Responses to Major Comments on Technical Support Document

Public Health Goal

For

Benzene

In Drinking Water

Prepared by

Reproductive and Cancer Hazard Assessment Section and
Pesticide and Environmental Toxicology Section
Office of Environmental Health Hazard Assessment
California Environmental Protection Agency

June 2001

TABLE OF CONTENTS

TABLE OF CONTENTS..... II

INTRODUCTION..... 1

RESPONSES TO MAJOR COMMENTS RECEIVED 2

COMMENTS FROM FORMAL EXTERNAL PEER REVIEWERS 2

Lawrence Livermore National Laboratory 2

University of Southern California..... 4

COMMENTS FROM INTERESTED PARTIES 7

Chevron Oil Co., on behalf of the Western States Petroleum Association (WSPA) (transcribed from oral testimony at the November 1999 public workshop)..... 7

United States Environmental Protection Agency, Office of Water 8

United States Environmental Protection Agency, National Center for Environmental Assessment (NCEA). 9

Western States Petroleum Association (WSPA), submitting comments prepared by Exxon Biomedical Sciences, Inc. (January 2000) 12

Western States Petroleum Association (WSPA) (April 2000)..... 28

REFERENCES NOT CITED IN THE PHG DOCUMENT 34

INTRODUCTION

The following are responses to major comments received by the Office of Environmental Health Hazard Assessment (OEHHA) on the proposed public health goal (PHG) technical support document for benzene. We have selected the more important or representative comments for responses. Comments appear in quotation marks where they are directly quoted from the submission; paraphrased comments are in italics.

These comments and responses are provided in the spirit of the open dialogue among scientists that is part of the process under Health and Safety Code Section 57003. For further information about the PHG process or to obtain copies of PHG documents, visit the OEHHA Web site at www.oehha.org. OEHHA may also be contacted at:

Office of Environmental Health Hazard Assessment
P.O. Box 4010
Sacramento, California 95812-4010
(916) 324-7572

RESPONSES TO MAJOR COMMENTS RECEIVED

Comments from Formal External Peer Reviewers

Lawrence Livermore National Laboratory

Comment 1. "Accuracy. The information provided in the review of benzene is, based on my knowledge, quite comprehensive. Information is presented from peer-reviewed studies published through 1998. The literature provided is very inclusive through 1996, most of the data published in 1997 that I am aware of is included and a good representation of data published in 1998 is provided, except possibly late in 1998. In general, I find the information to be very accurate and to provide a good distillation of the toxicology of benzene in both animals and humans. The information is particularly well documented for the occupational studies underway in humans.

"The weakest area of the entire document is in the discussion of DNA adducts where some information on identification of both protein and DNA adducts, and the relevance to humans is not presented. However, there is enough information provided for the reader to develop an understanding of the mechanisms of how benzene leads to cancer and the likely genotoxic risk posed by benzene without it."

Response 1. The Office of Environmental Health Hazard Assessment (OEHHA) discusses protein and the deoxyribonucleic acid (DNA) adduction in several places in the public health goal (PHG), including the sections entitled "Biomarkers of exposure" (pages 11-12), "Carcinogenic mode of action" (pages 60-65) and "Low dose linearity" (pages 79-82). In response to these comments, the following changes were made: 1) The following sentence was added to the mode of action section: "Interestingly, recent work has suggested that NQO1 protects the marrow against hydroquinone- and 1,2,4-benzenetriol-induced toxicity through an unexpected mechanism, inhibition of high molecular weight DNA adducts (Wiemels et al., 1999)." 2) The additional reference of Singer and Hang (1999) was added to that section. 3) The following sentence was added to the low dose linearity section: "Recent studies have examined DNA and protein binding of very low doses of benzene (5 μg/kg body weight) administered to different species of mice and rats (Mani et al., 1999). The authors noted that their results were consistent with the hypothesis that the metabolic capability to metabolize benzene to toxic metabolites contributes to the difference in benzene's ability to elicit a carcinogenic response in different species." OEHHA reviewed the recent literature and located new published papers on DNA adducts that were not already cited in the above noted sections. These papers did not impact the interpretation of the data and were consistent with the findings presented in the document.

The commenter also states that the document does not discuss relevance to humans of the adduct data. However, OEHHA noted in the document that DNA adducts may give rise to mutations and chromosomal damage, and that binding to histones, topoisomerase II, and tubulin may give rise to chromosomal breakage or aneuploidy. The document describes the proposed mechanisms by which DNA adduction may be misrepaired leading to mutation (pages 60-64). It also describes observations of benzene-oxide DNA and protein adducts in humans, and benzoquinone-protein adducts in humans exposed to benzene (pages 62-63).

Comment 2. "Appropriateness of data. The data provided covers the most recent information on sources of benzene, exposure routes, exposure levels, metabolism among species (mice, rats, hamsters, non-human primates and humans), toxicity and dosimetry. This and the supporting information provides an understanding of the philosophy used to recommend risk and acceptable exposure levels, and the

methods used to arrive at recommendations. I find the selected data quite appropriate and consistent with our present understanding of benzene toxicity. The most significant concern on my part is in its carcinogenic effects which is the focus of the document. The non-carcinogenic concerns (cytopenia's, etc.) are, in my view, early stages in the development of leukemia and mechanistically are indications that damage is occurring."

Response 2. Comment noted.

Comment 3. Other Factors. "The discussion and presentation of data related to the public health goal for benzene is quite comprehensive. While interactions with other solvents that are directly related to benzene toxicity are discussed (hydroquinone, etc.), the potential for the interaction of benzene with other solvents/pollutants and how this can affect benzene toxicity is only weakly considered. Some mention of toluene and its inhibition of benzene metabolism/toxicity is made but no mention is made relative to other compounds known to be present in California's water supplies (methyl tertiary-butyl ether (MTBE), metals, etc.). Further, these factors have only been looked at using relatively high doses and such interactions need to address what is known or expected to occur based on our understanding of mechanisms, at drinking water levels. Toluene, for example, may have no effect when at low levels because the CYP2E1 enzyme known to be critical to both compounds metabolism will not be saturated.

"Secondly, while discussion is provided on individual variability, very little data is available on the likely variation in response to benzene exposure among individuals because very little is known about the range in activity of all the steps in benzene metabolism, including repair of the damage. Thus, a major uncertainty is how sensitive is the most sensitive individuals likely to be and how large will this sensitive population be."

Response 3. These are very important points. Workers and the general population are exposed to benzene and a wide range of other chemicals and physical agents that may affect the susceptibility or variability of individual responses to benzene. We agree that the presence of toluene and other chemicals that compete for cytochrome P4502E1 and other metabolic enzymes are not likely to have an effect at low levels, since metabolism would not be saturated (as discussed in the Risk Characterization section). The document attempted to summarize the sources of inter-individual variability and benzene-induced hematotoxicity (pages 64-71). The potential effects of co-exposure of benzene and ethanol, infectious agents, other known leukemogens, dietary sources of phenol, catechol and hydroquinone (the phenolic metabolites of benzene) and micronutrients were discussed. However, clearly more research is needed to understand the impact of multiple chemical exposures with benzene.

Comment 4. "Non-quantitative statements about dose or effect are frequently made throughout the document that should be quantified. For example, pg. 12, under absorption it states: "Humans exposed experimentally to low to moderate doses…". These doses should be given as a range or exact number."

Response 4. Additional quantitative information has been added. On page 12, following the sentence "Humans exposed experimentally to low to moderate air concentrations of benzene" the phrase "(approximately 1.7 to 32 ppm, Appendix C, Table 3)" was added. On page 12, following the sentence "These observations are consistent with animal inhalation studies using low concentrations of benzene" was added "(11 to 29 ppm, Appendix C, Table 2)." On page 84, following "low exposure levels" was added "(e.g., less than 10 ppm)." On page 112, to the following sentence, "However, the results are not statistically significant, which may be due to the low-level exposures" was added "(< 15 ppm-yrs)." On page 122, the phrase "high exposures to benzene in gasoline may be less hazardous" was changed to "high exposures to gasoline (300 to 2000 ppm) which contains benzene may be less hazardous . . ."

Comment 5. "I feel this is a very well-written and accurate document and I don't find many problems with it. The conclusion that the drinking water standard should be lowered is consistent with the recent data that toxicity is occurring under exposure conditions lower than previously observed."

Response 5. Comment noted. As a point of clarification, the PHG value developed here (0.15 ppb) is actually very similar to the previous health based toxicity value for benzene in drinking water (0.18 ppb, the proposed maximum contaminant level (PMCL)) recommended by the state of California (DHS, 1987). The current California drinking water standard, the maximum contaminant level (MCL) (1 ppb), is based in part on technical feasibility.

University of Southern California

Comment 1. "The only major aspect I feel needs some more thought is the use of the human inhalation data to model the human ingestion data. Personally, I prefer the use of the more extensive animal data on oral ingestion studies to model the human ingestion situation. Perhaps one way to resolve this would be to make the calculations based on the animal ingestion data as well, and to compare them to the calculations based on the human inhalation data. I understand that the authors would rather use human data, and I support this, but I am worried that the route of administration could affect the results. Hence, probably a fair thing to do would be to make the calculations both ways, and see if the results are close. If so, then I suggest averaging them."

Response 1. Per the reviewer's suggestion, we have added a comparative analysis of cancer potency estimates from animal ingestion studies and human inhalation studies of benzene (see responses to comment 3).

Comment 2. "Appropriateness of the data set selected for deriving the public health goal (PHG) and the supporting information. The data sets the authors retrieved and summarized are highly appropriate for deriving the public health goal. This data includes the scientific literature on the metabolism of benzene, the toxicity of benzene in humans and animals, the genetic toxicology of benzene in humans, the carcinogenicity of benzene in animals, and the epidemiology of leukemia induction in humans. This data is extensive, the review the authors conduced is very comprehensive, and the conclusions they derived from these data sets are appropriate."

Response 2. Comment noted.

Comment 3. "In general, the authors evaluated the data they chose appropriately and they interpreted it correctly. The data they chose do support the dose-response assessment. The only question I have here is whether they would be better served for carcinogenesis to also utilize the animal data on ingestion of benzene to also derive the public health goal. The reviewer suggests that they conduct a parallel calculation using the animal ingestion data, and determine whether the PHG for cancer induction varies significantly from using the data on animal ingestion vs. using the human data on inhalation. Of course, the animal data on ingestion is most relevant to the human situation of drinking water contaminated with benzene, but the human inhalation data is useful because it is actual human data and factors out inter-species comparisons. The reviewer wonders if a comparison of the calculations by the two methods would be instructive. If they yield calculated concentrations that are close, then this would increase the confidence of the scientific and regulatory community in using these PHG values. If they are very different, then some further thinking on this issue would be required."

Response 3. In the draft PHG, we included a section entitled "Risk estimates from animal carcinogenicity studies." In that section, OEHHA notes that potency estimates derived from animal studies have been reported elsewhere (ARB, 1984; Crump and Allen, 1984; U.S. EPA, 1985). These include cancer potency estimates derived from oral gavage studies as well as inhalation studies, including gavage studies conducted by the National Toxicology Program. Only one additional cancer bioassay of benzene (Farris et al., 1997) has been conducted since these assessments. However, since complete pathology was not performed on all of the animals in the Farris et al. study and since the reported tumor incidences did not appear appreciably different from earlier bioassays, OEHHA did not develop a cancer potency estimate from the Farris et al. study.

