paper.*

1. Measure Population Distributions

[6] "Decompositing" is a mathematical procedure used by OPP to produce estimates of pesticide residue levels in single items of produce based on the distribution of residues measured in composite samples where the residues measured in the composite samples represent average residues in a group of generally ten or more items.

Comment. One commenter encouraged the Agency to measure population distributions of acute pesticide exposures rather than modeling them, indicating that measurements will prove more accurate and more precise than any modeled estimate. Registrants could acquire and submit scientific measurements of exposure to a pesticide already in use by directly sampling a survey population, either by exposure monitoring or by measuring biomarkers of exposure.

Response. OPP recognizes the value of measuring actual pesticides and their metabolites in populations in assessing real-world exposures to pesticides. These data tend to be most useful in assessing and judging the accuracy of exposure models, rather than regulating pesticide use *per se*. While actual measurements of pesticides measured in body fluids may represent the best data for evaluating exposures, there are many situations in which biomonitoring is limited in what it can measure and the use of models is necessary. For example, modeling permits the exposure assessor to consider[7]:

Unmarketed Pesticides in Development: By definition, the population has not been exposed to these pesticides and biomonitoring would not be useful;

Temporal Flexibility: Modeling can be used to assess future time periods and the hypothetical "what if" situations necessary for risk mitigation activities. This is not possible with biomonitoring;

Source Attribution: Modeling can attribute specific exposures to specific pesticides and (through its flexibility and "what if" capabilities) specific use practices. With biomonitoring alone, it is not possible to indicate whether risk is attributed to, for example, use of Pesticide X on blueberries grown in the Northeast, the use of Pesticide X on apples in the Pacific Northwest, or the use of Pesticide Y (if these pesticides have similar metabolites) on almonds in the West;

Inclusion of More Chemical Species: Biomonitoring is available for only a limited number of pesticides for which analyses can be performed; and

Representation of Long-Term Conditions: Modeling can assess long-term time scales needed for assessment of chronic exposures. Personal monitoring studies for such assessments are more invasive and burdensome.

[7] Derived from "Total Risk Integrated Methodology (TRIM) Exposure-Event Module Development" TRIM.Expo Technical Support Document. DRAFT Report. EPA OAQPS. August 27. 1999.

Despite these limitations, EPA has considered biomonitoring data to establish the prevalence of pesticide exposures in the general population. For example, a 1995 Centers for Disease Control and Prevention study found a metabolite of a common household insecticide in the urine of 82 percent of the 1000 people monitored (Hill *et al.*, 1995) This study was conducted to establish reference concentrations for adults in the general population of the United States. The EPA's National Human Exposure Analysis (NHEXAS) program and the U.S. Public Health Services' National Health and Nutrition Examination Survey (NHANES) studies will provide useful biomonitoring and other information when they become available.

A three-day workshop, sponsored by the International Life Sciences Institute, was held in October, 1999 to discuss model evaluation and validation and included discussing NHEXAS/NHANES data and how it might fit into OPP's modeling scenarios. The data that are being generated under the auspices of NHEXAS and NHANES are likely to provide useful information that will be used in conjunction with other data to address the exposure issue.

D. Data Quality/Uncertainty

✔ Limited Size of USDA CSFII Database
✔ Precision Limits of USDA CSFII Database with Respect to Upper Percentiles
✔ Identifying and Handling Outliers
✔ Uncertainty

Overview. *A number of individuals commented on the data quality and uncertainty aspects of EPA's probabilistic assessments. Specifically, issues were raised concerning whether the size of USDA CSFII food consumption survey was adequate for regulating exposures at the 99.9th percentile as well as how high-end values (both from the USDA survey and from residue field trials and monitoring studies) could impact OPP's risk assessments.*

1. Limited Size and Potential Usefulness of the CSFII Survey

Comment. Many commented on the limited size and potential usefulness of the USDA CSFII survey. They said that these limitations should preclude any decision to use a level as "extreme" as the 99.9th percentile of exposure. Several commenters cited USDA remarks that the most recent 1994-96 CSFII sample is not of sufficient size to report intakes at levels as high as the 99th percentile for all sex-age groups and foods and the statement that "the USDA does not believe that the consumption data is reliable to predict percentiles in excess of approximately 95%." The commenters recommended that OPP follow USDA and model developer guidance and use the consumption database within the limits that have

been recommended. Another commenter specifically mentioned these guidelines as ones that are issued by the National Center for Health Statistics for presenting estimates of upper percentiles of consumption and nutrient intake distributions.

Several commenters brought up the issue of inadequate precision in the upper percentiles. They stated that even though CSFII data do permit estimation of high-end consumption in the tails of the distribution, the characteristics of the food consumption distribution result in confidence intervals that become wider and less symmetric as the percentile increases from 95th to 99th for the intake of many foods such that the true intake becomes harder and harder to estimate with precision. Another commenter echoed these remarks and provided some background details on how the sample sizes were originally selected by USDA. The commenter stated that the sample group sizes were chosen to meet precision levels from a nutritional standpoint [8], not from the standpoint of adequate precision in determining the range of individual consumption of individual foods. Although great care was taken to survey enough persons in each group to represent various subpopulations for the nutrition-oriented purposes of the survey design, the number of sampled persons, the commenter argued, would have had to have been much higher if the selected precision levels had been expressed in terms of individual consumption of individual food items. The commenter also indicated that the survey did not attempt to sample a statistically representative number of infants under one year old; it simply included each infant in a household that also included one other sampled person. Likewise, the survey was not designed to be statistically representative with regard to the distinction with regard to nursing infants or non-nursing infants.

Several commenters expressed concern about the perceived OPP policy position that it is possible to multiply an (unreliable) consumption distribution by an (unreliable) residue distribution and obtain (after summing all exposures from a given day over an individual) an exposure distribution which is transformed into something that is reliable. Another commenter indicated that consumption data are reliable to about the 90th-95th percentile according to sample size and that this distribution is multiplied by residue values that could be considered to be reliable up to that degree or less. Simply because one multiplies one distribution by another to make more exposure points does not make the data that went in or the results that come out any more reliable or accurate.

---

[8]Specifically, sample sizes were not selected to define high-end nutritional parameters. Rather, the sample group sizes were selected such that the coefficients of variation for mean saturated fat and iron intakes would be 3% or less for each of the 20 all-income sex-age domains and to be 5% or less for each of the 20 low-income domains. The study's size was selected not to define high-end nutritional parameters, but rather to have adequate precision for an estimate of the mean consumption in two nutritional categories (fats and iron) that come from a variety of foods, not a precision in determining the range of individual consumption of individual foods.

One commenter stated that there is a need for input data to be statistically reliable at levels greater that the 99.9th percentile for the exposure to be reliable at 99.9%. To have confidence in the estimated exposure occurring at the 99.9th percentile, the commenter stated that one must have even greater confidence in the input distributions. The commenter provided an example from which he concludes "[c]ertainty in the outcome for the 99.9th percentile (a joint probability of 0.999) requires that the individual probabilities of occurrence be certain at the 99.95th percentile ((0.999 = 0.9995 x 0.9995), when the residue concentration and food consumption are represented by a parametric distribution function and are not correlated." He went on to state that since an acute dietary exposure assessment involving multiple commodities is a summation of this simple case, it follows that the requirement for input confidence at the 99.95th percentile holds for each commodity and residue distribution considered. The commenter concluded that one cannot improve the certainty in a predicted result by combining less certain inputs, and the certainty in a predicted exposure can be no greater (and in fact will be less) than the certainty of the inputs for residues and consumption. The commenter indicated that he had used both bootstrapping and two-dimensional Monte Carlo analysis to model residue and consumption distributions to evaluate this effect and had concluded that high-end uncertainty in input distributions is retained or compounded in high-end uncertainty in the output exposure distributions. The degree to which the uncertainty is compounded will be a function of the nature of the input distributions themselves as well as the number of food items considered in the acute dietary assessment.

Response. The reliability of estimates generated using probabilistic techniques at the 99.9th percentile has been an area of major controversy and confusion.

First, this issue has become more confused by USDA's statement that the CSFII data cannot be used to reliably predict *consumption* percentiles in excess of 95% (*i.e.*, USDA has stated that consumption estimates at percentiles greater than the 95% have a large amount of uncertainty associated with them). USDA has stated that for certain foods in certain sex-age groups, the CSFII consumption data are not of sufficient size to report intakes at levels as high as the 99th percentile since there are too few observations to report statistically sound or reliable estimates. USDA subscribes to the analytical reporting guidelines developed by the National Nutrition Monitoring and Related Research Program when reporting its Continuing Survey of Food Intake by Individuals (CSFII) food intake data. These guidelines serve as the requirement for the reliable reporting of survey data and represent conditions that yield the most sound statistical conclusions. In part, they require a minimum sample size for reporting and annotation of less reliable values due to sample size limitations. Thus, USDA highlights those consumption

percentiles for which there is considerable more uncertainty than is desired.[9]

Any concern about the adequacy or size of the USDA CSFII data base should be directed, however, not at whether the data sets are adequate to define high-end percentiles of *consumption* of a specific food by a specific population subgroup, but rather whether the databases are sufficiently large to adequately characterize the distribution of daily pesticide *exposures* from all food that an individual eats in any given day. OPP will use the 99.9th percentile of *exposure* (not consumption), which incorporates both the (admittedly) full distribution of CSFII consumption data and the distribution of field trial or USDA/FDA monitoring data. The distinction between consumption and exposure is critical. For any given subgroup (*e.g.*, children 1-6) and any given commodity (*e.g.*, kiwifruit), there may indeed be too few 1-6 year old children eating kiwifruit to make a reliable estimate of "high-end" kiwifruit *consumption* by 1-6 year old children. USDA and OPP are in full agreement on this issue. Nevertheless, this does not necessarily mean that estimates of "high-end" *exposure* of all children 1-6 (kiwifruit and non-kiwifruit eaters alike) are also unreliable. The reliability of the estimates of high-end exposures would, in part, be determined by an array of other factors including the number of other potentially treated commodities eaten, the percent of the commodities eaten which data predict would contain residues, and the residue levels in the treated commodities that are eaten. As acknowledged by another commenter

> ...Whether the potential for significant overstatement of risk lessens as more foods are added depends on the residue values for the foods that are added and on the consumption values for those particular foods.

---

[9]Specifically, the Interagency Board for Nutrition Monitoring and Related Research (IBNMRR) recommends that "the quantity values at a tail percentiles, P, (*i.e.*, $P \leq 0.25$ or $P \geq 0.75$) should be marked with an asterisk when the minimum of nP and n(1-P) is less than eight times a broadly calculated design effect."

Certainty is not something which is a concept which is "turned on" and therefore"present" below a specific value and "turned off" and therefore "not present" above that value. It represents, instead, a gradation: as one moves away from the mean value of a distribution toward the extremes of the distribution, uncertainty increases (the phenomenon of "ever widening confidence bands"). The least certainty, admittedly, exists at the tail ends of the distribution. However, as recognized by the IBNMRR:

> *It is important to remember that these guidelines* [on sample size] *are not absolute. They represent conditions that yield the most sound statistical conclusions. Violating these sample size guidelines (or other criteria included in the larger report) introduces a greater degree of uncertainty about the soundness of the analytic conclusions, but not does not necessarily mean that a particular analysis is invalid. Subject matter knowledge, as well as the survey design and the analytic approach, are required to judge the merit for each use and interpretation of data for a particular survey or surveillance system.*

This commenter continued:

> OPP also argues that the process of combining (multiplying) consumption values and residue values also tends to cure problems caused by attempting to make high-percentile estimates from a consumption database that is too small...For any given erroneous or otherwise unrepresentative consumption value, multiplying it by a residue value will yield a correspondingly unrepresentative exposure value.[10] Will an overestimated exposure value be sufficiently diluted by other exposure values so that it has no effect on the estimated value at the 99.9th percentile? Again, we think the answer will depend on how many consumption values are unrepresentative and by how much; there is no categorical answer.