In response to the reviewer's suggestion, we have rewritten the section to provide a comparative analysis of the oral cancer potency estimates obtained from the animal studies (including those studies employing oral exposures) to the estimates obtained from the Chinese Worker Cohort, Pliofilm Cohort. A table (Table 29) has been added to easily compare the upper-bound oral cancer potency estimates from the most sensitive tumor sites in animals to those for leukemia in humans. The following conclusion is reached: " ...cancer potency estimates from the animal studies, including those based on oral exposures, are consistent with the estimates obtained from the human studies. This comparison provides strong support for the validity of the proposed PHG for benzene."

Comment 4. "This document is so extensive, that no further information needs to be considered in reaching the proposed PHG, in the opinion of the reviewer. In fact, there is so much data here, that some effort should be made to condense this document in specific places where possible, of course at the discretion of the authors and where suggested in the specific comments section below."

Response 4. We condensed the summary and several other sections in the document that the reviewer identified as lengthy.

Comment 5. "The only uncertainty that the reviewer wishes to be considered is to conduct a parallel calculation on the concentration yielding de minimis cancer risk (one/one million) using the animal ingestion data, since this matches the route of administration of benzene to humans via drinking water. Then, please compare this calculated concentration to that derived from using the human inhalation data for benzene. The difference between these two calculations would be instructive and constitute an uncertainty that should be addressed in the opinion of the reviewer."

Response 5. The cancer potency estimates derived from animal gavage studies are consistent with the oral potency estimates derived from the human occupational studies. The document has been revised to make the point more explicit. Interestingly, the potencies from human inhalation and animal ingestion studies are nearly identical, after correction for differences in absorption by route of exposure. Also, inhalation exposures resulting from benzene-contaminated tap water use (e.g., showering and dishwashing) are predicted to result in a roughly equivalent dose as would occur through direct ingestion (Table 3).

Comment 6. "Overall, this is an excellent document. It is very strongly researched and the assessment of the literature retrieved was done in a very intelligent fashion. The authors clearly command the areas of benzene metabolism, toxicity, and carcinogenesis, epidemiology, and risk assessment. All the sophisticated calculations and strong literature review and analysis lend a strong academic strength to this document. This reviewer's confidence in the totality of this document is very high. The document could be improved in a number of places by condensation where possible to allow the reader to focus more on the most important points."

Response 6. Comment noted.

Comment 7. "Specific Comments. Summary. In this section, some concise explanation should be made as to why the PHG is now proposed to be 0.00014 mg/L, when the current California MCL is 0.001 mg/L. Is the current California MCL set too high?"

Response 7. The basis of the MCL (technical feasibility) has been noted.

Comment 8. "Introduction. On page 3, para. 4, the authors should discuss whether there is any difference in the tumor site and incidence when benzene is given to animals by inhalation vs. by ingestion to strengthen their use of the inhalation data in humans rather than the ingestion data in lower animals. Clearly, this is a tough choice, and the authors chose to use humans to reduce this uncertainty. However, they should be careful, since the route of administration as they well know can have a profound effect on the site and incidence of tumors. The reviewer understands the choice the authors made. They should perhaps rationalize it more strongly in this section to convince the reader that this choice was correct, or at least the best possible choice among alternatives."

Response 8. We have added the following sentence to the introduction, describing the choice of basing the PHG on human data. "Fourth, cancer potency estimates derived from animal studies of orally administered benzene were essentially the same as potency estimates derived from the occupational cohort studies (Table 29)." An additional sentence was also added to the risk characterization section discussing the consistency of the animal ingestion-based estimates with the human inhalation-based potency estimates.

Comment 9. "Metabolism and Pharmacokinetics, Absorption - This section is well-written. However, there is no referencing of the scientific papers from which these facts come."

Response 9. This paragraph serves as a summary and directs the reader to an extensive discussion of the topic in Appendix C. No revisions are needed.

Comment 10. "Genetic Toxicity…the thought should be imparted here that clastogenesis is more important than benzene-induced mutagenesis. There is not a lot of data on benzene-induced mutagenesis in cultured mammalian cells. Most of benzene's carcinogenicity is thought to be mediated through chromosome breakage.

"Carcinogenicity - Page 39, para. 3: Did the authors calculate a relative risk of leukemia from these Turkish data? … This section is really long. It is probably best to accept the length in order to obtain the comprehensive coverage. However, if possible, the authors should try to shorten this section." *The sections on Benzene and Childhood Leukemia, Carcinogenic Mode of Action and Inter-Individual Variability and Benzene-Induced Hematotoxicity are comprehensive, but also very long. Some attempts should be made to condense them.*

Response 10. In response to the consistent call for condensing the length of the various sections, OEHHA has condensed the discussions of the human cancer data, childhood leukemia, and inter-individual variability. However, OEHHA was unable to condense the carcinogenic mode of action section to any considerable degree without loss of content. With regard to the specific comment on genetic toxicity, this concern was addressed in the original document. "Benzene is both a mutagen and clastogen, although its clastogenic activity is likely to be of greatest concern (see Mode of Action section)." Also, a cancer potency was estimated from the Turkish data (Aksoy 1980, 1985, 1994). Please

see Table 28 and the section entitled “Leukemia risk estimates from other cohort studies of benzene-exposed workers.”

Comment 11. “Carcinogenic effects: The discussion from pages 73-81 is very long; some effort could be made to condense it somewhat. The discussion on low dose linearity is very necessary and is done very well. This is an excellent section. It is, however a little long and could be condensed somewhat. The conclusion of no strong evidence to support a threshold for benzene carcinogenesis is appropriate.”

Response 11. Comments noted. We have condensed the discussion of the “Approaches taken in previous assessment” as recommended by the reviewer.

Comment 12. *Calculation of PHG: the reviewer would like to see the authors calculate a C value based on the animal ingestion data for both carcinogenic and non-carcinogenic effects*.

Response 12. OEHHA did not calculate a public health-protective concentration, C, for benzene in drinking water based on the cancer potency estimates derived from oral animal carcinogenicity data. However, in response to the reviewer's comments, a comparison of the cancer potency estimates derived from animal and human data was provided (Table 29). OEHHA summarized that comparison by stating "Thus, the cancer potency estimates from the animal studies employing oral exposures, 0.04 to 0.2 $(mg/kg\text{-}d)^{-1}$, are consistent with the estimates obtained from the human studies, 0.1 $(mg/kg\text{-}d)^{-1}$ (Table 29). This comparison provides support for the validity of the PHG for benzene." In other words, the value of C would be similar based on either the human or animal datasets.

Comment 13. “Risk Assessment: Page 114, para three, line 2: The reviewer suggests the authors consider using the terminology “sublinear,” rather than “supra-linear.” “Supra-linear sounds too much like “Super-linear” and confuses the message the authors are trying to convey. In general, this section is very good, and brings the document to a very strong conclusion and explains some of the underlying complexities of the calculations. However, it is somewhat long and should be condensed to hold the attention of the reader.”

Response 13. For greater clarity, we have provided a definition of supralinear in each section that the term is used. For example, “supralinear” was replaced with “supralinear (i.e., the trend in relative risk falls below a linear trend with dose).” We acknowledge that the terms sublinear and supralinear are used differently in different fields, and among different groups.

Comments from Interested Parties

Chevron Oil Co., on behalf of the Western States Petroleum Association (WSPA) (transcribed from oral testimony at the November 1999 public workshop)

Comment 1 (paraphrased). *WSPA feels that acute non-lymphocytic leukemia (ANLL) should have been used as the tumor type for analysis, instead of total leukemia (i.e., all subtypes of leukemia as a related class of diseases). Work by Crump (1994) and by Irons, Ross and Stroebel were cited to indicate that ANLL is the tumor type most associated with benzene exposure*.

Response 1. The PHG document provides an extensive review of the evidence on which human cancers are associated with benzene exposure. OEHHA agrees with WSPA that ANLL is the predominant tumor type in most studies of benzene-exposed workers. However, strong evidence exists to suggest that benzene causes other forms of leukemia as well. Indeed the two studies selected as best and used as the

basis of the PHG, the Pliofilm Cohort (Crump, 1994) and the Chinese Worker Cohort (Hayes et al., 1997), indicated increased relative risks (RR) of non-ANLL of 1.9 and 2.0, respectively. Additionally, the draft PHG outlined data indicating that benzene likely alters the early stem and progenitor cells leading to all subtypes of blood cells. Moreover, practical considerations of diagnosis and reporting also argues for the use of the more general category of total leukemia. (Also see responses to related comments from WSPA's written submission (below)).

Comment 2 (paraphrased). *WSPA believes the available data warrant the use of a non-linear approach to low dose extrapolation of the cancers induced by benzene. WSPA cited the Crump (1994) analysis of the Pliofilm cohort as evidence that a non-linear relationship was best.*

Response 2. The PHG document details the available data on low dose linearity. These data were applied to the criteria for selection of a linear or non-linear approach as outlined in the U.S. Environmental Protection Agency Guidelines for Carcinogen Assessment, which indicated a linear approach was most justifiable. OEHHA has also concluded that the Crump (1994) analysis best supports a linear approach. Crump (1994) used two sets of exposure estimates for the analysis of the benzene-exposed workers (Pliofilm Cohort): (1) Crump and Allen (1984) exposure estimates and (2) Paustenbach et al. (1992) exposure estimates. For the Paustenbach estimates, both linear and non-linear models adequately fit the data, with the non-linear model providing only slightly better fit. OEHHA believes that Paustenbach et al. estimates are likely to be unreasonably high (Utterback and Rinsky, 1995, also see responses to WSPA's post-workshop comments below). OEHHA therefore did not use the Paustenbach estimates to calculate leukemia risk. With respect to the Crump and Allen exposure estimates, the Crump (1994) analysis states: "Whereas dose-responses were essentially linear when the Crump and Allen (1984) exposure matrix was used…."

Comment 3 (paraphrased). *WSPA believes that the National Cancer Institute (NCI) study of Chinese workers is not suitable for risk assessment. WSPA states that U.S. EPA agrees with its unsuitability since it did not use the Chinese data in its 1996 reassessment of benzene (finalized in 1998). A summary of potential concerns with respect to exposure estimation, including co-exposures, confidence in exposure assignment, appropriateness of the control group, and control for smoking was also provided.*

Response 3. Many of the issues outlined have been addressed in the PHG. Smoking was not controlled for in the Chinese worker study, but smoking was also not controlled for in the study of the Pliofilm workers. OEHHA is not aware of any retrospective cohort study of benzene that has controlled for smoking status.

With respect to U.S. EPA's analysis, NCI's dose-response assessment of the Chinese workers (Hayes et al., 1997) was not available to U.S. EPA at the time that U.S. EPA drafted their assessment in 1996. Moreover, U.S. EPA (1998) did not re-analyze any data, but merely compiled estimates reported in the literature. Cancer potency estimates using the Chinese data were not available in the published literature at the time U.S. EPA released its report. The suitability of the NCI study of Chinese workers for risk assessment, including issues of co-exposures and exposure assignments, is discussed below in responses to WSPA's written comments.

United States Environmental Protection Agency, Office of Water

Comment 1 (paraphrased): *In Appendix C of the draft PHG, OEHHA provided an equation for gastric absorption of benzene and briefly described the results of the Sabourin et al. (1987) study and other*

studies to support the use of 100 percent absorption of benzene by humans at low doses. However, a more detailed description of the Sabourin et al. study along with a table of actual absorption data from the Sabourin et al. study would be more appropriate and convincing than the equation; an example of such an approach was provided.

Response 1. We agree. Appendix C has been revised. OEHHA adapted the language provided by U.S. EPA for this purpose.

Comment 2. "In the discussion of the Sabourin et al. (1987) inhalation absorption data, benzene that was absorbed but exhaled as unmetabolized benzene was not regarded as having been absorbed. Based on the oral data, this leads to an underestimation of absorption especially at higher doses where a greater percent of the total absorbed benzene is exhaled as unmetabolized benzene."

Response 2. We agree. This point was directly made in the last sentence of the Sabourin et al. (1987) study description. "Benzene taken up, but subsequently excreted in exhaled air is not counted in the absorbed fraction; the resulting retention values thus somewhat underestimate total absorption."

Comment 3. "In looking at the human data on benzene absorbed via the inhalation route, the equation used was as follows: % absorbed = 100 x $(C_{inhaled} - C_{exhaled})/C_{inhaled}$. The benzene in the exhaled air is a combination of the benzene not absorbed by the lungs and that which was absorbed and then exhaled unchanged. The text correctly notes that the inability to account for the fraction of the benzene in the exhaled air that was absorbed leads to an underestimate of the % absorbed. Given this fact the data in Table 2 seem to support a percent absorption that is greater than 50 %. The text should acknowledge the fact that use of 50 % for the inhalation route is a conservative estimate of the dose…."

Response 3. OEHHA disagrees. The section of Appendix C entitled Discussion of Inhalation Absorption discusses the variability and uncertainty of the available data related to this point. OEHHA notes "There is very good overall agreement among the studies, with most supporting an absorption factor close to 50 percent." OEHHA also notes that human-based physiologically-based pharmacokinetics modeling conducted by Bois et al. (1996) indicated a mean absorption value of 57 percent. Therefore, we feel the draft PHG document's statement, "Thus, occupational and environmental exposure studies suggest that an absorption fraction of 0.50 is a good estimate," to be appropriate.

Comment 4. "The IRIS document that seems to have been the basis for the Appendix C has results from the humans studies (pp. 159, 160) summarized in a table. The Table is mentioned in the PHG text (p. 159) but is not included."

Response 4. Table 2 was provided in the original draft but was located following the references. We have moved the table within the body of the text to avoid confusion.

United States Environmental Protection Agency, National Center for Environmental Assessment (NCEA).

Comment 1. "The differences between the USEPA and that of the California EPA (CalEPA) are as follows:

1) CalEPA sets one standard that is protective for cancer and non-cancer adverse health effects. It is called a public health goal or PHG.
2) CalEPA's PHG of 0.00014 mg/L to produce a risk of 10^{-6} is equivalent to USEPA's range of 1 to 10 ug/L. CalEPA's concentration level is 7.1 to 71.4 times lower than that of USEPA to produce the same risk of 10^{-6}. Their PHG is more protective than our risk range concentrations.
3) CalEPA included all types of leukemia in its estimate.
4) CalEPA derived its "best upper bound estimate" of lifetime risk by combining cohort (Hayes et al., 1994 [sic]) with that of Pliofilm workers (Rinsky et al., 1987; Paxton et al., 1994).
5) CalEPA extrapolation from inhalation to oral ingestion was similar to that of the USEPA, however, the calculation of exposure included direct ingestion, dermal exposure and indoor air (showering) plus the consumption of benzene-contaminated tap water in homes. These additions raised the intake to 4.7 L. The USEPA believes it is an overestimate of the water intake.
6) CalEPA assumed 50 % absorption via inhalation, 100 % absorption via ingestion and 2.0 L of water consumption. This assumption is in agreement with US EPA (1999)."

Response 1. In the benzene PHG document, OEHHA developed a non-cancer toxicity value, akin to the U.S. EPA's RfC, as well as a cancer-based estimate for benzene in drinking water. The cancer-based estimate, being lower, was set as the PHG. The PHG is akin to U.S. EPA's maximum contaminant level goal, which by default U.S. EPA sets to zero for carcinogens. OEHHA's cancer-based health criterion would also be adequately protective against non-cancer effects.

OEHHA selected a single, best estimate for the PHG (0.15 μg/L) but also provided a range of reasonable estimates consistent with the data and alternate assumptions (0.08 to 1.2 μg/L). U.S. EPA (1998) did not generate cancer potency estimates *per se*, but selected a range of cancer potency values from the literature. If one compares the range of water concentration estimates based on the cancer potencies selected by U.S. EPA (1 to 10 μg/L) to OEHHA's range of PHG values (0.08 to 1.2 μg/L), the ranges overlap. Indeed, as U.S. EPA points out, the best estimate selected by OEHHA (0.15 μg/L) is about 7- to 70-fold lower than the range selected by U.S. EPA. There are three primary reasons for the differences in the two sets of health values for benzene:

1. U.S. EPA used 2.0 L water consumption per day. OEHHA estimated 4.7 L/day, which accounts for inhalation and dermal absorption from showering and other tap water uses. The 4.7 L/day was based upon exposure studies of households in which the water source was contaminated with gasoline. The estimated value was also nearly identical to the estimate obtained from CalTOX, a multimedia total exposure computer program (DTSC, 1999). OEHHA believes that U.S. EPA underestimates the potential risk because it does not account for water-related exposures other than ingestion. This difference in approach leads to a 2.4-fold difference in the estimated health values for benzene in drinking water.

2. U.S. EPA's range of estimates was based upon application of linear risk models to the Pliofilm cohort (Crump, 1994). Crump (1994) utilized the two sets of exposure estimates developed by Crump and Allen (1984) and Paustenbach et al., 1992, but did not use estimates described by Rinsky et al., 1987. No estimates using the Rinsky exposure estimates and a linear risk model are available in the published literature for the latest update of the Pliofilm cohort. OEHHA believes the Paustenbach estimates are likely to be high (Utterback and Rinsky, 1995) and did not use them in its risk estimation. Use of the Rinsky estimates with all data points, or the use of the Crump or Rinsky exposure estimates after removing the highest exposed workers (as OEHHA has done), yields cancer potency estimates that are approximately 2-fold higher than estimates based on all data points using the Crump estimates and 4-fold higher than estimates using the Paustenbach estimates. The fact that OEHHA included in its range of reasonable cancer potency estimates, estimates based on the

China study has little bearing on the comparison, since the best potency estimates from the Pliofilm cohort and the China study differed only by about 25 percent.

3. OEHHA based its best estimate for the PHG on the upper 95 percent confidence bound on the cancer potency estimate, whereas U.S. EPA utilized central estimates as the basis of its potency range. OEHHA notes that there is considerable human inter-individual variability in the susceptibility to benzene-induced hematotoxicity (reviewed in the PHG in a section entitled Inter-individual Variability and Benzene-induced Hematotoxicity). OEHHA feels well-justified in using the 95 percent confidence bound as its best estimate and in using both central and upper bound estimates in its range of reasonable estimates of the cancer potency. It is interesting to note that if U.S. EPA had included the 95 percent confidence intervals in its range of acceptable estimates (Table 4 of U.S. EPA, 1998), the top of U.S. EPA's range of estimates would have been more than two-fold higher than OEHHA's best estimate.

In point three of the comments, U.S. EPA suggests that there is a difference in the selection of tumor type from the U.S. EPA and California Environmental Protection Agency (Cal/EPA) assessments. U.S. EPA (1998) selected a range of lifetime risk estimates based on the Pliofilm cohort data and linear risk models (Crump, 1994), which ranged from 7.1×10^{-3} to 2.5×10^{-2} ppm^{-1}. The 7.1×10^{-3} ppm^{-1} risk estimate is based on acute myeloid and monocytic leukemia combined (AMML), while the 2.5×10^{-2} ppm^{-1} (actually reported as 2.4×10^{-2} ppm^{-1} in Crump, 1994) is based on total leukemia (all forms of leukemia as a related class of diseases). OEHHA based its lifetime risk estimates on total leukemia, but noted that "Cancer potency estimates for other cancer endpoints, such as acute non-lymphocytic leukemia, generally fell within this range." Thus, it appears that U.S. EPA and Cal/EPA are in agreement on this issue.

Comment 2. "This document has several deficiencies from EPA's perspective. First and foremost among these is the lack of any discussion of derivation of the inhalation unit risk range from the USEPA 1998 document entitled, "Carcinogenic Effects of Benzene: An Update". Although the IRIS benzene 1999 summary is cited as a source of information regarding the level of evidence of carcinogenicity (page 124), yet no discussion is evident in the CalEPA document regarding the derivation of the USEPA unit risk range from this same document. ... The information about the derivation of USEPA's risk range is strangely absent."

Response 2. The U.S. EPA (1998) document is cited numerous times in the PHG document, and the studies evaluated by U.S. EPA, including the one from which U.S. EPA selected its cancer estimates, were described. U.S. EPA is correct in noting that OEHHA did not report the risk range described by U.S. EPA in its 1998 Update. In response to this comment, OEHHA added a paragraph in the section entitled "Approaches taken in previous assessments" which describes the U.S. EPA (1998) document and the range of lifetime cancer risk estimates selected.

Comment 3. "In its estimate of the risk of leukemia to the general population, the authors averaged the geometric mean of the Chinese cohort study of Hayes et al. (1997) with that of the Pliofilm workers of the Rinsky et al. (1987) cohort to obtain a number that became the basis for development of the PHG (page 112). The USEPA felt that there were still sufficient concerns about the use of Chinese cohort study in the derivation of the cancer unit risk range."

Response 3. Comment noted. OEHHA, however, does not share the same degree of concern as U.S. EPA regarding the Chinese study (Hayes et al., 1997). For detailed discussion of this topic, the reader is directed to the responses to written comments prepared by Exxon Biomedical, Inc and submitted by WSPA (below).

Comment 4. "The CalEPA also assumed there was a substantial component of exposure achieved through benzene ingestion and dermal absorption chiefly through showering based upon two studies by Beavers et al. (1996) and Lindstrom et al. (1994) that more than doubled the contribution of these routes of exposure to the benzene water equivalence used in the derivation of their risk estimates. These studies are based upon one subject each. Several concerns (smoking and other potential exposures) and the use of only two subjects to derive benzene exposure estimates via dermal and ingestion routes make the use of these studies questionable at best. Furthermore, the USEPA has cited several studies that indicate that dermal absorption accounts for 1% of applied dose (page 10, USEPA, 1999)."

Response 4. U.S. EPA (1999) estimates that the proportion of one's total exposure to benzene-contaminated water due to dermal absorption is small (less than one percent), based on observations in several studies (page 10, USEPA, 1999). However, OEHHA estimated that approximately 20 percent of a typical person's exposure from benzene-contaminated tap water use would be through dermal absorption primarily from showering and washing (Table 3). This estimate is based on human dermal transfer factors derived from the very studies that U.S. EPA cites (DTSC, 1999). Estimates of the percentage of the dose due to dermal transfer from household drinking water contaminated with benzene included: 11 percent (Lindstrom et al., 1994, a U.S. EPA study), 30 percent (Beavers et al., 1996), and 16 percent (CalTOX (DTSC, 1999)).

Moreover, the U.S. EPA water number does not incorporate the fact that exposures via inhalation from benzene-contaminated tap water will occur from volatilization to indoor air from showering and washing. Available estimates (Lindstrom et al., 1994: Beavers et al., 1996 and CalTOX (Table 3 of PHG)) all indicate that roughly an equal dose will be received via inhalation as from direct ingestion. Indeed, in the IRIS documentation (U.S. EPA, 1999, page 15), U.S. EPA states "...this estimate is a risk factor for ingested benzene, and is not sufficient to account for total exposure to drinking water. For development of a drinking water safe concentration, the risk due to inhalation of volatilized benzene from drinking water and to dermal uptake must be added to ingestion risk (Beavers et al., 1996; Lindstrom et al., 1994)."