To date, OPP has rarely found that a single high-end , e.g., >99th percentile, purportedly uncertain *consumption* value for a single commodity is completely (or even significantly) responsible for "driving" the risk at the 99.9th percentile of *exposure*. OPP has laid out in its proposed 99.9th percentile policy the steps that would be taken to determine if this has occurred. In those cases where this is demonstrated to occur, OPP describes how it will incorporate this consideration into the risk management decision and any required risk mitigation measures.

However, OPP recognizes the *potential* for overestimation of exposure in those instances where the high-end ("tail") exposures are derived from high-end (and "uncertain") consumption estimates. In the *Percentile Policy* document issued for public comment, OPP indicated that it would investigate the consumption patterns of those individuals who are present in the high-end tails of the exposure distribution. That is, for individuals identified as comprising these high-end exposure tails, OPP would conduct a sensitivity analysis to determine if a high-end (and therefore potentially uncertain) consumption value was responsible for "driving" the exposure estimate for this individual. If the preponderance of persons located in the exposure tails of the distribution were consuming unusual amounts of food and OPP believed that these amounts were unreasonable or unrealistic, then a decision could be made that the reliability of the 99.9th percentile of exposure was suspect and an appropriate risk management decision could be made. To date, when these sensitivity analyses have been performed, OPP has <u>not</u> found that the major contributors to RfD-exceedences are foods for which unusually high amounts are consumed. OPP's experience so far indicates the amounts consumed that lead to predicted exceedences are not unusual, but instead

[10]OPP notes that this is only true if the unrepresentative consumption value is multiplied by non-zero residue value. For example, if only 10% of the crop is treated (*i.e.*, there is a 90% probability of the consumption value being paired with a zero residue value), the "unrepresentative" consumption value will yield a correct exposure value (*i.e.*, 0 mg/kg/day) 90% of the time and an invalid exposure estimate 10% of the time.

are very frequently quite reasonable (*e.g.*, two or three apples or peaches).

In an attempt to determine if the upper percentiles of consumption as reported by CSFII 1989-92 are unusual or otherwise deviate substantially from that which would be expected, OPP has performed a number of statistical and graphical analyses of the consumption data to investigate these alleged "peculiarities." Specifically, OPP has looked at the consumption data in terms of both the distribution of consumption values for a particular commodity (to determine if the high-end consumption values deviate substantially from that expected based on the pattern of the lower consumption values) and in terms of the absolute values of the high-end consumption values (*i.e.*, are the reported consumption values for a particular commodity unreasonable at percentiles greater than the 95th). In looking at a number of fruits and vegetables that are commonly found to be risk drivers, OPP notes that the high-end percentiles of consumption are: (1) frequently quite reasonable in and of themselves; and (2) frequently follow the pattern of consumption which is displayed by USDA's lower percentiles[11]. There obviously are exceptions, but (as pointed out earlier) these exceptions will be investigated, considered, and evaluated on a case-by-case basis.

Based on evidence seen to date in these analyses, OPP does not believe that it is scientifically appropriate to make a universal declaration that all reported food consumption values that are greater than the 95th percentile are suspect and should therefore be disregarded. In fact, one commenter spoke directly to this issue:

> ...we strongly oppose any unscientific "doctoring" of the CSFII or PDP databases supporting the Agency's Monte Carlo acute dietary risk assessments. As we have shown, the quality control procedures used by USDA in developing these data resources have produced very high quality data, at significant expense to the taxpayer. It would therefore be unconscionable for the Agency to acquiesce to proposals that are intended to make high-end exposure estimates "go away" because they deviate too greatly from what some want to label as "usual" or "representative" patterns of exposure and risk. It is precisely children with high but predictable "normal" exposure who are at risk and whom the FQPA is designed to better protect.

---

[11]Specifically, OPP has investigated the values of the reported high-end consumption of a variety of commonly consumed foods as well as the distribution (or pattern) of consumption of these foods. The former was investigated by looking at the reported values which comprise the top percentiles of consumption and noting if these were potentially highly unusual. The latter was investigated by plotting the logarithms of the reported individual consumption values (on a mg/kg bw/day basis) on a Q-Q plot and assessing whether the reported high-end consumption values differed markedly from the pattern displayed by the lower percentiles (*e.g.*, 80th through 99th).

2. Underlying Precision at High-End Percentiles

Comment. The concerns and issues raised by many commenters were well-represented by one commenter who succinctly summarized the important concepts underlying precision at high-end percentiles and issues and options facing OPP:

> Factors that can affect representativeness include sample sizes and other aspects of survey or study design, reporting/measuring correctness, recording accuracy, and correspondence of the data to real-world conditions. It must be remembered that the value of each of the highest calculated exposure numbers – the ones that together define the 99.9th percentiles value – is simply the mathematical result of multiplying one set of consumption values from the CSFII by one set of residue values chosen by the computer from a set of analytical measurements. Thus, the nature of the high-percentile portion of a Monte Carlo distribution will be affected primarily by *how high* the higher consumption and residue values are, and *how many* of each of these high values there are, compared to the total. A high combined value can result either because the consumption value was high or because the residue value was high. The more high values of either sort included in the inputs, the more high output results there will be, and the higher the input values are, the higher the highest output values will tend to be. If a significant number of the highest output values result from residue and/or consumption values that are unrepresentative, the risk assessment will be distorted...
>
> ...It is quite possible that the high values – the ones that determine where the 99.9th percentile falls – result from the inclusion in the consumption database of: (1) one or more erroneously reported, implausibly high consumption values, and/or (2) one or more consumption values that were correctly reported but represent unusually high consumption. By unusually high consumption, we mean daily consumption that would not plausibly occur at a rate approaching one time out of a thousand but that found its way into the CSFII database simply because the one individual who ate a huge amount of some food was interviewed on the day after his huge "eating event." As we will show by our discussion of the CSFII database, it would not be at all difficult for either of these to occur. In such a case, the model results will indicate that the current use pattern presents an unacceptable risk, when in fact the risk is an implausible artifact of a particular survey subject's extremely unusual and highly infrequent eating behavior or the result of a mistaken estimate...
>
> ...If for whatever reason a person tended to over-report consumption, several amounts in that person's report may be exaggerated. Because each consumption value is used by the DEEM program and many iterations of the matching are performed, each over-reported amount will be matched a

> number of times with any high residue values present in the database. Whenever that occurs, the model will generate overstated exposure values that may cause the modeled risk at a percentile level to be higher than it would be if properly reported inputs had been used. The more this happens, the greater is the potential for exposure distortion and consequent over-regulation.
>
> The best way to minimize the possibility that an extremely rare but correctly reported value will distort the high-percentile predictions of Monte Carlo assessment is to use very large sample populations. The anomalous value is then extremely unlikely to occur in the sample set at a much higher rate than it occurs in the total population. Unfortunately, it is not feasible to increase the size of the CSFII enough to accomplish this. Feasible ways of avoiding this kind of problem are: (1) using a lower percentile value that is not as prone to being heavily influenced by anomalous individual consumption numbers, and/or (2) examining the extremely high individual values to determine their effect on the outcome and whether they should be discounted, and then taking the findings into account in deciding how to regulate. These are the *only* ways of dealing with the implausibly high values that are the result of exaggerations in the responses of samples persons or other errors.

In this same vein, however, another commenter cautioned the Agency on too readily discarding the food and residue databases of high-end values. The commenter stated that the policy statement framed two related and important issues - - what constitutes an "unusually high" food consumption or pesticide residue level? And when is such a level "representative"? The commenter continued:

> This interim policy could, if finalized and aggressively exploited, set the stage for purging food consumption and pesticide residue databases of high-end values. This would be an unacceptable result, which could seriously compromise the public health goals of the FQPA. For obvious reasons, whether such values are "unusually" high or not, they will "drive" exposure and risk estimates because of simple mathematics. By the very nature of a Monte Carlo analysis, combinations of high-end food consumption and high-end residue values will occur at a frequency representing the likely odds of such occurrence in the real world, while such occurrences will account for a very small percentage of simulated eating day episodes, they nonetheless do occur and must therefore be taken into account by the Agency...
>
> ...The Agency's interim policy seems to accept the notion, often raised by the pesticide industry and agricultural sector members of TRAC that high-end Monte Carlo risk estimates are inflated because of outlier values or mistakes in data entry or coding that lead to gross overestimation of food consumption levels, pesticide residues, or both. If this were the case, the Agency would indeed need a procedure to identify such data points so that they could be excluded or adjusted in a scientifically defensible manner.

> [We] have studied the available food intake and residue data extensively... [W]e find no evidence of such an upward bias in the case of foods that account for the bulk of the diet of infants and children and also for the vast majority of the risks associated with dietary exposure to organophosphate (OP) insecticides.

Response. OPP fully recognizes these issues and concerns and is cognizant of the offered cautions. As we have stated in the original *Percentile Policy* document,

> The Office also recognizes that unusually high intakes can potentially "drive" calculated exposure and risk estimates and believes that it may be inappropriate to base risk management decisions on unusual consumption values, particularly if these consumption values dominate high-end exposure estimates. Therefore, OPP is proposing that risk characterizations include a sensitivity analysis that will take advantage of a recent upgrade to the DEEM software program, which is now capable of generating a "Critical Exposure Contribution" (CEC) analysis when run in the acute Monte Carlo mode. The CEC provides insights into the sources contributing to the exposure estimated for the most highly exposed people in the exposure distribution... The display includes key demographic information (gender, age, body weight), the food(s) consumed, amount consumed, the residue value, the total daily exposure estimate, and the exposure estimate by food. Thus, the CEC provides the Agency with comprehensive information on foods (and food-forms) that account for the largest portion of the person's estimated exposure. If OPP finds that the high-end exposures are principally driven by suspect high-end consumption values, the Agency's risk mitigation decisions can appropriately consider and weigh these factors.

3. Specific Thoughts and Ideas

Comment. Numerous comments reflected specific thoughts, ideas, and recommendations relating to the issues surrounding the USDA survey and reported high-end consumption values. These dealt specifically with the outlier question with respect to both the CSFII consumption data and the residue data and offered suggestions on how these potential outliers may have originated and should be verified and ultimately handled in the regulatory decision. One commenter provided additional information about the techniques used by USDA in their 1994-96 survey which OPP intends to adopt in the near future (second quarter of calender year 2000) and general concerns about accuracy and bias in the survey and limited scrutiny that high reported consumption values received. For the 1994-96 survey, two days of consumption were targeted, each of which was the subject of a separate recall interview. The interviewer obtained answers to an 18 page questionnaire, inquiring three times about each food eaten. The commenter expressed concern with respect to the potential for overstated

quantities of foods of particular types and the reliance on portion size estimation guides such as measuring cups, spoons, and rulers to help persons estimate how much they ate. In addition, concern was expressed about the short nature of the interview (30-33 minutes) and the fact that many foods had to be dealt with. For each of the named foods, a series of questions had to be answered. In addition, each interview included a long list of additional questions about general dietary habits and lifestyle. Given the short time of the interview and the many questions asked, the commenter pointed out that there was little time to closely scrutinize the reported individual consumption amounts. The commenter acknowledged the QA/QC procedures used by USDA: the commenter recounted, for example, that if a sampled person reported unusually large consumption of an item, the interviewer was asked to confirm this with the interviewee and note this confirmation on the survey form. As an added check the coding software automatically questioned very high values, and coders or reviewers could question outlier entries and remedy data entry errors. The commenter acknowledged that in some cases, interviewers re-contacted sampled persons at the request of the reviewers. In a small number of cases, USDA reviewers excluded high values if they concluded that the evidence showed the reported value was not possible or if there had been no confirmation by the sampled person. The commenter pointed out, however, that high values were excluded only if the evidence showed that they were impossible or were not confirmed by the sampled person - - values were not excluded simply because they were very high or implausible.