Comment 5. (paraphrased). *NCEA notes several concerns with the use of the Tsai et al. (1983) study of benzene-exposed refinery workers as the basis of the non-cancer health concentration for drinking water. Specific concerns included the fact that the average exposure duration was only 7.4 years, an apparent healthy worker effect, and lack of clarity in what was considered a "normal" versus "abnormal" blood chemistry profile. NCEA feels that this study lacks sensitivity.*

Response 5. The chronic reference exposure level (REL) for benzene, recently finalized for the California Air Toxics Hot Spots Program, was used as the basis for the non-cancer health value for benzene in drinking water. The chronic REL received extensive public review and comment as well as formal external scientific peer review from the Air Resources Board's Scientific Review Panel. The chronic REL document has been reproduced as Appendix B in the PHG document. However, OEHHA notes that the no-observed-effect-level of 0.53 ppm derived from the Tsai et al. study is consistent with estimates derived from animal studies and is consistent with the lowest-observed-effect-level of 7.6 ppm from the Rothman et al. (1996) study, which U.S. EPA utilized in the development of its RfD/RfC.

Western States Petroleum Association (WSPA), submitting comments prepared by Exxon Biomedical Sciences, Inc. (January 2000)

Comment 1. **Cover letter:** "Clearly, in the fifteen years since the adoption of the current [benzene] value the amount of information, both epidemiological and mechanistic, relevant to an assessment of the leukemogenic potency of benzene has grown tremendously. Because of this increased knowledge base

WSPA believes that is scientifically appropriate and defensible to abandon two assumptions that were used in the original potency calculation and which remain essentially unchanged in the current assessment. These are the use of all the leukemias as the relevant endpoint of benzene-induced effects and the continued use of the default assumption that the dose response for benzene is linear in the low dose range. The attached documents will provide supporting documentation and reasoning to support this position.

"The Technical Support Document for the proposed benzene PHG has done a good job of surveying the available information on the leukemogenicity of benzene. Possibly because of the massive size of this database the technical support document has not included a critical appraisal of much of the published literature. This lack of critical evaluation is most evident in the acceptance at face value of the publications describing epidemiological and other studies of a cohort of occupationally-exposed Chinese workers as published by researchers from the National Cancer Institute (NCI) and the Chinese Academy of Preventive Medicine (CAPM). Although investigations of this cohort have great promise for advancing our understanding of benzene-induced leukemia and other hematopoietic disorders, there are a number of concerns that until resolved render this cohort useless for a quantitative assessment of the leukemogenic potency of benzene. Foremost of these concerns is with the exposure estimates that have been NCI/CAPM researchers. Although exposure reconstruction is an endeavor fraught with difficulty at best, a close comparison of the publications from this study reveal a number of inconsistencies. These contractions create serious question about the validity of the exposures as presented. The attached "Comments" document lays out many of these problems and supports the contention that the current exposures are potentially seriously underestimated. It is important to note that although the comments in the attached document refer to unpublished sources, a peer-reviewed paper is in press in the December issue of "Regulatory Toxicology and Pharmacology". The paper is by Otto Wong, entitled "A Critique of the Exposure Assessment in the Epidemiologic Study of Benzene-Exposed Workers in China Conducted by the Chinese Academy of Preventive Medicine and the US National Cancer Institute" and should be available in January 2000.

"In summary, WSPA agrees with the recalculation of a cancer potency value for benzene using the updated Pliofilm cohort. We believe, however, that the data from the cohort itself as well as supporting biochemical information about the mechanisms of secondary leukemogenesis and clastogenicity support the use of nonlinear dose response model as well as acute non-lymphocytic leukemias (ANLL) as the relevant biological effect for dose response calculation. It is also our position that the current exposure information from the NCI/CAPM is insufficient to support a quantitative estimate of the leukemogenic potency of benzene at this time."

Response 1. The response to this comment serves as an overview of the fundamental changes made to the draft PHG in response to the totality of comments received from WSPA. Detailed responses regarding the specific concerns raised are discussed below. In sum, OEHHA agrees that a considerable amount of research of benzene carcinogenicity has been published since California's last evaluation of benzene's cancer potency in 1987. Despite these efforts, benzene's carcinogenic mechanism of action remains elusive, likely because there are multiple mechanisms by which benzene can give rise to leukemias. We do not agree that this research now indicates that benzene causes only ANLL in humans. We agree that ANLL and myelodysplastic syndromes are most strongly associated with exposure to benzene, but there is strong evidence to indicate that benzene causes other forms of leukemia as well. Also, OEHHA does not agree that the current understanding of benzene mechanism of carcinogenicity provides strong evidence to move from a linear-to-low dose assumption. Indeed, there is considerable evidence to suggest that linearity is most appropriate. Thus, in light of these observations, and in applying this evidence to the carcinogenic guidelines for selecting between linear or non-linear approaches, OEHHA feels it is appropriate to employ a low-dose linear model. This conclusion was also reached by U.S. EPA (1998). However in response, OEHHA has noted in the risk characterization

section that some researchers believe that the true dose-response relationship for benzene-induced leukemia is non-linear or has a threshold of response.

With respect to the use of the Chinese Worker Cohort (Hayes et al., 1997), OEHHA does not share the same degree of concern as WSPA on the quality and merits of the study. As discussed below, many of WSPA's strongest criticisms, and the concerns raised by recently published critiques (Wong, 1999; Budinsky et al., 1999) demonstrate a misunderstanding of how NCI used the individual benzene exposure estimates in the dose-response study (Hayes et al., 1997). These are discussed in detail below. Some concerns remain unanswered, and OEHHA is aware that NCI is preparing a manuscript to help clarify some of the misunderstandings and criticisms that have been published. OEHHA concludes that the PHG is consistent with cancer potency estimates derived from either the Chinese Worker Cohort or the Pliofilm Cohort.

Comment 2. "For the China Study cohort, a linear model was also chosen, using the "exposure level in the lowest dose group as a point of departure for assuming linearity from that point to the origin," and dropping the highest dose group (25 or more ppm-year) to prevent a "likely underestimate of [the] true risks" (page 11). The "best" estimate of the lifetime of the lifetime risk was determined to be 0.056 ppm^{-1} (Table 27), with a reasonable range of 0.00063-0.087 based on other results using linear extrapolation of all dose groups, the absolute risk model, and estimates based on US EPA methods (EPA 1996)."

Response 2. The commenters have adequately characterized our approach, except with respect to treatment of the Chinese data. To clarify, the best potency estimate from the Hayes et al. (1997) data was based on a linear model after the top dose group was removed. Since the dose response curve is supralinear, inclusion of the highest exposed group may potentially underestimate the true risks at low doses. A separate approach was based on the U.S. EPA proposed carcinogen guidelines (U.S. EPA, 1996). In that case, a non-linear relationship was assumed and the lowest dose group was selected as the point-of-departure from which a straight line was drawn to the origin. This method resulted in the highest lifetime risk estimate (0.087 ppm^{-1}), which was included in the reasonable range of estimates, but did not serve as the basis for the PHG.

Comment 3. "The calculations and assumptions by OEHHA are clearly stated, and basically appear to have been incorporated correctly. However, some of the assumptions are very conservative, and two in particular warrant comment:

"Giving equal weight to the Pliofilm and Chinese data: The Pliofilm calculations are based on a reasonable number of cases and on individual exposure data. The data were appropriately analyzed (except as noted below) using several different analyses and assumptions. The Chinese data, while based on many cases, were analyzed by OEHHA based on only 4 data points (1 control and 3 exposure group relative risks)- in fact the estimate was made (arguably incorrectly) using only 3 data points because the highest dose group was deleted. To equally weight these two data sets is not reasonable, since the robustness of the underlying data is very disparate."

Response 3. OEHHA disagrees. With the use of linear models and Poisson regression, theoretically one would not expect the results to be very different whether grouped or individual data were used. Indeed, OEHHA explored this comparison for the Pliofilm data (see Table 22). Potency estimates derived from grouped data (four data points as Exxon calls it) were only slightly different (zero to ten percent across several data sets) from estimates derived from individual data. Thus, use of the grouped data is reasonable, although use of the individual data provides greater rigor.

Comment 4. "Dropping high exposure groups in the estimates from the Pliofilm and Chinese data: OEHHA appears to be seeking a linear relative risk response with increasing dose. These are two data sets, which do not exhibit linear response in the dose – but they do exhibit a linear response in the log of the dose. OEHHA's characterization of these data as "supra-linear" is a mischaracterization. In fact, the Pliofilm data shows substantial sublinearity in the low dose range according to many dose metrics.[3,4,7] The most likely dose response function is an S-shaped curve, displaying sublinearity in the low dose region and supra linearity in the high dose region. The same may hold for the China cohort, but it is impossible to evaluate this from the published data, since more than half of the cohort's person time is lumped in the lowest dose group (see Hayes et al. 1997).[5] To make the data conform to a linear response OEHHA drops the data which deviates from linearity. OEHHA justifies the dropping of the high dose Pliofilm data (page 97) by saying if the data are supralinear then including the top doses will underestimate the risk at low doses. Then it is stated that the same thing will happen in the Chinese data so it is justifiable to drop the data in both data sets. The data are not shown to be supralinear, and the authors do not discuss why they need to have a linear response. This is not a scientific approach to data analysis.

"The differences in the results for dropping the high dose groups are substantial (see below) – although the authors state that "the cancer potency estimates based on Crump (1994)[3] and Rinsky (1987)[4] exposure matrices are essentially the same with the top dose group removed" (page 97, end of 3rd paragraph).

"OEHHA Estimated Upper 95% CL for lifetime Risk (ppm-1) (Table 25 & 27)

	Rinsky	Crump	Chinese
All data	0.048	0.022	0.011
Top removed	0.044	0.045	0.056

"It is invalid to splice up the studies and use only the data that fits with OEHHA's preference for linearity. Instead, non-linear modeling should have been used (see further comments below)."

Response 4. OEHHA disagrees. We reviewed the evidence for low dose linearity (see section entitled "Low dose linearity") and compared that evidence to guidance for selecting a linear and non-linear approach. In OEHHA's opinion, the evidence does not support moving from a linear approach. In this regard, U.S. EPA agrees (U.S. EPA, 1998). Moreover, as stated in the PHG, "visual inspection of the dose-response relationships of the two studies selected as the basis of this assessment, the Pliofilm Cohort and the Chinese Worker Cohort, suggest a linear relationship at low doses (Figures 5 and 6)."

Exxon's assertion that the Pliofilm data suggests a sublinear response curve is not supported by data. The small number of cases and the small number of person-years at risk at lower exposures provide extremely low statistical power. As stated above, the Crump (1994) analysis of the Pliofilm data noted that for the Paustenbach estimates, both linear and non-linear models adequately fit the data, with the non-linear model providing only slightly better fit, and when using the Crump and Allen exposure estimates, "dose-responses were essentially linear . . ." (Crump, 1994). OEHHA believes that these data do not support a conclusion of "substantial sublinearity."