One commenter stressed that the specific language of the FQPA demonstrated that Congress intended the Agency to use scientific judgement when selecting data on which to base regulatory decisions and that FQPA required that the validity, completeness, and reliability of data be assessed before being used in a regulatory area. The commenter stated that OPP thus must consider the validity, completeness, and reliability of exposure estimates along a population exposure distribution before selecting a point on that distribution as a basis for regulation and, specifically, that OPP must consider the validity, completeness, and reliability of the estimate at the 99.9^{th} percentile before selecting it as a regulatory endpoint. The commenter indicated that the data selected for analysis should be representative of consumer behavior and should not include implausible reported consumption values. He provided several examples, including an instance of consumption of over 3 kg beef in a single eating occasion by a 13-18 year old male in the 1994-96 CSFII and consumption of 300 g of grapes by an infant in the 1989-91 CSFII database. These data, the commenter stated, are not representative of the populations and using such data in a risk assessment can lead to inaccurate estimates of risk, especially when estimating upper percentile exposure estimates for a population.

Response. OPP generally agree with both comments. OPP's risk

assessments begin with the presumption that USDA CSFII consumption data are reliable based on the survey design and the QA/QC procedures followed. Rather than having the risk assessor purge the CSFII data sets of any reported consumption values that are regarded as questionable or suspect *before* any risk analysis is done, OPP believes that it is more appropriate to first perform the analysis with the entire data set (*i.e.*, fully intact) to determine if the putative "outliers" are indeed responsible for the risks at the upper percentiles which exceed our level of concern. OPP believes that in most cases they are not. It is far easier for resource reasons for OPP to investigate outliers using the software's CEC approach (discussed earlier) than it is to purge data beforehand of suspect points. Moreover, in the interest of transparency and consistency, OPP believes it is appropriate to perform analyses on the entire CSFII dataset rather than alter the dataset at the outset of each analysis. As indicated earlier, if high-end values are responsible for risk estimates at the upper percentiles which exceed our level of concern, this will be properly characterized in the risk assessment and this information conveyed to the risk manager for an appropriate decision. The examples of high consumption offered by the commenter would be scrutinized if they were responsible for driving any given risk assessment performed by OPP.

4. Mis-estimation of Extreme Percentiles Likely with Finite Samples

Comment. One commenter indicated his belief that an extreme percentile such as the 99th is most often driven by "outliers" for a single input. For example, if one examines a model's output and focuses on the extreme tails, one finds, for example, that the tails consist of people who are "typical" except for one (and generally only one) very atypical dietary input (*e.g.*, they ate 5 lbs of apples that day). He continued, stating that even if the USDA CSFII data were to be screened to eliminate all incorrect extreme values, there would still be a number of correct values which are inappropriately extreme (i.e., we are quite likely to have at least one outlier that is "real", but still "wrong" in the sense that it represents a much too extreme percentile of the actual quantity of interest). The commenter provided an example illustrating this point: if we were to have a sample of 1000 observations, the largest observation would be expected to be between the 99.93 and 99.997 percentile (with 95% confidence)[12]. Thus, the largest value is about the 99.9th percentile most of the time, but there is approximately one chance in forty that the largest value is actually as high as the 99.997th percentile (which is considerably stricter than the nominal 99.9th percentile the Agency apparently desires). If this same pattern emerges over 20 or so inputs (each estimated from a sample of 1000), then the probability of one or more inputs representing in actuality a level greater than the 99.99th percentile is 40%. The commenter stated

[12] These values were likely calculated by the commenter using non-parametric order statistics, and have not been re-calculated or verified by OPP.

that this mis-estimation of extreme percentiles is likely with finite samples and recommended that much larger sample sizes or, in the absence of very large data sets a clear *a priori* idea of what is considered to be "large" consumption values. He suggested, for example, that if the 99.9th percentile consumption value were to be more than (for example) five times the mean, this reported consumption could be considered an outlier.

Response. As indicated before, OPP would specifically examine the tails by means of the CEC output. If these tails consistently consisted of individuals who are "typical" except for one very atypical dietary input (e.g, one person eating 5 lbs of apples, another eating 3 lbs of peaches, a third eating 6 lbs of green beans, etc.), this would be fully characterized in the risk assessment and considered by the risk manager. To date, OPP generally has not found that the persons in the tails of the distribution are there due to extreme consumption values.

With respect to the commenter's concern that an analysis with 20 sets of inputs (e.g., foods) of 1000 observations each would lead to a 40% probability that at least one of those observations is at the 99.997th percentile and represents too extreme a percentile of the actual distribution to consider, OPP has found that frequently only one or two (and perhaps occasionally three) foods serve as "drivers" of the risk assessment. With only two or three foods that are generally responsible for the bulk of the risk, the probability of having at least one of these consumption values at the 99.997th percentile is reduced from 40% to between 5% and 7%[13]. The commenter is correct in that the greater the number of different foods which enter an assessment, the greater the chance of having extreme (but still real) outliers being reported. However, this is tempered by the fact that it is usually one or two foods which serve as risk drivers.

5. Addressing Outliers in CSFII

Comment. One commenter stated that outliers should be addressed in the food consumption data sets by looking for: (a) extreme high-end single consumption values, and (b) looking for extremely high daily caloric intakes. Still another commenter recommended that bounding estimates be applied in judging the appropriateness of high consumption levels for individual risk considerations, suggesting that a one in 500 or one in 1000 eating occasion event (or perhaps two or three standard deviations from the population mean instead) be considered a rare, atypical event which deserves separate evaluation. Alternatively, the commenter suggested that more parameters (*e.g.*, subgroups) be established to identify cluster characteristics to focus risk reduction efforts on targeted groups, rather than disrupting a less risky general population. Presumably, the commenter

[13] Calculated as $1 - (39/40)^{20} = 40\%$ vs. between $1-(39/40)^2 = 5\%$ and $1-(39/40)^3 = 7\%$

believes that it would be appropriate to identify the characteristics or factors for small but very specific groups that may consistently be present in the tails of the exposure distribution.

Another commenter, taking the opposite approach, indicated that outliers should be identified and those consumption databases that are too limited for robust description of population extremes should be identified and flagged as well. The commenter suggested that if databases are sufficiently rich, modeling of distributions can identify outliers with consideration of those data points that lie outside of prescribed confidence limits for the modeled distribution. Consideration of the improvement in model fit when suspected outliers are removed may additionally aid in identifying outliers. The commenter suggested that the significance of removing the outlier on the resulting outcome should be tested and that statistical tests can be devised to evaluate the significance of data censoring on exposure predictions. Sensitivity analysis may be used as well to determine those input distributions where the presence of outliers may most significantly affect the exposure assessment. The commenter stated, however, that while evaluation of outliers and use of the CEC are valid and useful approaches to better understanding sources of uncertainty at the extremes of exposure distributions, these techniques do not address or correct the fundamental flaws in logic associated with use of an extreme and uncertain endpoint as a regulatory threshold.

One commenter concurred with the Agency's use of USDA's CSFII survey, stating that it is the highest quality food consumption dataset available and that the draft science policy paper presented a clear description of the extensive quality control procedures that the USDA has developed over many years and now relies upon to assure that the consumption values in the CSFII are an accurate representation of the true distribution of actual eating patterns and habits. He concurred with OPP's statement that "the USDA CSFII database has been properly evaluated and contains accurate and reliable consumption values that, by FQPA standards, are acceptable for use in OPP's assessment of human dietary exposure to pesticide residues." The commenter stated that he was

> aware that other groups are warning the Agency that implausible outlier values in the CSFII render Monte Carlo results "very unstable" at the high end of the exposure and risk curves. To determine whether there is any validity to this claim, we assessed the distribution of the actual reported food consumption levels in the 1994-1996 CSFII for one to five year olds.

The commenter proceeded to perform an extensive analysis of the 1994-1996 CSFII data of the type recommended by the other commenters. Specifically, consumption of fresh apples, apple juice, fresh peaches, and fresh pears by children aged 1 to 5 was investigated. A total of 5,372 valid eating days were available for

analysis – including 889 fresh apple eating day episodes, 1,130 apple juice eating day episodes, 72 fresh peach eating episodes, and 97 fresh pear eating episodes. For each of the 5,372 CSFII eating days, the commenter calculated the grams of the four key children's foods consumed on a per kilogram of body weight basis and then ranked the results and calculated a variety of descriptive measures to characterize more fully the distribution of values. Specifically, the commenter calculated for each food the mean level of grams of food consumed per kilogram of body weight as well and the 95th, 99th, 99.9th percentiles of consumption as well as the highest reported consumption. These are reproduced in Table 1 for each of the four food forms investigated:

Table 1. Distribution of 1994-1996 CSFII Food Consumption Levels for Four Key Foods, Measured in Grams of Food per Kilogram of Bodyweight for Children Ages 1 to 5

	Apples	Apple Juice	Peaches	Pears
Maximum Value	26.7	136.7	11.5	29.2
99.9 Percentile	22.8	121.5	11.5	29.2
99 Percentile	18.0	78.1	11.1	18.3
95 Percentile	13.8	52.9	9.4	14.4
Mean	6.8	21.3	5.6	7.5
Minimum	0.2	0.8	1.1	0.6
Total Eating Days	889	1,130	72	97

Source: Compiled by Benbrook Consulting Services, based on 1994-1996 CSFII Consumption Data.

The commenter concluded that there are clearly no "odd-ball" outlier values in the CSFII food consumption survey data for these four major risk-driver foods consumed heavily by infants and children. Of the 889 records in which apple consumption was reported, only two entailed consumption of more that 400 grams of apple in a day. A 15.88 kg four year old was responsible for the highest level of consumption on a per kilogram of body weight basis (26.7 g /kg bw), with another four year old child weighing 18.14 kg responsible for consuming a total of 414

grams of apple in a day. This level represents consumption of three medium sized apples. While this may be a high level, the commenter stated that children can eat certain favorite foods at various stages of growing up in comparable or even greater quantities. The commenter also pointed out that there are only modest differences (two to six fold) between the 99.9th percentile of consumption and mean consumption. For apples, there is only a 3.37 fold difference between the 99.9th percentile of consumption (22.8 g/kg bw) and the mean level of consumption (*i.e.*, 6.8 g/kg bw).

The commenter concluded that he was

> confident that when the consumption values for other commonly consumed commodities are subjected to the same sort of analysis, the results will be comparable. The Agency and USDA can readily confirm this prediction by issuing a ranking and summary of reported food consumption episodes for the 20 or so major foods making up most of the diet of infants and children. This could be done each time a new set of data is released through the CSFII.

and that

> We believe the USDA's statistical procedures are catching and truncating any implausible values, and that the distribution of consumption levels per kilogram of body weight will be tight in cases with few reporting eating episodes. But to allay fears that a truly unusual value in a rarely consumed food might skew upward an estimate of risk, even for a very few individual eating day risk estimates, we concur that the agency should put in place an empirical filter to trigger an assessment of such unusual cases. We recommend further assessment of high-end consumption values if two conditions are met. First, one of these two triggers should apply –
>
> · the 99th level of consumption exceeds the mean by six-fold or more, or
> · the 95th level of consumption exceeds the mean by four-fold or more.
>
> Then, the EPA should require an affirmative judgement from an expert panel of dieticians and food consumption specialists that high-end consumption levels meeting one or both of the above triggers are, in fact, implausible. One obvious set of cases where such levels would be plausible, and should not be altered, is a food typically served and consumed as a garnish in relatively low quantities – leading to a relatively large number of low-consumption episodes (and hence a low mean) -- which some children eat as a main course, perhaps in an ethnic dish or seasonal favorite of a family.

Overall, the commenter indicated that he supported assuring that the underlying food consumption and residue databases are themselves sound prior to their

incorporation in a Monte Carlo.

Another commenter made a similar statement that “since the passage of FQPA, the EPA has been distracted by accusations that outliers in the food consumption data might have compromised the Agency’s probabilistic risk models.” He performed an exercise similar to that described above, albeit with a slightly broader range of foods. The commenter stated that a recent analysis of dietary risk from OP pesticides to children showed that the biggest sources were apples, peaches, fresh green beans, apple sauce, apple juice, grapes, and pears (EWG, 1999) , and that, together, these seven foods are responsible for approximately 87% of the children who receive a daily dose of OP’s above the RfD. If consumption outliers render EPA’s models inaccurate, the commenter continued, then one should see extremely high levels of consumption of these foods. The commenter included the following table (Table 2) and concluded that the maximum reported consumption values are quite reasonable.

Table 2. Maximum Consumption Does Not Differ Greatly from Average Consumption

	Average Consumption	Maximum Consumption
Apple	3/4 Apple	3 Apples
Peach	1 Peach	2 Peaches
Fresh Green Beans	1.5 ounces	5 ounces
Apple Sauce	½ cup	2 cups
Apple Juice	1.2 cups	2.5 quarts
Grapes	3 ounces	1.1 pounds
Pear	2/3 Pear	2 pears

Source: CSFII (1994-96) and Gebhardt, 1991

The commenter stated that most of the food consumption levels differed by approximately 4-fold or less and that the only foods that showed greater than a 4-fold variation were grapes and apple juice although "even these day long consumption values serve as realistic maximum population values given the well-documented eating habits of small children." He concluded that "the EPA and USDA need to move beyond the issues of outliers and other statistical smokescreens."

Response. As detailed in the original *Percentile Policy* document, OPP believes that the USDA CSFII consumption database is a valid survey of U.S. dietary consumption practices and can be used for purposes of dietary risk assessment for pesticides. OPP believes that the analysis conducted by the commenters concerning the alleged presence of outliers is sound and further supports the approach of using the CSFII data "as is" with the added caveat of routinely performing CEC and sensitivity analyses to better characterize the risk and exposure estimates. As discussed earlier, OPP will consider a number of factors in determining if consumption outliers represent implausible values and/or have an undue effect on exposure assessment and these factors will be considered by the risk manager in a risk management decision. As previously stated, OPP will not remove any perceived outliers *a priori*. Instead, decisions will be made concerning any deviation from the 99.9th percentile regulatory threshold on a case-by-case basis which considers all available information, including the nature and extent of any perceived outliers. The extensive specific considerations recommended in the above comments will be appropriately evaluated.

6. Consider Subtle Biases in Model Inputs

Comment. One commenter indicated that models which focus on extreme percentiles need to take subtle dependencies and biases in model inputs into account and states that any complex Monte Carlo model will have inputs that are correlated. He provided an example (*e.g.,* if a person eats many apples, they will likely eat fewer pears) and acknowledged that associations such as this are appropriately handled by the CSFII data which records actual dietary intakes. The commenter listed some examples of correlated inputs that may be less obvious and should be incorporated into the assessment: individuals who consume large quantities of fresh fruits and vegetables may be particularly health conscious and these individuals might tend to preferentially purchase organic produce and thus be at much lower risk than their dietary intake of produce would imply; individuals may not accurately report dietary intakes and there may be a tendency to overstate consumption of fresh fruits and vegetables. The commenter recommended that studies to answer such questions be initiated. For example, for the CSFII data, one could determine if food sales data are consistent with reported consumption or one could perform a general survey to find out how much fresh produce people eat and whether they might preferentially purchase organic produce.

Response. OPP believes that it is important that all food (organic and conventional) be safe to eat such that there is no reason that high-end consumers of fruits and vegetables should feel that it is necessary to purchase organic produce to be sufficiently protected. To introduce an added variable (organic or conventional for each food form eaten) to the survey would substantially complicate what is already a very detailed and time-consuming interview. Nevertheless, if these data were provided, their impact could potentially be evaluated and it could be determined whether any significant changes in our exposure estimates at the 99.9^{th} percentile occur.

With respect to the possibility that interviewed individual over-report their consumption of fresh fruits and vegetables, OPP believes that USDA has taken adequate steps to minimize this bias. OPP does not believe that any putative over-reporting is of such significance to invalidate the survey or to require that wholesale adjustments to the data be made. In any case, OPP believes that any over- reporting is more likely to affect the *mean* consumption values, and have a lesser effect (if any) on the extreme tails of the exposure distribution at which regulation occurs since the putative over-reporting more likely occurs among those who “under-consume” the commodities of interest.

7. Consider Nature of Cumulative Distribution

Comment. One commenter indicated that the nature of the cumulative exposure distribution curve around the selected decision point needs to be considered when evaluating whether the uncertainty in the exposure estimate is too great to be meaningful for regulatory decision-making. Typically, the commenter stated, cumulative exposure response curves have a "hockey-stick" shape where exposures are close to zero over a considerable range of potentially exposed population and then skyrocket upward at the extremes of the distribution. The commenter stated that as the tails of the distribution is approached, acceleration[14] upward in the exposure response is observed and that acceleration along the exposure response curve is especially rapid as the extreme of the output distribution is approached. An example is provided: the relative acceleration for the interval from the 99.75^{th} to 99.9^{th} percentile is 1000-fold of that occurring between the 95^{th} and 99^{th} percentile. This rapid increase in estimated exposure with slight increases in the proportion of the population considered adds substantially, the commenter stated, to uncertainties in exposure assessment at the extreme of the distribution and these types of changes mostly reflect the extremes in the input data for consumption where unusual and uncertain patterns of food consumption are represented in the conditional probability distribution for exposure. The commenter suggested that statistical evaluation of slope over a range of exposures can possibly contribute to understanding of uncertainties at the extremes of distributions and that rapidly changing slopes about the decision-point of interest are indicative of high uncertainty in the exposure estimate at that particular point.

Response. The commenter has raised a number of issues in his remarks. OPP believes that the rapid increase in estimated exposure with slight increase in the proportion of the population (what the commenter refers to as "acceleration") is a natural outgrowth of the log-normal nature of the consumption and residue distribution curves which, together, define the exposure distribution curve. As acknowledged by the commenter

> [W]ith respect to residue distributions, it is well recognized that organic residues in the environment typically distribute in a manner best described as log-normal...Residue data, therefore, may be best represented by a lognormal distribution (where upper and lower bounds are truncated by a lower bound of zero and an upperbound fixed at the residue tolerance) ...With respect to consumption patterns, the expectation is that these data, too, will most typically be left skewed. Unusual or extreme consumption patterns cause left skewing of otherwise normal distributions for commonly consumed items. For infrequently or less consumed items where no consumption will be shown in most diets, the effect of extremes in

[14] The commenter defines acceleration as the second derivative of the cumulative exposure distribution curve

> consumption will lead to even greater skewedness. In the common event where data are too limited for the distribution of data to be unambiguously modeled, the assumption of lognormality substantially lessens error associated with the uncertainty regarding distribution selection than does the assumption of a normal distribution.

It is curious, then, that the commenter believes that the rapid rise in estimated exposure is necessarily due to anything other than the fact that when a log-normal consumption distribution (with occasional but still very real extremes) is multiplied by a log-normal residue distribution (again, with occasional but still very real extremes), the result will be anything but a right-skewed distribution with a more extreme right-tail than either of the two input distributions. It is this "more extreme right tail" that is described as undue "acceleration" by the commenter (which manifests itself as a sudden and rapid increase in estimated exposures when plotted on a cumulative distribution curve). It also appears that the commenter is ascribing this "more extreme right tail" as symptomatic of "uncertainties at the extremes of distributions" and "indicative of high uncertainty in the exposure estimate at that particular point." It seems the commenter may be confusing "uncertainty" with "variability." The rapid rise in estimated exposure at the tails of the distribution is merely reflective of expected *variability* within the population. To quantitatively assess the degree of *uncertainty* in the tails of the distribution, a more complex analysis (2-Dimensional Monte Carlo analysis) would need to be performed. OPP, in deciding the appropriate point on the exposure curve at which to regulate, must adequately consider the full range (or variability) of exposures to the population. The "acceleration" described by the commenter only means that there is a small group of persons at the tails of the exposure distribution (and this group rapidly grows ever-smaller as the predicted exposures become more extreme). This rapidly diminishing group size was fully taken into consideration when proposing the 99.9^{th} percentile as a threshold of concern (in fact, it is one of the reasons the 99.9^{th} percentile was proposed).

OPP agrees with the comment that, as a general rule, uncertainty does increase as the estimates of exposure become more extreme. OPP does not agree (as was discussed above) that the phenomenon of "acceleration" is a direct measure or symptom of this uncertainty. OPP also disagrees that a "statistical evaluation of slope over a range of exposures can possibly contribute to understanding of uncertainties at the extremes of distributions" and that "rapidly changing slopes about the decision-point of interest are indicative of high uncertainty in the exposure estimate at that particular point."

Finally, OPP does agree that the "slope" of the exposure distribution curve should also be considered in a risk management decision, but for different reasons. It would make little sense for a risk manager to consider two scenarios equivalent if in one scenario exposure was unacceptable at the 99.9^{th} percentile, but

acceptable at the 99.5th percentile (a steep slope) while the other was unacceptable at both the 99.9th percentile and the 95th percentile (a shallow slope). A shallow curve indicates many more people are potentially exposed at levels greater than the RfD (or PAD) and thus there is reason for greater concern than if the curve is steep.

8. Consider Statistical Weights

Comment. Another commenter provided input on the statistical weights used in the survey design, noting that the impact of statistical weights should also be considered when assessing the reliability of exposure estimates at the 99.9th percentile. The commenter stated that the upper tails of a population consumption distribution can be heavily influenced by the statistical weight that is assigned to the high-end consumer. If the high-end consumer happens to be from a statistically under-represented subgroup, then the upper percentile consumption estimates for the population can be misleading and not representative of the population. Thus, the commenter stated that when evaluating the reliability of the upper percentile estimates, the statistical weight assigned to high-end consumption values must be taken into account.

Response. OPP agrees. However, we note that the statistical weights used in the survey design are integral to the survey and are required for the survey to be considered representative of the population. Although statistical weights and their effect on high-end exposure estimates will be considered, OPP will be very cautious about discarding this important information.

9. Outliers in Residue Data

Comment. A number of individuals made specific comments on the Agency position on "outliers" from residue (as opposed to consumption) databases. One commenter believed that the Agency's stated position on "outliers" for residue values from field trials is appropriate. It would be proper, the commenter stated, to reject residue values on the basis of an experimental blunder, such as the use of the wrong formulation, an erroneous application rate, or a harvest time outside of the designated pre-harvest interval (PHI). "However," the commenter stated,

> it is inappropriate to reject residue values merely because they lead to risk estimates that inconvenience a chemical company. Considering that field trials are conducted at a limited number of sites, under climate conditions in effect at the time of the trials, residue databases on maximum label conditions should otherwise be assumed to be realistic. This should especially be the case for uses where the tolerance has been set on the basis of few samples at each field trial.