Comment 5. "In selecting studies on which to base its benzene potency estimate, OEHHA provides only a brief discussion of the rationale (p 92). From 20 epidemiology studies of cancer/benzene associations, the Pliofilm cohort and the NCI-CAPM China study data were chosen as a basis for the slope factor calculations. Selection criteria included emphasis on cohort over case-control design, large number of individuals at risk, good exposure estimation, and preference for studies with dose response information and a separate unexposed control group. Among the two datasets chosen, the Pliofilm cohort data are much more widely considered suitable for deriving cancer potency estimates for benzene exposure. The authors mention several attributes of the Pliofilm data, such as minimal co-exposures, clear dose response

relationships, high relative risks, and extensive scientific scrutiny over several years. Because of common use of the Pliofilm data in risk assessments, including several reassessments of the exposure data, the strengths and limitations of using these data in risk assessments are well understood in the scientific community.

"In contrast, the China study is untested in its usefulness for benzene risk assessment. While a cursory review might suggest that it has advantages over the Pliofilm cohort, particularly a significantly larger sample of over 75,000 workers and potential to study lower exposures in some segments, it also has disadvantages which limit the extent to which it can assess benzene-associated risk. The most limiting factors involve the exposure information.[8,9] These will be discussed below, followed by comments about additional study limitations.

"In using the China study, OEHHA is at variance with other agencies that have performed risk assessments on benzene. In particular, the US EPA cited the obvious weakness of the Hayes et al. (1997)[5] study involving co-exposures. EPA states: 'Although not specifically stated, concurrent exposures to many other chemicals, some hazardous, must have occurred because benzene was used as a solvent for paints, varnishes, glues, coatings, and other products that were part of the occupational environment for this cohort. These products contain a myriad of chemicals, of which some were undoubtedly carcinogens.' In addition, EPA states: 'During the earliest period, only 3% of the exposure estimates were based on actual measurements. Accuracy and precision of these subjective estimates was unknown. Such bias, if present, might have contributed to the supralinearity at higher benzene concentrations evident in the results of this study.'[10]"

Response 5. As stated in the draft PHG, the Chinese Worker Cohort (Hayes et al., 1997; Dosemeci et al., 1994; Yin et al., 1996, many others) and the Pliofilm Cohort study (Rinsky et al., 1981, 1987; Paxton et al., 1994a) both have advantages and disadvantages. For example, in both studies, as with most retrospective cohort studies, there is poor exposure information in the early years of the cohorts. In both studies, exposure estimates for early years were recreated based on scant early-year measurements, and extrapolation from later year measurements using other relevant information and assumptions. The exposure estimates in the early years of the NCI-CAPM study were based on 3 percent actual measurements but this was increased to 68 percent by the end of the study. By contrast, the percentage of estimates in the early years for the Pliofilm study is likewise very low; however, it was increased to 100 percent by NIOSH in later years. Also, the exposure assignments in the NCI-CAPM study were based on many sources of information, not merely the grab samples (Dosemeci et al., 1994).

It is true that the Pliofilm Cohort has had a longer track record for use in risk assessment. This fact does create somewhat greater confidence in the estimates derived from them. However, the Chinese Worker Cohort has been a subject of active study and publication since 1987. The first dose-response assessment of the Chinese data was published by Hayes et al. (1997); thus, it is not surprising that it is "untested for its usefulness in risk assessment." It should be noted that the Hayes et al. (1997) paper was published after U.S. EPA's initial draft of the Carcinogen Reassessment Update for benzene was released (finalized as U.S. EPA, 1998). As noted above, U.S. EPA (1998) did not estimate cancer potencies *per se*, but selected from a range of potencies from the literature. To OEHHA's knowledge, no cancer potency estimate for the Chinese workers has been published prior to the draft PHG assessment.

Concerns regarding the exposure estimates for Chinese workers are discussed in Comment 6.

Comment 6. "OEHHA claims that one reason to use the China study data is because there is "good exposure information (especially for most recent ten years of follow-up)." On the contrary, several major issues suggest that the estimates do not meet quality standards than can be reasonably expected in quantitative risk assessment. OEHHA mentions, but does not discuss, one such concern – i.e., the lack of [information on] co-exposures, which may confound benzene-leukemia associations. Other issues not

mentioned by OEHHA include questions about some of the exposure assessment procedures, as well as indications that the quantitative estimations of benzene exposure are not of sufficient validity to warrant their use in the dose response step in risk assessment. Of greatest concern is the likelihood that the exposures are underestimated, which would overestimate risk and call into question the validity of a risk assessment.

"Co-exposures: Workplace exposures other than benzene have not been evaluated and reported for this study population and this is probably the greatest shortcoming of the study. Indeed, the China Study cohort consists of 75,008 workers in 672 factories from 83 different types of industries in 12 cities.[11] In sharp contrast to the Pliofilm cohort, the workers in the NCI study undoubtedly have a wide variety of exposures other than benzene and these additional exposures need to be evaluated. For example, in an earlier study which included a majority of these 75,088 workers, Yin (1987)[12] mentions a glue mixture of chlorobutadiene and benzene 1:3. While this was only mentioned for one specific factory, many of the operations (e.g., insecticide manufacturing) surely involve potentially confounding exposures. A short list of chemicals that may act as confounders in these operations include: arsenic, styrene, butadiene, ethylene oxide, polycyclic aromatic hydrocarbons, and various agricultural pesticides. Since the investigators did not systematically collect, assess or analyze co-exposures, other potential workplace effects are not accounted for in the benzene risk estimates. To the extent that these exposures occur, [they] are not accounted for in the benzene risk estimates. To the extent that these exposures occur simultaneously, it is more likely that risk estimates per unit benzene exposure would be overestimated. Thus benzene-associated risk estimates must be interpreted with extreme caution."

Response 6. With respect to whether the quantitative estimations are of sufficient validity to use in risk assessment, OEHHA disagrees. In addition to the responses above, NCI and collaborators published an indirect exposure validation study (Dosemeci et al., 1996). Of 412 documented cases of having benzene poisoning among the Chinese benzene-exposed workers, a strong correlation was observed between risk of poisoning and intensity of recent exposure. A strong correlation was also observed with long-term cumulative exposure to benzene (Dosemeci et al., 1996). Ott (1999) criticized this validation study, stating that it only confirmed relative exposure assignments, but not absolute exposure assignments.

However, OEHHA believes that the results also provide some validation to the actual (absolute) ranges of exposure. Numerous researchers in 12 cities conducted ranking of exposures in the study. Without substantial agreement on what the actual levels of exposure were, such a strong association would not have been observed. It is less likely, as Ott (1999) proposes, that a dose-response relationship was found but was due to incorrect assignments based on a similar degree and direction of error in the 12 cities.

OEHHA agrees that there is some concern regarding the impact of co-exposures; however, we feel that that concern does not negate the findings of Hayes et al. (1997). There are several reasons to believe that the effect of potential co-exposures is small. For example, the large majority of the Chinese Worker Cohort (about 75 percent) were painters. Hayes et al. (1997) found a clear excess of leukemia among the painters (i.e., using benzene-containing paint). IARC (1989) reviewed the epidemiological evidence and found no increased leukemia risk among painters not exposed to benzene. Moreover, the excess relative risks for leukemia across different industries (e.g., painters, rubber workers, chemical workers, shoe manufacturers) was quite constant (RR = 1.3 to 2.5), with one possible exception being chemical workers (RR = 3.6). The chemical workers made up a small proportion of the worker cohort. Thus, if any of the excess leukemia cases among chemical workers were related to co-exposures then the overall impact would be small. One should also recall that the only clearly established occupational causes of leukemia are ionizing radiation and benzene exposure, although there is some evidence for associations of leukemia and 1,3-butadiene and styrene.

The commenter should also consider that high co-exposures to chemicals like toluene may in fact reduce leukemia risk, since these agents compete with benzene and its metabolites for cytochrome P4502E1. Thus, at high doses, where competitive and saturable metabolism is occurring, these co-exposures are

likely mitigating the bone marrow damage of benzene. Urinary metabolite data from Rothman et al. (1998) suggest that competitive metabolism of benzene and its metabolites is occurring in the range of 30 to 60 ppm benzene in humans. Thus, it is a distinct possibility that the supralinear dose response curve for leukemia among the Chinese workers may be due in part to co-exposures.

Comment 7. "This [co-exposures] is a particularly important issue because other studies have reported that LH cancer other than acute myelocytic leukemia (AML) may be associated with mixed solvent environments.[13] In this study, the highest risks for LH cancers were with solvents *other than benzene*. In the NCI study, the magnitude of some of the exposure estimates for benzene indicates that occupational hygiene practices were poor. This observation cannot be limited to hygiene practices for benzene exposure, but would apply to the other chemicals mentioned above, as well. Thus, the potential for high exposure to other substances needs to be taken into account in the NCI studies. Under current circumstances, it is misleading to label the NCI/CAPM cohort as only a "benzene" cohort."

Response 7. In the section "Human cancers associated with benzene" we stated: "However, only one epidemiological study reported statistically significant increases in relative risk of lymphoma with benzene exposure. Hayes et al. (1997) reported that benzene-exposed workers in China had significantly evaluated relative risks of non-Hodgkin's lymphoma compared to unexposed workers. However, the increased risk was not consistent among different industries, an observation that reduces the strength of this association."

Also, see responses to Comment 6.

Comment 8. "Exposure Assessment Procedures: Understandably, assessments were difficult to perform and document under field conditions in China, and the need for 12 different field centers is cumbersome. Standardized procedures and training were reported to be in place, but it is no documentation of consistency across field centers. Several possible procedural shortcomings may have lead to overestimates of risk. For example, it is not clear whether field staff were blinded with regard to leukemia or other benzene effects among workers whose jobs and sites they were assessing, and there might have been a tendency for larger estimates in places where benzene health effects were known. A more broad concern is the potential for widespread underestimates of benzene exposure, which would lead to overestimates of effects. Underestimated benzene exposures could result from factors such as undocumented dermal exposures and greater reliance on large samples versus personal monitoring. The report (Dosemeci et al., 1994)[11] gives little attention to the quality and applicability of the monitoring data used to derive exposure estimates. A major concern is that "virtually all of the benzene measurements were based on short-term area sampling." The effective use of area samples requires knowledge of, at a minimum, the strategy for the short-term area collection and several workplace factors (e.g., worker distance from the area sample; ventilation conditions and others)."

Response 8. In the NCI-CAPM study, NCI and collaborators developed standardized job titles for 60 benzene exposure-specific job-titles in 11 major activity groups. Factory-specific exposure estimates were made for each benzene-exposed job. These were developed for five-year calendar periods, by a factory exposure assessment team consisting of industrial hygienists, safety officers, and other employees, following a predetermined exposure assessment procedure (Yin et al., 1994). The researchers utilized 8477 air measurements of benzene along with many other pieces of information to estimate exposure, including information on benzene use, percentage of benzene in benzene-containing materials, and changes in engineering controls. Each factory-specific job title was assigned a benzene-exposure level in one of six concentration ranges (<1 ppm; 1-5 ppm; 6-10 ppm; 11-25 ppm; 26-50 ppm; and >50 ppm) for each of the following exposure periods: 1949-59, 1960-64, 1965-69, 1970-74, 1975-79, 1980-84, 1985+.

The exposure assignments were reviewed to evaluate consistency between the estimated level and the sources of exposure information.

OEHHA agrees that the benzene exposure estimates relied in part on short-term "grab" samples, which are less reliable than personal monitoring. However, one should recall that for the Pliofilm cohort from 1936 to 1968 only "grab" samples were available. It was only until the early 1970s that 130 four- and eight-hour measurements of benzene exposure were made (Paustenbach et al., 1992). The availability of personal monitoring data of the Pliofilm workers, even though it is only for later years, does increase confidence in the exposure assignments.