Response. OPP agrees with the commenter. With respect to outliers from field trials, OPP will generally only discard residue data if they are the result of an experimental blunder or are clearly implausible. If the outlier was believed to be valid and used earlier in establishing a tolerance, it would be necessary for the pesticide registrant to first demonstrate that the tolerance is invalid and should be lowered.

10. PDP as the Primary Data Set

Comment. Another commenter stated that with respect to use of USDA's Pesticide Data Program (PDP) that PDP data provide the highest quality, most up-to-date residue data covering the foods most heavily consumed by infants and children. The commenter agreed with the Agency – for foods tested by PDP (even if sampled in just one year), PDP data should be used as the primary dataset when carrying out Monte Carlo assessments. As stated by the commenter:

> The advantages of PDP - - reflecting food as eaten, after storage, washing, and preparation - - outweigh the disadvantages of smaller sample sizes than what might be accessible by combining several years of FDA surveillance monitoring data, or other data sources of more debatable relevance and quality...The larger the PDP dataset for a given food, the greater the confidence that can be placed in the data. For this reason, if the condition stated below is met, we support the merging of up to three years of PDP data for a single food. The condition is that data should not be merged if there were substantial changes in pesticide use patterns – acres treated, rates of application, or timing of application between years. We suggest a "substantial equivalency test" -- accept no more than a 25 percent change from one year to the next in any of these three indicators of pesticide use patterns. USDA pesticide use data, augmented by reports from extension specialists in the field for the years when USDA does not collect fruit or vegetable data, can be used to apply this test in years when USDA does not collect fruit or vegetable use data.

With respect to outliers and PDP data, the commenter continued, some allege that a few grossly exaggerated pesticide residue values are driving high-end risk outcomes in Monte Carlo analyses. The commenter analyzed 57 food-pesticide combinations in the 1997 PDP sampling. For each, descriptive statistics were computed and analyzed as in the case of the CSFII consumption data. The commenter concluded that:

> Given the design of the program, we do not believe there are any circumstances that would lead to a composite residue level that does not, in fact represent actual levels of residues in the food supply. While some very high residues may result from illegal pesticide use, the FQPA makes no

distinction between residues from legal and illegal uses.

> Based on the clear mandate of the FQPA, we urge EPA to include all such exposures in its cumulative risk assessments. Such cases will contribute relatively infrequently to exposures among children exceeding their PAD or RfD on a given day, but still may warrant attention as the agency sorts its way through risk mitigation options for a given set of active ingredients and/or foods contributing excessively to acceptable exposures and risks.

Response. OPP agrees. With respect to invalid residue measurements from PDP, these would be expected to be very rare given the QA/QC practices of the PDP program. Without overwhelming evidence of sampling or analytical error or clear implausibility, OPP will not discard high-end residue results from the PDP program. Furthermore, OPP monitors changes in use practices, percent of crop treated, and other factors which may be expected to have a substantial effect of residues using databases from USDA's National Agricultural Statistics Service and proprietary subscription services and discussions with agricultural extension personal. If changes in use/usage data are seen or suspected, OPP will incorporate this information into its risk assessment.

E. Clarification Of Issues and Ideas

✔ Population vs. Individual Risk
✔ Risk Equation
✔ Log-Normal Distribution of Exposures
✔ Introduction of PAD
✔ Appropriate FQPA Groups
✔ Interpretation of Percentiles
✔ 99.9^{th} Percentile: To What does it refer?

Overview. *A number of the comments received requested clarifications of some of the issues which were addressed in the* Percentile Policy *document. Others had apparently misinterpreted some issues or suggested that OPP itself had misinterpreted these issues. In any case, OPP believes that the clarifications listed below will assist readers in interpreting and judging our policy guidance.*

1. Population vs. Individual Risk

Comment. One commenter provided detailed information about population risk vs. individual risk, warning OPP that it is important when examining questions about public health to avoid confusing the mathematics that support the description of population risk with application of the current policy to

individual risk. The commenter stated that individual risk is a mathematically different question than population risk and should therefore be handled in a different manner. Although population risk is not specifically defined by the commenter and how OPP was confusing the mathematics between the two risk measures was not detailed, the commenter stated that direct application of population estimates to individuals is inappropriate. He stated:

> As high levels of computer capacity have become more and more available to researchers in the last decade, scientists have been shifting more from using population summary descriptive values to utilizing the option of including all individual data to represent a population. Most of the mathematical techniques were developed to support summary values, and have just been expanded over recent years into more probabilistic applications. As Monte Carlo techniques are utilized in new applications more and more, some technical uses will prove to be more useful and scientifically acceptable than others.

The commenter encouraged further dialogue and public discussion with academia and industry on the issue of population vs. individual risk, stating that it was a large topic which deserved thorough examination, review of public literature, and discussions that are beyond the scope of a single public response to a proposed OPP policy. The commenter stated that if the intent of OPP is to protect more individuals than under the previous policy, other options could be employed to achieve the same goal and that it is unnecessary to use a scientifically unreliable point for application of the statutory safety standard.

Response. The commenter has raised a valuable point concerning the philosophical differences between regulating population risk *vs*. regulating individual risk. OPP considers both population risk and individual risk; both are critical to a sound, effective, and fair regulatory policy. As stated in the Agency's Exposure Assessment Guidelines (U.S. EPA, 1992)

> In preparing exposure information for use in a risk assessment, the use of several descriptors of both individual risk and population risk often provides more useful information to the risk manager than a single descriptor or risk value. Developing several descriptors may require the exposure assessor to analyze and evaluate the exposure and dose information in several different ways. ... The questions that need to be addressed as a result of the purpose of the assessment determine the type of risk descriptors used in the assessment.

Individual risk is defined here in the context of risk borne by individual persons within a population, whereas population risk refers to the extent of harm for the population (or population segment) being addressed. Individual risks are frequently calculated for some or all of the persons in the population being studied

and are then put into the context by indicating where these risks fall in the distribution of risks for the entire population. Population risks, on the other hand, may deal for example with how many individuals might be probabilistically estimated to be above a certain risk level (*e.g.,* the portion of the population which exceeds the RfD or an effect-based level such as the LOAEL). In response to the commenter's concern, OPP maintains that there is not an intrinsic conflict between the mathematics of individual risk and population risk and that, by using probabilistic methods, consideration of one implicitly leads to consideration of the other. Despite the fact that OPP develops a distribution of risks over an entire population or subpopulation, these are still distributions of *individual* risks. And despite the fact that these are individual risks, these are *distributions* of individual risk over the *entire population* (or subpopulation). Hence, individual risk and population risk are not two conflicting concepts using separate mathematical techniques, but rather two synergistic approaches which should be considered jointly in arriving at reasonable regulatory alternatives.

2. Incorporating Summary Descriptive Values

Comment. One commenter seemed to suggest that OPP uses population summary descriptive values for input parameters in estimating the distribution of individual risks using probabilistic techniques. That is, OPP's probabilistic estimates incorporate summary descriptive values (such as averages) into its estimates of exposure distribution.

Response. OPP does not agree. This would be inappropriate because a principle tenet in probabilistic exposure assessment that exposure occurs to an individual and the integrity of the data concerning this exposed individual should be consistently maintained throughout the assessment. It is appropriate in a probabilistic assessment to consider the *full distribution over many individuals of the exposure of interest* (and then only if the risk assessor could ensure that the associated correlations and linkages are adequately accounted for). What is important in a probabilistic assessment of individual risks is not the average (or other summary statistic measure) exposure, but rather the exposure experienced by *each specific individual*. The fact that a "composite" of individuals may "average" a given exposure is not useful and cannot be appropriately incorporated into the probabilistic risk assessment.

3. Question on Risk Assessment Equation

Comment. One commenter suggested that OPP's explanation of risk assessment and particularly the risk equation (in section I.C.2) showed "clear confusion in the mind of the writer about risk and toxicity." The explanation included the following formula:

$$RISK = f(\text{toxicity, exposure})$$

The commenter inferred that the formula meant that one multiplies the toxicity value for the pesticide by the amount of pesticide to which an individual is exposed. The writer correctly pointed out that since the paper uses the RfD as the measure of "toxicity," it would be incorrect to multiply the RfD by the exposure to obtain an estimate of risk since the RfD is really a measure of non-toxicity.

Response. The *Percentile Policy* document attempted to convey the concept described by the commenter. The reason the above notation was specifically chosen was to avoid that confusion and make the risk equation as generic and widely applicable as possible. The above notation (called *function notation*) is not meant to imply that the two quantities represented by "toxicity" and "exposure" are necessarily multiplied together, just that toxicity and exposure are two quantities which determine risk. When toxicity is measured in terms of cancer-causing potential (*e.g.*, in terms of a slope factor as a Q*), then toxicity and exposure are multiplied together. When toxicity is measured in terms of an RfD, then the reciprocal of the RfD (which is representative of "toxicity") is multiplied by the exposure to obtain the risk. To avoid confusion and recognizing that function notation is not frequently encountered or always readily understood, the revised policy document includes a footnote which clarifies this concept and states that risk is determined by "combining" (rather than "multiplying") "a value representative of the toxicity" for the pesticide by the amount of pesticide to which an individual is exposed.

4. Concerns Over Bell-Shaped Curve Used as Example

Comment. A number of commenters were dismayed that a sidebar and an example in the appendix showed a bell-shaped curve which purportedly represented dietary exposure to pesticides. One stated that the decision to use the 99.9^{th} percentile is "based on the assumption that human exposures and human sensitivities are normally distributed." The commenters correctly pointed out that distribution of exposures to pesticides in food is a highly skewed distribution and should be portrayed as curve with a large hump to the left side of the graphic. One of the commenters specifically expressed "extreme concern and disappointment" with the information presented under Section 3 (Databases used in Probabilistic Dietary Exposure Estimates) and with the referenced Appendix entitled Primer on Interpretation of Exposure Distribution Curves. The commenter stated:

> Any implications that exposure distribution curves - - and thus the source inputs of residue concentrations and consumption - - are normally distributed is extremely misleading and misinformed. If the Agency's understanding of

> acute dietary risk is indeed based on the assumptions of normal statistics, then the rationale underlying their interpretations of extreme outliers can be expected to be largely incorrect...Any notion that exposure distributions should be interpreted as normal distributions indicates a fundamental flaw in the Agency's appreciation and interpretation of high-end exposure estimates.

Response. OPP risk assessors are fully aware that exposure distributions of residue distributions in the environment are typically lognormal and in many cases is willing to make this assumption (see, for example, U.S. EPA, 1998a and U.S. EPA, 1998b). OPP's experience with consumption data, too, also indicates that consumption distributions are likely to be right skewed distributions (tailed to the right) and can, in many cases, be adequately modeled as a lognormal distribution.

OPP's inclusion of the primer on statistics grew out of early experiences trying to explain to the public its approach to dietary risk assessment. Use of the bell-curve example in the primer on interpretation of exposure distribution curves permits ready recalculation of tabled exposure values from first principles of "classic statistics." If this example were to have used a more realistic log-normal distribution of exposures, then recalculation of the appropriate values by the public would have been considerably more complex and would have likely led to greater confusion among the general public by even those that are familiar with statistical principles. The primer to which the commenter objects was included partly as a result of questions received during and after a series of public meetings where attendees had numerous questions on how the exposure at the 99.9 percentile was calculated and how it was subsequently related to a determination of acceptable or unacceptable risk. Although it was apparent that many people had sufficient background knowledge of statistics, OPP realized in subsequently reviewing the original material presented, that insufficient information was included for even astute readers to place the 99.9 percentile exposure estimate in the appropriate risk assessment context. Thus, it was decided in this paper to include a primer on this topic.

OPP believes that it still makes sense for this primer to rely on basic (normal distribution) statistics so as to be accessible to the widest audience. Although using a lognormal distribution (as suggested by the commenter) and all its attendant mathematic complexities would have been more realistic, OPP believes that this would have severely hampered understanding and would have contributed minimally to comprehension of the necessary principles at the basic level. Nevertheless, OPP has added an explanatory note to the sidebar that the classic bell curve does not often represent pesticide exposures which a generally highly skewed. OPP also added an additional note to the primer which indicates

that environmental distributions are typically lognormal (right-skewed with a long right tail) and that the analysis conducted in the primer could be extended to such cases, albeit with considerably more complexity.

5. Frustration with the Term "Population Adjusted Dose"

Comment. Several commenters were frustrated with the introduction of a new term, the PAD (population-adjusted dose) which OPP is now using to describe the value produced when the RfD is divided by an FQPA factor. Commenters questioned why OPP needed to introduce still another acronym which is simply a modified Reference Dose (RfD).

Response. OPP understands that this additional term and acronym may cause some confusion, at least at the beginning, and introduced it only with reluctance. The major reason for another term is to clearly indicate whether an FQPA factor based on increased sensitivity was included in the OPP-determined acceptable dose or not (*i.e.*, if the RfD had been modified or not). OPP is concerned that without a clear distinction, it would be difficult to readily differentiate between unmodified RfDs and RfDs modified by an additional factor addressing increased sensitivity.

6. Use of the Subpopulation "Women of Childbearing Age"

Comment. Another commenter stated his belief that the subpopulation of "women of childbearing age" should not be listed as one of the groups affecting the PAD value, stating that FQPA specifically mentions children, not women of childbearing age. The commenter did not believe that the "women of childbearing age" subgroup should be used as an example in this paper unless there is evidence indicating that infants and children are directly affected by exposure to this subgroup.

Response. When the toxic effect of concern relates to fetal development, OPP applies the FQPA factor to the "women of childbearing age" population subgroup since this subgroup represents the conduit for fetal exposures. OPP believes that this is an appropriate application of the FQPA factor.

7. Description of 'What's Safe' Is Misleading

Comment. Another commenter believed that the statement that "...when the 99.9^{th} percentile of estimated exposure is equal to or less than the PAD, the vast majority of people would not be exposed to pesticides in their food at unsafe levels" is misleading. The commenter stated, firstly, that this statement assumes that the analysis accurately estimates exposure while in reality the many

assumptions built into the analysis are designed to significantly overestimate exposure. The commenter also pointed out that even if one were to "assume that the method somehow produces an accurate estimate, 0.1 percent of the people might be exposed to levels at least 100 (and sometimes 1000) fold lower than that shown to have no effect in laboratory animals" and that "this is different than actually being exposed to 'unsafe levels.'"

Response. None of the assumptions "built into" the analysis are designed to significantly overestimate exposure. When "real world" residue levels are collected from market basket surveys, any standard assumptions are abandoned and the real-world values are routinely inserted into OPP's analyses and are believed to represent exposure estimates that best represent actual exposure levels to the population. Little additional refinement of these types of estimates could be performed and OPP believes that these types of estimates accurately reflect exposures to the U.S. population. Also, OPP's statement is still accurate even when less refined data is available (*e.g.*, crop residue field trials): even when these less refined estimates are used, "the vast majority of people are [still] not exposed to pesticides in their food at unsafe levels."

OPP agrees with the commenters second point and has stated in the original *Percentile Policy* document that exceeding the threshold of concern does NOT automatically mean that people are being exposed to unsafe levels of pesticide residues in food or that an individual will necessarily experience any adverse effects. OPP recognizes that "exceeding safe levels" does not necessarily imply "unsafe levels." Therefore, OPP has modified the original statement to read as follows: "...when the 99.9th percentile of estimated exposure is equal to or less than the PAD, the vast majority of people would not be exposed to pesticides in their food that exceed safe levels."

8. Comparative Ratios Between 99th and 95th Percentile Are Too Simplistic

Comment. One commenter took exception to the statement that "the size of the population exceeding the PAD at the 99th or 95th percentiles would be 10 and 50 times, larger, respectively, than the number at the 99.9th percentile" stating that these ratios are "overly-simplistic" and that they do not match the DEEM predictions seen[15]. The commenter indicates that "ratios between the 99.9th percentile estimate and 95th percentile estimate may be less than 10-fold."

[15]Interestingly, another commenter stated that with regard to the percentile value for regulation, "numerical reality is relatively straightforward, and is described in the draft policy paper in Section III."

Response. The estimates of "10 and 50 times larger" at the 99th or 95th percentiles than at the 99.9th percentile refer to *population size* and not to estimated exposure levels. The commenter is entirely correct that the ratio between the 99.9th and 95th percentile *exposure* estimates may be less than 10 fold. By mathematical definition, however, the size of the exposed populations must be 10 and 50 times larger for the 99th and 95th percentiles, respectively[16].

9. Which Exposures Does 99.9 Refer To?

Comment. Several commenters indicated that it was unclear which exposures the 99.9 percentile refers to. Specifically, the commenters indicated that the following statement was very confusing and unclear in the *Percentile Policy* document:

> If the 99.9th percentile of acute dietary exposure (together with exposure from other non-dietary, non-occupational sources), as estimated by probabilistic (*e.g.*, Monte Carlo) analysis, is equal to or less than the Population Adjusted Dose (PAD) for the pesticide, OPP will determine that the safety standard of FFDCA sec. 408(B)(2)(A) is met with respect to acute dietary risk. However, if the analysis indicates that exposure at the 99.9th percentile exceeds the PAD, OPP will conduct a sensitivity analysis to determine to what extent the estimated exposures at the high-end percentiles may be affected by unusually high food consumption or residue values. To the extent that one or a few values from the input data sets seem to "drive" the exposure estimates at the high end of exposure, OPP will consider whether these values are representative and should be used as the primary basis for regulatory decision making. The Office will also examine the consequence of removing such high-end food consumption or residue values when estimating the 99.9th percentile of exposure.

Methods until now, they stated, have aggregated chronic dietary background exposure to short term residential exposures. This is a significant and substantial departure from EPA policy and practice to date. If OPP is shifting its position or

---

[16]For example, for a population size of "x", the ratio of the size of the population exceeding the 99.9th percentile to the size of the population exceeding the 95th percentile can be calculated as follows:

$$\frac{(1 - 0.95)x}{(1 - 0.999)x} = \frac{0.05x}{0.001x} = \frac{0.05}{0.001} = 50$$

The ratio of the size of the population exceeding the 99.9th percentile to the size of the population exceeding the 99th percentile can similarly be calculated to be 10.

policy, then it should be publicized in its own paper and open to comments and input.

Another commenter cautioned OPP against making specific highly-exposed subpopulations the focus of modeling efforts since this would result in regulatory decision making based on percentiles fare beyond the 99.9th percentile of any "general population." There is a natural tendency to want to focus on groups who are "really at risk," but it is not clear, the commenter stated, that there has been adequate consideration of how this focus of groups of special interest will interact with the 99.9th percentile. One could imagine an analysis, the commenter pointed out, "that focused on the 99.9th percentile of children who eat apples, live near apple orchards (and are thus at risk from spray drift), and drink water that has a high probability of being contaminated with the pesticide of interest." While the 99.9th percentile of this group might have an unacceptably high exposure, the commenter argued that the "at risk" population is arguably very small.

Response. OPP is clarifying the statement in the policy document. The 99.9^{th} percentile policy at this time applies only to probabilistic exposures to pesticide residues in food. At present, estimates of exposure through drinking water and residential uses are not sufficiently developed to warrant inclusion in a probabilistic assessment. Thus, the 99.9^{th} percentile policy will only apply to exposures through food. To produce an aggregate assessment, these will be combined with estimated exposures through drinking water and residential uses by means of OPP's current interim aggregation policy. When exposures through drinking water and residential uses are sufficiently refined to be incorporated into probabilistic evaluations, these will be aggregated and assessed and OPP may use a different percentile threshold.

F. Suggestions for the Future Directions

✔ Evaluation Steps/Investigative Work
✔ Mechanics of DEEM simulation
✔ Compounding of Uncertainty Factors
✔ Use of 1989-92 CSFII vs. 1994-96 CSFII
✔Overestimation of Residue Values

Overview. *A number of commenters provided information which was unable to be directly "categorized" into one of the above headings, but nevertheless were useful to the Agency for potential consideration in the future.*

1. Future Investigative Work

Comment. One commenter provided comments on OPP's planned

additional investigative work (the three steps) concerning the DEEM model and commended OPP "for recognizing and acknowledging so forthrightly that there may be significant problems with using a percentile as high as the 99.9th." However, he stated it was not clear how OPP would be able to eliminate (or do a sensitivity analysis) the high-consumption events given that the DEEM program currently in use is unable to do this.

Another commenter provided input on the several steps which had been proposed to evaluate the dietary methods, suggesting that "since results of the analyses are greatly affected by the number of crops and residue values used in the analysis, these steps should be repeated for numerous (30+) compounds with different labels and underlying residue data sets before any conclusions can be made" and that "[c]onclusions based on single compound or label are likely to be inaccurate or misleading."

Response. OPP appreciates the commenters' recommendations. The first commenter is correct in doubting the ability to eliminate the high-consumption events given that the DEEM program currently in use is unable to do this. At the time this protocol was proposed, OPP expected that the software would be modified to perform this task. Subsequent investigation revealed that this would both be a difficult activity to perform and would have limited meaning due to potential problems with the mathematical "reweighting" which would necessarily have to occur. Thus, it was decided that this task could not be performed.

Nevertheless, a number of activities designed to provide a better understanding of the most critical elements of the methodology were performed (several of which did address the issue of the "high consumption" individual and its effect on the tails of the distribution). A brief description of these follows:

- <u>How many iterations are necessary to achieve reasonable stability in the exposure estimates?</u>

  Both specific (controlled) investigations of this issue and our experience with numerous recently conducted exposure analysis performed as part of the reregistration process demonstrate that 1000 iterations are more than adequate to produce reasonably stable estimates of exposure at the 99.9th percentile with the DEEM software. Overall, we have found that exposure estimates produced following 1000 iterations do not vary by more than approximately 1-3% for any given subgroup. Thus, exposure estimates are generally reliably stable following 1,000 iterations, even for large data sets.

❖ Given an adequate number of iterations, are DEEM exposure estimates reasonably reproducible at the 99.9th percentile?

To test the reproducibility of DEEM exposure estimates at the 99.9th percentile, a total of 120 DEEM runs (of 1000 iterations each) were performed on multiple computers over different days. The results of these tests and the statistical analysis are illustrated in Figure 1.

Each of the 120 points in the figure represents a separate DEEM run. We note that estimates of exposure following in each case 1000 iterations are stable and reproducible and follow the expected normal curve. In this specific case, the mean exposure at the 99.9th percentile was 3035.4 x 10^{-6} mg/kg bw/day with a standard deviation of 15.1 x 10^{-6} mg/kg bw/day. This means that for this specific case, approximately 95% of the time the estimated exposure from DEEM will be within 1% of the true value and 99% of the time will be within 1.3% of the true value. The analysis thus demonstrates that, provided an adequate number of iterations are performed, the stability and reproducibility of the DEEM software exposure estimates are not significant sources of concern in risk assessment.

❖ Is there evidence of consumption extremes (*e.g.*, 98th, 99th, 99.9th percentiles) being inconsistent or unrealistic or unrepresentative?