OEHHA agrees that the reporting of the exposure estimation procedures in the China study (Dosemeci et al., 1994) was not as extensive as for the Pliofilm study. NCI is preparing a manuscript to help clarify some of the questions and criticisms of the NCI-CAPM study that have been made in the literature (Budinsky et al., 1999; Wong, 1999) (personal communication, Dr. Hayes, February 2000).

Comment 9. "Dosemeci et al (1994)[11] applied several adjustments to arrive at exposure estimates. ... However, the adjustments in this study need to be scrutinized for possible, misestimation of exposures. For example, the adjustment for exposure frequency appears to be a very severe adjustment, which on average would have reduced an eight-hour exposure reading to nearly half. This is due to the fact that the report mentions an average exposure duration of 4.2 hours per day. The assumptions about the remaining 3.8 hours per day are also important, and it appears that this assumption was that employees were totally unexposed during this time. This is doubtful, given some of the high short-term readings documented in Dosemeci et al.[11] Also, the method for adjusting exposure estimates for other historical changes is not described in sufficient detail by the authors. It appears that if monitoring data were absent or "inconsistent with "other" (undescribed) exposure information, then monitoring data from other calendar periods were used. If high readings were judged to be inconsistent with "other" data more often than low readings were, this could result in an underestimate of exposure and an overestimate of risk per unit benzene exposure."

Response 9. The above interpretations of the Dosemeci et al. (1994) and Hayes et al. (1997) papers are incorrect. Individual exposure assignments were made for each job-factory-calendar period-specific job title, as described in responses to comment 8. These were compiled and the average exposures, representing a wide range of factories and industries, were reported in Dosemeci et al. (1994). In the Hayes et al. (1997) paper, the individual exposure assignments for each worker, not the averaged measure, were used for the dose-response assessment. This same misinterpretation was made repeatedly in the published criticisms by Budinsky et al. (1999) and Ott (1999). OEHHA agrees that some of the details of the exposure assessment for the Chinese study were not completely described in the original study report (Dosemeci et al., 1994). NCI and collaborators are preparing a manuscript for publication to address these and similar concerns. However, a lack of reporting of all of the exposure assessment details should not be equated with "misestimation."

Comment 10. "OEHHA states on page 92 that an advantage of the China Study is that "exposures remained relatively constant for a large portion of the cohort." It can be argued that the Dosemeci exposure estimates are surprisingly equivalent across industries and era; this is a counter-intuitive time trend and suggests problems with the exposure estimates."

Response 10. The commenter appears to have read too much into the word "constant." Hayes et al. (1997) developed an exposure measure, where follow-up was censored 1.5 years after the individual's exposure level changed out of one of three broad categories (i.e., <10, 10-24, and ≥25 ppm benzene) for the first time. For clarification, OEHHA has added the ranges of the benzene exposure categories to the description of the Hayes et al. (1997) study.

Comment 11. "The methods used to estimate exposure in this study appear to be designed to evaluate a relative dose response relationship, rather than an absolute relationship. In fact, the only validation of exposure estimates done for this cohort only validated the relative ranking of exposure categories.[15] Exposure assessments done in this way are certainly less accurate and the data are less valid for use in the dose step of a risk assessment. For example, the final benzene exposure estimates were assigned one of six categories (<1 ppm, 1-5 ppm, 6-10 ppm, 11-25 ppm, 26-50 ppm, and >50 ppm). Yet despite this assignment, the exposure data are summarized as a continuous variable in Tables II, III, and IV and analyzed as continuous variables by Hayes et al., 1997.[5] There is no description of how an open-ended category (e.g., <1 ppm, >50 ppm) would be assigned a single number. For the other four categories, presumably a midpoint value was assigned. If so, there is no discussion of the possible impact of assigning the midpoint of the range to exposures estimates. Depending on how the data were arrayed, the midpoint assignment could produce either under- or overestimates of risk per unit benzene exposure. In summary, a marked preponderance of the issues identified above would likely overestimate the risk, or underestimate the exposure due to benzene. Thus, it is of questionable validity to use the exposure estimates at face value in risk assessment exercises."

Response 11. Each benzene-exposed worker was categorized annually into one of five benzene exposure levels (< 1 ppm [category midpoint: 0.05 ppm], 1- 5 ppm [category midpoint: 2.2 ppm], 6-10 ppm [category midpoint: 7.0 ppm], 11-25 ppm [category midpoint: 15.0 ppm], 26-50 ppm [category midpoint: 32.6 ppm], and > 50 ppm [category midpoint: 108.8 ppm]) (personal communication, Dr. Hayes, February, 2000). The category midpoints were used to accumulate exposures over time. Accumulated exposure estimates were used to categorize each individual as summarized in the dose-response assessment (Hayes et al., 1997). Dosemeci et al. (1994) reported mean exposure estimates compiled over all factories and cities. Thus, the high exposures observed in some individual factory exposures are not apparent when viewing the overall summary measures (Dosemeci et al., 1994). Since there is expected to be a distribution of exposures representing a large number of individuals within each category, the approach seems reasonable and not likely to represent a large source of error in the exposure assignments. Thus, the exposure methods employed in the Chinese study are less accurate for each individual compared to the Pliofilm cohort, but on a grouped basis are appropriate.

Comment 12. *There is a possible bias in cohort selection.* "If health-related sources are used to a greater extent to identify study (versus comparison) subjects, the bias will be in the direction of overestimating benzene-associated health risks."

Response 12. OEHHA agrees that the reporting of the cohort selection is somewhat vague. NCI and collaborators are preparing a manuscript for publication to address these concerns, which were also raised by others (Budinsky et al., 1999; Ott 1999).

Comment 13. "It should be remembered that although the China study cohort is extremely large, the critical element is the number of LH cancer cases among the study and control groups. With 71 hematologic neoplasms (Hayes et al., 1997)[5] including quite small numbers of cases in the comparison group, the number of cases in subgroup analyses are not high and many results are statistically unstable. Therefore, the seemingly larger sample in the China study is misleading and should not be thought of as a significant justification for using these data."

Response 13. Approximately 3 percent of the Chinese Worker Cohort had died and are reflected in the cases reported by Yin et al. (1996) and Hayes et al. (1997). For the Pliofilm Cohort approximately 28 percent of the workers have died and are reflected in the cases reported by Paxton et al. (1994). The

low rate of mortality in the Chinese study reflects the relatively young age of this cohort. The low number of cases in the control group does create some concern. However, the unexposed control group represents 405,000 person-years at risk and the exposed workers represent 698,000 person-years at risk (Hayes et al., 1997). These numbers should be compared to the Pliofilm cohort, which represents a total of 40,392 person-years (Table 19).

Comment 14. "Data from the China study should undergo the same scrutiny and testing that other studies have done. This process, which takes a number of years, is the only way to increase confidence in decisions about if and how to use the data in quantitative risk assessments. The exposure data in particular are in need of more extensive examination, in areas such as: 1) analysis of other exposures; 2) exposure "confidence" ranking; 3) subdividing and closer examination of the lower exposed group (e.g., <1 ppm); 4) effect of field center; 5) effect of categorization; 6) effect of geometric means; 7) effect of assigning zero exposure intensities; and 8) comparisons with personal monitoring data."

Response 14. The investigations of this cohort of benzene exposed workers in China have been a source of numerous publications since 1987 and continue to provide new insights into the carcinogenic effects of benzene. It is true that the dose-response assessment was only recently published (Hayes et al., 1997) and there is currently an ongoing debate regarding the details of the study (Budinsky et al., 1999; Ott 1999; Hayes et al. (in preparation)). OEHHA believes Hayes et al. (1997) to be a valid study suitable for risk assessment and has derived cancer potency estimates from it.

Comment 15. "On pages 90-92 OEHHA "examines the scientific evidence with respect to the question of benzene-induced low dose linearity". However, this review is a selective review of the literature that supports linearity and a critical review of pieces of the literature that do not support linearity. In some cases, there are factual errors in the write-up.

"For example, in the fourth paragraph on page 90, OEHHA correctly points out that few genotoxicity studies have carried out dose response analyses, especially in humans. They state that an increased frequency of DNA strand breaks and chromosomal aberrations have been found in workers exposed in the low ppm range. However, OEHHA does not point out that other studies done in the low ppm range have failed to detect genotoxicity (e.g., 17; 18; 19). OEHHA then describes the results of two recent studies done on a subgroup of workers from China that examined chromosomal effects and states that these studies did not find a threshold, but could not "extend the dose response range to exposure levels lower than have been associated with leukemia." However, the authors do not point out another study done on the glycophorin A locus, also performed on a China cohort subgroup that did show a threshold for exposure [20]."

Response 15. The original draft of the section entitled "Low dose linearity" was organized to discuss the evidence as it related to the criteria laid out in the carcinogen guidelines for choosing between linear and non-linear approach. In response to the reviewer's comments, the description of the studies has been modified to read: "Increases in the frequency of DNA strand breaks and chromosomal aberrations have been found in groups of workers exposed to concentrations of benzene in the low ppm range (TWA), while others have failed to find an association."

With respect to the genotoxicity results from the glycophorin A assay, Exxon is incorrect in its portrayal of the data. Exxon cited an early abstract (reference 20) which was later fully reported in Rothman et al. (1995). As discussed in the section of the PHG on Genetic Toxicity in humans: "A strong positive trend over all dose groups (p=0.005) was observed between cumulative exposure and the frequency of NN mutant variants. No correlation was observed for the N∅ mutants."

Comment 16. "Benzene has not been shown to be a point mutagen. As delineated in the paper by Dr. Julian Preston in Appendix A, among the various types of genotoxicity, only point mutations can result from a single biochemical event. As such, a "one-hit" linear-at-low-dose response curve could be justified for a strong point mutagen. However, this is not applicable to benzene's genotoxic mode of action. Benzene has only been associated with genotoxic events that would be governed by multiple "hits", hence implying low dose sub-linearity."

Response 16. OEHHA believes that the assumption that clastogenicity equates to non-linearity is not consistent with available scientific evidence and current thinking about how carcinogenic mechanisms of action in general relate to tumor dose-response relationships (Gaylor, 1992; Hoel and Portier, 1994; Elder and Kopp-Schneider, 1998; Lutz, 1998). The cited paper (Appendix A of WSPA's submission) includes a discussion of radiation-induced leukemia as a model of an agent that induces primarily clastogenic effects and demonstrates a non-linear tumor response, referring to the unpublished work of Mendelsohn (1996) to support the notion that "the relative risk for radiation-induced leukemias is nonlinear as a function of radiation dose, and that exactly the same nonlinear shaped curve and relative risk values are obtained with chromosomal aberrations as a function of dose." However, this conclusion is not universally held. A number of highly authoritative groups have published evaluations of radiation-induced leukemia and concluded the contrary. For example, the National Academy of Sciences Committees on Biological Effects of Ionizing Radiation (BEIR) IV and BEIR V (NRC, 1990) as well as the researchers from the Life Span Study who have followed the atomic bomb survivors in Japan (Preston et al., 1994) all utilized linear-to-low dose risk models in assessing the leukemia risks from radiation exposures. Also, OEHHA would like to point out that the excess relative risk dose-response curves for most data sets of radiation-induced leukemia appear linear with low doses and supralinear with increasing radiation doses (NRC, 1990).

Comment 17. "DNA adducts may produce point mutations, and some have argued that low doses of benzene can cause DNA adducts. However, findings using the DNA-specific ^{32}P-postlabeling technique has not shown the formation of DNA adducts, except at levels far exceeding those of environmental concern. Preliminary results that used the very sensitive accelerator mass spectrometry (AMS) technique suggested that benzene may produce these lesions at low concentrations.[21] However, these results are likely due to protein contamination. In Appendix B, Dr. Vijay Reddy (a former colleague of Dr. Kurt Randerath, who developed the ^{32}P-postlabeling technique) supports this statement. In addition, Dr. Reddy points out that the lack of correlation between DNA binding levels and tumor susceptibility for various tissues in rodents suggests that these adducts are not involved in any subsequent carcinogenic response."