This issue was, at least in part, addressed by one commenter who performed an extensive analysis of reported consumption of apples, apple juice, peaches, and pears by young children and a second commenter who performed a similar analysis with apples, peaches, fresh green beans, apple sauce, apple juice, grapes, and pears (as taken from USDA's CSFII). This analysis, combined with OPP's own analyses and experience, suggests that extreme consumption events are not pervasive in the CSFII dataset and, except on rare occasions, are not singly or primarily

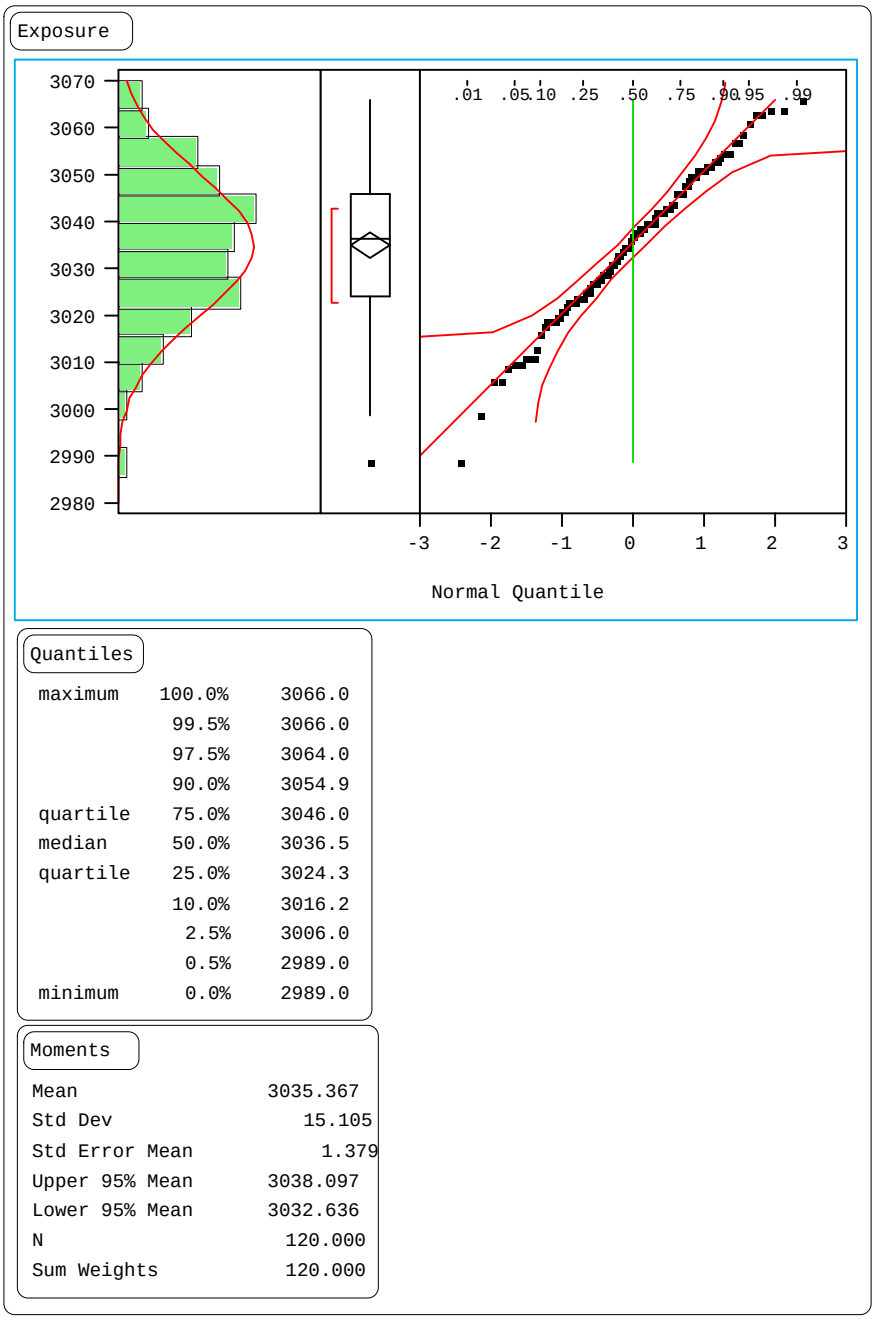

Figure 1. Results of 120 DEEM Runs to Determine Reproducibility of DEEM Exposure Estimates

responsible for driving high-end exposure estimates. Nevertheless, as originally stated in the *Percentile Policy* document and reiterated here in OPP's response to comments, risk assessors in OPP will fully characterize any exposure estimates which appear to be driven by high or unusual reported consumption.

2. Generate Better Consumption and Residue Data

Comment. One commenter recommended that future development efforts generate more reliable, more certain data for input which might include more controlled, experimentally based human consumption determinations and expansion of the PDP residue data. The commenter suggested that a Blue-Ribbon panel of biostatisticians and ecologists be engaged to sort out the individual risk vs. population risk mixing and propose better techniques for evaluating individual risk.

Response. In response, OPP supports using the most reliable data available and has been a consistent advocate for expansion of the PDP program. With respect to improvement in the consumption estimates generated by USDA using more "controlled, experimentally-based" determinations, it is unclear to OPP what form this would take and there would be the fear that such controlled, experimentally-based determinations might, in and of themselves, alter the responses or behaviors that are being measured. Given the unparalleled size, extent, and nature of the USDA CSFII, the extensive QA/QC which it undergoes, and the fact that it reports on consumption of thousands of statistically-selected individuals on a repeated basis at a taxpayer cost of millions of dollars, OPP believes that it is unlikely that any realistic alternative will be proposed.

3. Issues Concerning the Details and Mechanics of DEEM

Comment. One commenter raised a number of insightful (and very valid) issues concerning the details and mechanics of the DEEM simulation. For example, he correctly pointed out that the DEEM simulation sums all the eating events of an individual of a given food form within a day and then assigns this total consumed to a randomly-selected residue value. For example, DEEM would use the total grams of fresh peach consumed by an individual on a given day and multiply this total by a randomly-selected residue value. Specifically, the commenter stated:

> Examples of these concerns can be seen in how the acute models handle the assignment of residue values to eating occasions within a 24 hour period. People commonly eat three or more times during a day. If a commodity such as an apple is eaten, it may be consumed at only one or as many as all eating

occasions occurring on that given day. Although different apples with different residue values are probably eaten at each of these different eating occasions, the current model assigns all the day's apple consumption the same residue chosen during the Monte Carlo process. This may be adequate for population descriptions, but if individual risks are being characterized this biases the daily value estimates for these individuals. Exposure estimates for the high daily consumption individuals would result from the assumption that the residue value was the same in each apple eaten. If the chosen residue value was a high-end value, all apples would be represented with that residue value. It is highly unlikely that an individual would consume more that one 99.9th percentile residue containing apple in a single day. In fact, the daily exposure value should be a summation of the individual eating occasion consumption amounts combined with the varying residue levels on the apples, rather than consumption amounts summed and then combined with a single residue value.

Response. OPP recognizes this issue and notes that there are two basic ways a model can account for this. One way (as is currently done and described above by the commenter) is to assume, for example, that all fresh apples eaten by a given individual in a given day contain the randomly assigned residue concentration. A second method assumes that each fresh apple eaten by a given individual on a given day is randomly picked *and is independent of the residue concentration in any other apple consumed by that individual on that day*[17]. The first method is more consistent with consumed apples being from one source and sharing the same treatment history (*i.e.*, each fresh apple eaten by a given individual in a given day is from the same bag of apples purchased from a grocery store, each baked apple eaten by an individual in a given day comes from the same apple pie, etc.) and is most appropriate when the residue values selected are composite values (an average of many items). The second method is more consistent with each apple consumed in a given day by a given individual being independently acquired from different sources (*i.e.*, no apple consumed necessarily shares the same treatment history). The DEEM version currently used by OPP uses the first method, but a recent software update permits the second method (where only items consumed during a given eating occasion are assumed to share the same treatment history) to be used as well. OPP will accept analyses performed with both models. In these cases, OPP will consider the result from both analyses in making a decision and characterize the results. To date, comparison of results using both methods do not suggest that differences are significant.

[17]Under the first method, analyses are performed by food form. Thus, residues in each of the various food forms (*e.g.*, fresh apples, baked apples, canned apples, etc.) are chosen independently. That is, a single residue is chosen for the day's fresh apple consumption, a second *independent* residue is chosen for the day's baked apple consumption, and a third *independent* apple residue value is selected for the day's canned apple consumption.

4. Timing Between Eating Occasions Not Incorporated

Comment. The commenter pointed out that the consideration of timing between eating occasions is not incorporated into the current model. Depending upon the pesticide of interest, the bioactive analytes may have all been eliminated by the time of the next eating occasion and exposed individuals will have "returned to baseline". In this case, summation of exposures over a 24-hour period would not be appropriate.

Response. The commenter is correct both about the way in which the Monte Carlo methodology of DEEM handles the timing of consumption and the theoretical consequences. This is an area in which OPP is developing guidance it hopes to issue in the future. Specifically, OPP is considering use of a "rolling average" which would be more flexible in incorporating the temporal course of exposure and toxicokinetics. Extensive toxicological data would need to be submitted to demonstrate that an individual "returned to baseline" prior to the next exposure event. As OPP moves toward considering cumulative exposure, this is likely to be a less significant issue in future as the "time between exposure events" will reflect exposure events to any pesticides with a common mechanism of toxicity. In addition, a current assumption is that there is no "carry-over" of exposure from one day to the next and that appropriate consideration of the timing of exposure events could, under certain conditions, lead to increases in exposure level estimates.

5. Limitation in the CSFII Survey

Comment. The commenter identified a limitation in the CSFII survey in that it fails to identify whether a high consumption eating event is a non-daily, but frequent, event or whether is represents a rare event in an individual's lifetime. If a person did report a consumption event, he stated that we do not know with what frequency this occurs and correctly points out that neither the consumption survey design nor the current model adequately addresses this question.

Response. OPP agrees that this limitation on the data in CSFII is a source of uncertainty. OPP, however, notes that this uncertainty could go in both directions (*i.e.*, the consumption survey can miss high consumption days just as easily as it hits them). In fact, for small subpopulations and items which are consumed by only a limited portion of the population, the survey is more likely to *miss* high-end consumption events than to overreport them[18]. The CSFII survey is

[18] This is one of the reasons for the asymmetric confidence intervals around the high-end consumption estimates. One cannot be reasonably certain, for example, that if only 100 individuals in a random survey have

a "cross-sectional" survey in that the survey interviews are conducted (or cover) a limited time period (generally 1-3 days). The survey is large enough that it likely captures consumption events that are only rarely experienced, but still indeed occur. It is unlikely that the survey has captured all (or even the most extreme) consumption events and some undoubtedly will have been missed. This is the very nature of a survey. *OPP believes that the survey adequately represents the individual one-day consumption patterns of the U.S. population*. OPP acknowledges that the survey does not capture all high-end consumption events, but the events it does capture are adequately weighted by the survey design. OPP also believes that reported consumption events should not be arbitrarily discarded *a priori* simply because they appear to exceed what some regard as "the norm" or an "expected" high-end value (however defined). If any risk assessment is driven by a high-consumption event, that event will be carefully evaluated by the risk assessor and be fully characterized in the assessment for the risk manager.

6. Inter- and Intra-Species Uncertainty Factors

Comment. One commenter provided input on the issue concerning the 10X animal to human factor and the 10X inter-individual variability in human sensitivity factor. The commenter stated that if a 10X safety factor is needed to ensure safety at 95%, it seems clear that a smaller safety factor would be in order at the 99.9th percentile. The commenter provides an argument as to why, when the percentile of regulatory concern changes from 95th percentile to 99.9th percentile, the 10X intra-individual safety factor is implicitly increased 50 fold.