Response 17. OEHHA has carefully reviewed Appendix B of WSPA's submission. The purpose of the document is not directly stated, but it attempts to refute the report from scientists at Lawrence Livermore National Laboratory that administration of benzene to rodents resulted in linear binding to DNA over a wide range of doses, including very low doses (Creek et al., 1997). We feel that the submission fails to refute the published findings, but does highlight the need for additional research. With regard to protein contamination, Creek et al. (1997) as well as other researchers cited in Appendix B (Arfellini et al., 1985; Mazzullo et al., 1989) utilized standard, accepted techniques for separation of DNA, RNA and protein. Moreover, because the binding to the various sites will vary proportionally with dose, even if a portion of the binding measured by AMS and standard radiobinding techniques were due to protein binding, it would still hold that dose-dependent DNA binding would occur to low doses.

A more plausible explanation of the discrepancy between the AMS and ^{32}P-postlabeling results lies in the limitations of the ^{32}P-postlabeling assay. The chromatography used for ^{32}P-postlabeling separates a narrow range of suspected adducts, primarily benzoquinone-DNA adducts in Reddy's studies. Important adductive species such as benzene oxide (Norpoth et al., 1988; Lindstrom et al., 1997), quinol thioethers (Bratton et al., 1997), or other reactive metabolites, may not be detected. Indeed, recent studies suggest

that high molecular weight DNA adducts of hydroquinone and 1,2,4-benzenetriol may be critical (Wiemels et al., 1999). The high levels of phenol and hydroquinone in all tissues due to background sources, unrelated to benzene, make it difficult to measure the phenol and hydroquinone-derived DNA adducts above the high background, although some have done so (Levay et al., 1996).

Appendix B also draws attention to a "lack of correlation between binding levels and tumor susceptibility for various tissues [that] raises the question of appropriateness of DNA binding data for use in risk assessment." This was illustrated in a table abstracted from data in several studies. Comparative data abstracted from this table for male Wistar rats and male BALB/c mice administered a single intra peritoneal (i.p.) injection of benzene are shown below.

DNA binding (pmol/mg protein) from a single 49 or 50 mg/kg i.p. dose of benzene

	marrow	liver	stomach	spleen	kidney	lung
Arfellini et al. (rats)	0.70	0.14		0.65	0.56	0.14
Mazzullo et al. (rats)		1.16	3.32	2.76	2.57	1.16
Arfellini et al. (mice)	0.36	0.30		0.63	0.27	0.04

These data cannot be used to infer tumor susceptibility for various tissues based on macromolecular binding. First, the studies are from a single dose, not chronic dosing. Secondly, there are no carcinogenicity studies of benzene by the i.p. route to compare tumor tissue susceptibility. By the oral route, male Wistar rats exhibited tumors of the Zymbal gland and oral cavity only (Maltoni et al., 1989). No other cancer study of male Wistar rats has been reported. Thus, technically all the tissues assayed by Arfellini et al. and Mazzullo et al. are non-responsive sites. No cancer study of male BALB/c mice has been conducted, although oral cancer studies of other strains of male mice have shown increased responses in Zymbal gland, preputial gland, Harderian gland, lymphoma, and lung. Also, OEHHA notes as much as eight-fold difference across common tissues studied in the two binding studies in rats, which further confounds interpretations.

Comment 18. "OEHHA cites three studies [22-24] as showing positive responses in the *lacI* gene from the "Big Blue" transgenic mouse model. It is assumed that these studies provide support for a one-hit mutation mechanism. However, these observations can be a result of point mutations or small deletions. Further, there are technical concerns…. Thus, these studies do not provide strong support for a mutational mechanism."

Response 18. There are many differences of opinion about mutational mechanisms, as noted in the response to Comment 16, and as discussed in the PHG document.

Comment 19. "Since both point mutations and adducts can be all dismissed as having a role in benzene-induced AML, chromosomal loss, translocations and aberrations are the most likely genotoxic lesions that may be involved in benzene-induced AML. Thus, the types of mutations that are the most plausible are the types that support a multiple hit hypothesis and this argues against linearity."

Response 19. OEHHA disagrees. OEHHA believes that the assumption that clastogenicity equates to non-linearity is not consistent in all cases with available scientific evidence and current thinking about how carcinogenic mechanisms of action in general relate to tumor dose-response relationships (Gaylor, 1992; Hoel and Portier, 1994; Elder and Kopp-Schneider, 1998; Lutz, 1998).

Comment 20. "On the top of page 90, OEHHA reviews dose response curves for benzene metabolism and metabolic formation. Even though benzene metabolism is clearly necessary for benzene-induced leukemogenesis, the shape of metabolic curves is practically irrelevant to the shape of a leukemia dose response curve. Thus, this exercise is puzzling at best, as there are undoubtedly multiple steps between metabolism and leukemogenesis. One could make a convincing argument that a saturable metabolism dose curve that displays strong supra-linearity is consistent with a sub-linear leukemogenesis dose response curve, especially if metabolic saturation is a key event that triggers compensatory mechanisms or alternate metabolic pathways."

Response 20. OEHHA agrees that metabolism is many steps away from leukemia. However, the linear formation of toxic metabolites following exposure to benzene in vivo is highly germane to a discussion of benzene toxicity. Indeed, Section 4 of Appendix C of WSPA's own submission entitled "Low-dose non-linearity of biological mechanisms" discusses at length how non-linear metabolism relates to low-dose cancer risk.

Even if there are multiple steps in the process, if the rate-limiting step is linear with dose then the results can appear linear with dose. Linear formation of reactive metabolites of benzene is relevant to observations of linear formation of macromolecular adduction in the bone marrow of rats and mice (McDonald et al., 1994) and in various tissues over a wide range of benzene doses (Creek et al., 1997). Also, in studies of human hematopoietic progenitor cells in culture, it has been found that extremely low concentrations of hydroquinone, likely relevant to environmental exposures, are able to disrupt developmental response to growth factors (Irons and Stillman, 1996). The reported dose-response curve appeared to be roughly linear with the log of dose, at low doses. On a linear scale, that relation would be strongly supralinear.

Comment 21. "In the fourth paragraph on page 91, it is stated that "in most epidemiologic studies examining dose response relationships, the data appear linear to low doses of benzene. For example, visual inspection of the Pliofilm and Chinese cohort suggest a linear relationship at low doses (Figures 3, 5, and 6)." However, the dose response curve for AMML in the Pliofilm study is highly non-linear. For mean ppm-yr exposures of 11, 151, 602, and 1341, relative risks of 0, 2, 9, and 83 are found[3]. This is based on leukemias of known cell type. The non-linear nature of the Pliofilm study is reinforced in Rinsky, 1987[4] and Schnatter, 1996b[7]."

Response 21. OEHHA disagrees. The data presented (from Crump, 1994) are based on the Paustenbach exposure estimates, which OEHHA believes are likely to be high (Utterback and Rinsky, 1995; and see responses to comments below on WSPA's April 2000 submission). Also, as stated by Crump (1994), two cases in the low-dose group were undefined as to their leukemia subtype. Thus if these two cases were included in the analysis, which would make them estimates for ANLL, then the relative risks for Crump's exposure category cut points are 2.4, 2.0, 9.1 and 82.8, respectively (Crump, 1994). The cut points selected provided a distribution of cases that were 2, 1, 2 and 5 cases, respectively for each of the exposure categories (Crump, 1994). The selection of the cut points of the exposure categories can have significant effect on the risk ratios (Selvin, 1996). For comparison, Table 19 of the PHG shows the results using different cut points for the Paustenbach exposure estimates that provide a more uniform distribution of the cases (total leukemia) across the categories. For mean cumulative exposures of 16.7, 108.9, 290.9 and 745.3 ppm-years, the relative risks of total leukemia were 2.14, 2.45, 5.73, and 10.41, respectively (Table 19, Figure 3), providing a strong linear fit (Table 20).

Comment 22. "OEHHA gives only marginal credence to studies that have examined low benzene exposures in petroleum workers.[26-28] In these studies, there is essentially no exposure response relationships observed, despite the examination of multiple exposure metrics. Since OEHHA discarded

high exposure data in the Pliofilm and China cohorts, they indicate an interest in lower exposures. The above-referenced petroleum worker studies should have been used as consistency checks, as described in Hertz-Piccioto, 1995.[29] This is consistent with OEHHA's focus on lower exposures."

Response 22. OEHHA agrees that it is important to utilize available data to provide a consistency check, especially data on cohorts exposed to low levels of benzene. This is the reason OEHHA compared the potency estimates from the Pliofilm and Chinese Worker Cohorts to potency estimates for leukemia from cigarette smoking. (See the PHG section "Comparison of benzene-induced leukemia risk to smoking-induced leukemia risk"). Cigarette smoking is the highest non-occupational exposure to benzene (Wallace, 1996). OEHHA compared the leukemia risk among two studies, each following over one million smokers, to the risk of leukemia from benzene exposure (i.e., based on the potency estimate derived in the PHG). The potency estimate from the benzene-exposed workers was only 32 percent of the potency estimate for leukemia from cigarette smoking, which is reasonable since cigarette smoke contains trace amounts of other potential leukemogens. As stated in the PHG, "Thus, we conclude that the leukemia lifetime risk estimates developed from the benzene-exposed workers are consistent with the smoking data and do not appear to overpredict benzene-induced risk."

As presented in the original draft PHG document, OEHHA developed, for comparison purposes, potency estimates for seven other benzene-exposed cohorts (see Table 28 of the PHG). We concluded that "Generally, the studies with very high exposures (Aksoy, 1994; Fu et al., 1996) provided the lowest potency which would be expected from linear extrapolation from the high end of a supralinear curve. Likewise, the studies providing the lowest exposure levels (Ott et al., 1978; Wong, 1987; Jakobsson et al., 1993; Ireland et al, 1997; Lynge et al., 1997) provided the highest potency estimates which is consistent with linear extrapolation from the low end of a supralinear curve." The cancer potency estimates from studies examining lower exposures were consistent with the potency estimates derived from the Pliofilm Cohort and Chinese Worker Cohort.

Exxon cites three studies as showing "essentially no exposure response relationships": a nested case-control study among Canadian Petroleum Distribution workers (Schnatter et al., 1996), a "meta-analysis" of a combined cohort consisting of more than 208,000 petroleum workers in the United States and the United Kingdom (Raabe and Wong, 1996), and a case-control study of petroleum marketing and distribution workers in the United Kingdom (Rushton and Romaniuk, 1997). These three studies are discussed in the PHG document.

With respect to the Rushton and Romaniuk (1997), OEHHA believes that the results from this study are consistent with the findings in the Pliofilm and Chinese Worker Cohorts. Although not statistically significant, the Rushton and Romaniuk (1997) study reported increased relative risks (RR = two to three) of total leukemia or acute myeloid and acute monocytic leukemia combined (AMML) with cumulative exposures <45 ppm-yr compared to <0.45 ppm-yr. In both the case-control studies (Schnatter et al., 1996; Rushton and Romaniuk, 1997), it was not clear how the cut points were selected for the exposure categories. The selection of the cut points of the exposure categories can have significant effect on the risk ratios (Selvin, 1996). For example, in one presentation of the data Rushton and Romaniuk use quartiles with cut points of <0.45, 0.45-4.49, 4.5-44.9 and ≥ 45 ppm-yr which corresponded to 22 cases (OR=1.0), 47 cases (OR=1.42), 20 cases (OR=2.48) and 1 case (OR=1.35), respectively. It would be interesting to see what the dose-response relationship would be if the exposure categories were selected to parse out the cases evenly, as is standard practice.