Response. The 95th percentile estimated exposure obtained from DRES model output was not really the 95th percentile value as it assumed 100% crop treated and tolerance (or near tolerance residues). In most cases, it was far higher than the 99.9th percentile and possibly higher than even the actual 100th percentile, the highest exposure in the actual population. This was demonstrated for a specific case in the original document when it was shown that exposures at the "95th percentile" using the DRES assumptions of 100% crop treated and tolerance level residues were an order of magnitude higher than estimates of the 99.9th percentile exposure generated using DEEM.

7. Switch to 1994-96 Consumption Data

---

reported consuming a specific food commodity on a given day, that a high-end (*e.g.*, 98th or 99th percentile) consumption of the entire population has been truly captured. In the 1989-92 CSFII survey, for example, the 99.9th percentile estimate for consumption of baby food apple sauce by infants is only 163 g, and for the 1994-96 survey the corresponding estimate is only 180 grams or approximately 1 ½ small baby food jars. It is unlikely that this adequately approximates the true 99th (or larger) percentile of the infant population and this type of uncertainty would be reflected in an asymmetric confidence interval with a long tail to the right.

Comment. Several commenters stated that the Agency is currently using the 1989-91 CSFII data and encouraged the Agency to switch to the 1994-96 data as soon as possible. The 1994-96 data is thought by USDA to contain fewer errors and other problems than the 1989-91 survey.

Response. OPP agrees with the commenter and is rapidly proceeding in that direction. We note that USDA and EPA have recently completed a recipe translation of the 1994-96 CSFII data so that as-eaten food forms (*e.g.*, cheese pizza) can be translated to their component parts on a raw agricultural commodity basis (*e.g.*, wheat, tomatoes, milk, etc.). This translation has been peer reviewed and is currently undergoing final review in OPP. It is expected that these publically-developed data will be quickly incorporated into the DEEM dietary exposure software the Agency is currently using to perform its risk assessments. At about the same time, OPP expects to be able to incorporate into the 1994-96 CSFII the results of the Supplemental Children's Survey, a USDA-sponsored survey which will add approximately 5000 additional children in various age groups from infants to adolescents (up to 18 years) to the existing 1994-96 sample survey. This will result in nearly doubling the number of surveyed infants and children upon which OPP's exposure estimates are based. The incorporation of the new USDA/EPA food translation (recipes), the addition of the Supplemental Children's Survey, and the switch-over to the 1994-96 CSFII are expected to occur simultaneously during the second quarter of calendar year 2000.

8. Overstatement of Residue Levels

Comment. One commenter commented on the systematic overstatement [of residues] due to use of field trial data or tolerance levels and encouraged OPP to develop approaches for scaling down exposure levels derived from field trial data to derive anticipated residue values that are comparable to existing PDP exposure levels. He cited a recent submission which compared tolerance values (1.5 ppm) to average field trial values (0.399 ppm), to highest PDP monitoring data value for a composite sample (0.36 ppm), to the highest market basket study value (0.052 ppm). Instead of relying on field trial data when no PDP or FDA monitoring data or market basket survey data are available, the commenter encouraged OPP to develop an alternate approach by examining crops for which there are both field trial data and PDP data, noting the ratio range of the values, and developing some sort of rough rule for short-term use to adjust field trial values downward for crops where PDP data are lacking.

Response. This is a valuable suggestion. OPP is already taking steps toward making more extensive use of PDP data and actual pesticide use and usage data. Given recent criticism concerning the use of USDA's consumption survey and the precision of OPP's high-end exposure estimates, OPP is cautious about

relying for our risk assessments on any "rough rules of thumb." We have recently released a Standard Operating Procedure which details under what conditions the PDP data can be extended to other similar crops which would be expected to have a similar distribution of residue values (U.S. EPA, 1999b). The "surrogate table" lists literally dozens of commodities to which limited PDP data can be extended. In some cases, PDP data can be extended to entire crop groups.

In addition, OPP has recently released for public comment two science policy papers entitled "Guidance for the Conduct of Bridging Studies for Use in Probabilistic Risk Assessment" (U.S. EPA, 1999c) and "Guidance for the Conduct of Residue Decline Studies for Use in Probabilistic Risk Assessment" (U.S. EPA, 1999d) which detail how information from crop field trials can be extended to better reflect actual use practices.

Another recently released science policy paper entitled "Data for Refining Anticipated Residue Estimates for Organophosphate Pesticides" is also available which provides information on additional ancillary studies such as cooking and processing studies which can be used to further refine residue estimates (U.S. EPA, 1999e). In short, OPP is expanding the number and nature of the methods available to produce refined residue exposure estimates. It is important that all of OPP's risk assessments use the best data available and that any alternate approaches be fully investigated, widely applicable, and transparent.

G. Beyond the Scope

A number of comments were beyond the scope of the current document. These are addressed below.

1. Decompositing Techniques

Comment. Several commenters commented on the introduction of "spurious" high-end residue values from use of decompositing techniques. They encouraged the Agency to further investigate the use of MaxLIP (Maximum Likelihood Imputation Procedure) which, they stated, was demonstrated to provide better estimates of single serving residue distributions from composites and also featured additional capabilities to adopt this methodology. Another commenter expressed support for use of a decompositing procedure and believed that the affected Agencies can reach concurrence on a decompositing protocol that produces a realistic distribution of residue values in individual samples.

Response. This issue is beyond the scope of the *Percentile Policy* and is covered in another science policy paper which was issued separately for comment (U.S. EPA, 1999f). OPP has presented the MaxLIP program along with an

alternative program (RDFgen) to the SAP and expects to receive a formal report from the SAP on this topic in May, 2000.

H. Incorporating Toxicology as a Probabilistic Distribution

A number of commenters responded to the question relating to incorporation of probabilistic toxicity components into our risk assessments. These are detailed below.

1. Consider Toxicity as Well as Exposure

Comment. One commenter in response to Question 7 stated that it was essential to consider the effect (toxicity) as well as the exposure from a probabilistic perspective for valid risk assessment using Monte Carlo approaches. This has been very clearly recognized, the commenter stated by the SAP. Methodologies for this type of analysis are being established within regulatory circles. The Office of Water under the Clean Water Act and the Great Lakes Initiative has developed refined methodologies for dealing with interspecies and intraspecies variability in toxicological effect thresholds. Further refinements, the commenter continued, have been advanced specifically for pesticides by the Aquatic Risk Mitigation and Dialogue Group and are forthcoming from the EPA ECOFRAM process. These precedents in the area of ecotoxicological risk, the commenter recommended, should be considered for their application and extension to human toxicologic considerations

One commenter, citing an SAP recommendation, suggested that toxicity distributions be used when conducting risk assessments and that "the use of simple NOEL values based on arbitrary doses used in the toxicology study has the effect of artificially lowering the aRfD and thus adding yet another layer of conservatism in the analysis."

The commenter also stated that the impact of uncertainty inherent in toxicity data should also be considered. There is uncertainty in the estimate of the NOAELs and LED10's and uncertainty in the uncertainty factors applied to the NOAELS and LED10's. When a high-end exposure estimate is used for comparison to the RfD, the uncertainty in the both the exposure estimate and RfD should be taken into account when judging risk. The commenter remarked that EPA's Scientific Advisory Panel recently recommended that margins of safety built into the toxicity evaluation process be considered when selecting a basis for regulation and endorsed the idea of using the entire distribution of exposure and toxic effects rather than a single "worst case" endpoint.

Response. OPP agrees with the commenters but notes that at this time there are not yet standard procedures for the Agency to implement a probabilistic

component to toxicity assessment.

III. References

Hill, Robert, H., Susan L. Head, Sam Baker, Marianne, Gregg, Dana B. Shealy, Sandra L. Baily, Cynthia C. Williams, Eric J. Sampson, and Larry L. Needham. 1995. Pesticide residues in urine of adults living in the United States: reference range concentrations. *Environmental Research* 71:99-108.

Interagency Board for Nutrition Monitoring and Related Research (IABNMRR). 1995. Third Report on Nutrition Monitoring in the United States: Volume 1. U.S. Government Printing Office, Washington, D.C. 365 pp.

U.S. EPA 1992. Federal Register Notice. "Final Guidelines for Exposure Assessment." 57 FR 22888. May 29, 1992.

U.S. EPA, 1996. Office of Pesticide Programs. "Acute Dietary Exposure Assessment Office Policy. June, 1996.

U.S. EPA 1998a. Federal Register Notice. "Assigning Values to Nondetected/Nonquantified Pesticide Residues in Human Health Dietary Exposure Assessments" (dated 11/30/98). Draft. 63 FR 67063-67066. December 4, 1998.

U.S. EPA 1998b. Federal Register Notice. "A Statistical Method for Incorporating Nondetected Pesticide Residues into Human Health Dietary Exposure Assessments" (dated 11/30/98). Draft. 63 FR 67063-67066. December 4, 1998.

U.S. EPA, 1999a. Federal Register Notice. "Choosing a Percentile of Acute Dietary Exposure as a Threshold of Regulatory Concern." Draft. 64 FR 16962. April 1, 1999.

U.S. EPA, 1999b. Office of Pesticide Programs. Health Effects Division. "Translation of Monitoring Data". HED Standard Operating Procedure (SOP) 99.3. March 26, 1999.

U.S. EPA, 1999c. Federal Register Notice. "Guidance of the Conduct of Bridging Studies for Use in Probabilistic Risk Assessment" (dated 7/29/99). Draft. 64 FR 42372. August 4, 1999.

U.S. EPA, 1999d. Federal Register Notice. "Guidance of the Conduct of Residue Decline Studies for Use in Probabilistic Risk Assessment" (dated 7/29/99). Draft. 64 FR 42372. August 4, 1999.

U.S. EPA. 1999e. Federal Register Notice. "Data for Refining Anticipated Residue Estimates Used in Dietary Risk Assessments for Organophosphate Pesticides" (dated 3/26/99). Draft. 64 FR 16967. April 7, 1999.

U.S. EPA. 1999f. Federal Register Notice. "Use of the Pesticide Data Program in Acute Dietary Assessment" (dated 5/5/99). Draft. 64 FR 28485. May 26, 1999.

IV. List of Commenters

Abbotts, John. 6/7/99.

Byrd, Daniel M. III, CTRAPS (Consultants in Toxicology, Risk Assessment, and Product Safety). 6/07/99.

Clarke, David P. Chemical Manufacturers Association. 6/7/99.

Cockrell, William Patrick. Florida Farm Bureau Federation. 6/7/99.

Craigmill, A.L. and Michael A. Kamrin, University of California and Michigan State University. 5/05/99.

Dong, Michael, California Department of Pesticide Regulation. 4/16/99.

Environmental Working Group. 6/7/99.

Fix, Lori, Agriculture Division. Bayer Corporation. 6/7/99.

Franklin, C.A., Pest Management Regulatory Agency. Health Canada. 5/11/99.

Gersich, F.M. Dow AgroSciences, 6/3/99.

Ginevan, Michael E. M.E. Ginevan and Associates. 6/6/99.

Groth, Edward and Lawrie Mott. Consumers Union and Natural Resources Defense Council. 6/6/99.

Jenkins, W.B. North Carolina Farm Bureau Federation. 6/7/99.

Laurie, Jack. Michigan Farm Bureau. 6/7/99.

Layton, Raymond J. DuPont Agricultural Products. 6/1/99.

Maslyn, Mark, FQPA Implementation Working Group. 6/7/99.

Pennsylvania Farm Bureau. 6/7/99.

Phillips, Jennifer L. Rhone-Poulenc Ag Company. 6/3/99.

Priestley, Frank. Idaho Farm Bureau Federation. 6/7/99.

Reigart, J. Routt, Medical University of South Carolina. 5/26/99.

Tomerlin, J. Robert. Novigen Sciences International. 6/4/99.

Whitacre, Dave. Novartis. 6/8/99.