Although the Raabe and Wong (1996) meta-analysis represents a very large number of workers, it has the drawback that no exposure or dose-response estimates are available. Thus, one cannot assess potential healthy-worker effects, which have been substantial in some benzene-exposed cohorts (Wong, 1987a) but not in others (Crump and Allen, 1984).

In response to these comments, we added a paragraph in the risk characterization section stating that some researchers believe that the true dose-response relationship for benzene-induced leukemias is non-linear at

low doses or exhibits a threshold of response and that this was considered in OEHHA's overall assessment of the available information.

Comment 23. "In the final paragraph on page 91, OEHHA points out that benzene-induced leukemias may derive from preceding syndromes characterized by hematotoxicity and bone marrow suppression. This would support a non-linear or threshold model. OEHHA then states: "there is significant evidence to indicate that such frank effects are not a required step in benzene-induced cancer". Quite surprisingly, there is no reference for this statement. In fact, this is still an open research question, evidenced by the fact that there has not been a single case of clear benzene-induced leukemia that has undergone sufficient clinical observation before disease onset such that frank hematotoxicity and bone marrow suppression could be ruled out."

Response 23. The offending sentence has been changed to read "However, extensive evidence of from epidemiological studies of leukemia arising in groups of individuals with relatively low exposures to benzene (see Carcinogenicity, page 38) would argue against the notion that a high-dose toxicity preleukemic state is a required step in leukemogenesis."

Comment 24. *The section addressing "comparison of approaches" (p. 89) contains a factual error in stating that "Investigations by Crump (1994)[3] indicated that linear models fit the Pliofilm data better than non-linear ones."* "In fact, the relevant statement from the article 3 is: 'Whereas dose responses were essentially linear when the Crump and Allen (1984)[30] exposure matrix was used, there was evidence of intensity-dependent non-linearity in dose responses using the Paustenbach[31] exposure matrix'. Also, Crump (1994)[3] himself states 'the preferred estimates from this analysis are those based on the Paustenbach et al. Exposure matrix'. This would indicate that the analyses that should be used from this report are the non-linear models based on the Paustenbach exposure matrix."

Response 24. OEHHA believes the Paustenbach estimates are likely to be high. The Crump (1994) analysis indicated that non-linear models provided only a slightly better fit than non-linear models when using the Paustenbach exposure estimates, and when using the Crump and Allen (1984) exposure estimates, linear models fit much better than non-linear models. However, in response to this comment, the sentence has been changed to read "Investigations by Crump (1994) indicated that linear models fit the Pliofilm data, using the Crump and Allen (1984) exposure estimates, better than non-linear ones." Additional information, including other epidemiological studies, metabolism, and mechanism of action, also were evaluated with respect the selection of linear models as appropriate. In this respect, US EPA (1998) came to the same conclusion.

Comment 25. "In summary, in considering whether to employ a linear or non-linear dose response curve for benzene leukemogenesis, OEHHA performed a selected review of the literature that supported only linearity. There is ample evidence to move away from a dose response curve that is premised on a mechanism that does not apply to benzene. Furthermore, a non-linear curve is supported by empirical epidemiologic studies of Pliofilm as well as petroleum workers."

Response 25. OEHHA disagrees; we evaluated the available evidence pertaining to low dose linearity and applied that evidence to criteria in the carcinogen guidelines for selecting among linear and non-linear approaches. Sufficient evidence does not exist to move away from a linear approach. In this regard, U.S. EPA agrees (U.S. EPA, 1998). Moreover, a significant amount of information suggests that benzene induces leukemia in a dose-response relationship that is linear to low doses.

Comment 26. "OEHHA derives potency estimates for total leukemia, rather than ANLL. The latter is clearly the only leukemic cell type for which a causal relationship is scientifically plausible. In attempting to justify this choice (pp. 61-65), OEHHA does not make a convincing or valid argument that total leukemia is the appropriate health effect. Again, rather than objectively summarizing the literature in this regard, OEHHA provides selective citations that present only one side of the debate. While OEHHA appropriately eliminates lymphomas and multiple myeloma from consideration, there is now also a good rationale for using ANLL, rather than total leukemias, for benzene risk assessments."

Response 26. OEHHA disagrees. Although ANLL is the predominant subtype of leukemia associated with benzene exposure in humans, there is strong evidence to associate benzene with other leukemia subtypes. The rationale for this assertion was discussed in the draft PHG document in the section entitled "Human Cancers Associated with Benzene Exposure." To clarify the impact of this decision, OEHHA added the following sentence to the Risk Characterization section. "The impact of this decision on the magnitude of the risk estimates is minimal, since cancer risk estimates based on ANLL/MDS (Table 27, Chinese Worker Cohort) or AMML (Table 18, Pliofilm Cohort) (Crump, 1994) differed by about 20 to 25 percent from those based on total leukemia."

Comment 27. **"**Appendix C entitled "Biologic Mechanistic Considerations Relevant to Benzene-Induced Leukemia" thoroughly examines the question of benzene and different leukemia cell-types. Drawing upon several lines of research, it becomes clear that there is insufficient evidence to regard leukemic cell types other than ANLL as causally related to benzene."

Response 27. OEHHA reviewed the evidence presented in Appendix C of WSPA's submission, and found no compelling additional data to support the above statement.

Comment 28. "Myeloid progenitor cells, which give rise to granulocytes, have the enzymatic distribution that makes them susceptible to benzene. Myeloid progenitors have high levels of myeloperoxidase, an activating enzyme and low levels of quinone reductase, a deactivating enzyme.[32] Earlier stem cells do not have this enzyme distribution.[33] Thus the myeloid progenitor is the likely target for benzene, thereby making ANLL the relevant cell type.

"The myeloid progenitor cell is also a target for direct proliferative and cytotoxic effects.[34] In addition, benzene metabolites induce a transient recruitment of early myeloid progenitor cells into active cycling and proliferation.[35] The inappropriate recruitment phenomenon makes these early myeloid progenitor cells more susceptible to genotoxic events either due to natural replication errors or secondary to benzene exposure."

Response 28. As noted in the draft, benzene exposure among worker populations is most commonly associated with ANLL. These observations noted in the comment may explain, in part, why acute myelogenous leukemia is the predominant leukemia subtype observed in these studies. However, as discussed in the section entitled "Human Cancers Associated with Exposure to Benzene," benzene can affect cells from both myeloid and lymphoid cell lineages and has been also associated in some studies with increased relative risks of non-ANLL leukemias (Crump, 1994; Hayes et al., 1997) and non-Hodgkin's lymphoma (Hayes et al., 1997).

Comment 29. "On page 63, in the second full paragraph, OEHHA briefly notes that different growth factors and cytokines are required for blood cell maturation within specific lineages. They also note that AML and CML arise from different lineages, and that these facts might affect the dose response for (and also the capacity to cause) different leukemias. … OEHHA seems to dismiss this line of evidence based

on the fact that ALL, AML, and CML share some genetic alterations (page 63). However, this is not relevant since it pertains to cancers with an unknown etiology, not benzene-induced leukemias."

Response 29. The sentence "However, shared genetic alterations among ALL, AML and CML have been known for some time, suggesting DNA damage to early stem cells (Bloomfield et al., 1978; Yunis, 1983; Kurzrock et al., 1988)" has been removed.

Comment 30. "Finally, OEHHA cites the Ireland 1997 study[42] as suggesting that death certificates are a poor source of leukemia cell type information, with an accuracy of less than 75%. However, if this is true, it also pertains to the general population rates. Unless one assumes this error is different in exposed workers versus unexposed workers, it should not affect dose response relationships markedly. In fact, the error would tend to weaken a dose response for a proven relationship (i.e., ANLL), and produce an error in either direction when there is no proven underlying causal association (i.e., for other cell types)."

Response 30. OEHHA agrees that the misclassification would also pertain to background population rates. The relevant sentences have been removed.

Western States Petroleum Association (WSPA) (April 2000)

Comment 1. *This 57-page submission related to a single issue, the use of the Paustenbach et al. (1992) exposure estimates of the Pliofilm Cohort, in the following context:* "…WSPA is concerned that OEHHA has apparently chosen to not consider the arguments based on the Paustenbach et al., exposure estimates. WSPA realizes that OEHHA staff does not have the luxury of time to delve deeply into individual papers, however, in this case we believe that the issues at stake are of sufficient importance that a more thorough consideration is warranted. The choice of exposure estimates not only determines the quantitative outcome of a potency calculation such as OEHHA is undertaking, but in this case also has a strong influence in the argument surrounding the choice of a linear or non linear risk model. … The primary point of these additional comments are to present an in depth examination of the Utterback and Rinsky critique of the Paustenbach exposure estimates." Three unpublished manuscripts were attached.

Response 1. OEHHA thanks WSPA for providing this information; we were not aware of the existence of these unpublished evaluations. OEHHA has examined carefully the submitted information and has tried to evaluate the polar viewpoints (e.g., Utterback and Rinsky, 1995 versus Attachments A, B and C of WSPA's submission) expressed on the validity of the Paustenbach et al. (1992) exposure estimates. Below, we discuss the major points of contention, and evaluate the arguments for and against the exposure assessment methods and adjustments made by Paustenbach et al. (1992).

The Paustenbach et al. (1992) exposure estimates. Paustenbach et al. (1992) published a reevaluation of the exposure estimates for the Ohio rubber hydrochloride workers (the Pliofilm Cohort), which had been extensively studied and continues to be followed by NIOSH (Rinsky et al., 1981; 1987; Paxton et al., 1994). Prior to the Paustenbach publication, two separate exposure estimates were available: one by the original study authors using standard epidemiological methods (Rinsky et al., 1981, 1987) and one by Crump and Allen (1984) who adjusted the Rinsky estimates by relating them to changes in the TLV over time. The Paustenbach et al. (1992) reevaluation took into account factors such as additional historical records, information gained from worker interviews, dermal exposures, short-term high-level exposures, installation of engineering controls, accuracy of monitoring devices, long work hours during World War II, and respirator use. The updated exposure estimates (Paustenbach et al., 1992) for the majority of the workers were higher than the original study author estimates (Rinsky et al., 1981, 1987) and the Crump and Allen (1984) estimates. The relative magnitude of these differences is borne out in the resulting cancer potency estimates. The potencies based on the Rinsky estimates, including the highly exposed

workers, were four-fold higher than those based on Paustenbach estimates, and the potencies based on the Crump estimates were about two-fold higher than those based on the Paustenbach estimates (see Table 20 of the PHG). In general, however, the three sets of exposure estimates differed primarily in how they estimated exposures in the early years of the cohort where few actual exposure measurements were available. This point is illustrated in OEHHA estimates of the cancer potencies following removal of the most highly exposed workers (>400 ppm-yr), which were the essentially the same using the Rinsky and Crump exposure estimates and two-fold lower using the Paustenbach estimates (Table 20).

Major points of contention related to the validity of the Paustenbach exposure estimates. Utterback and Rinsky (1995) severely criticized the basis of the Paustenbach et al. (1992) exposure estimates, suggesting the estimates were unreasonably high. Paustenbach, Beatty and Confer (Attachments A, B, C of WSPA's submission, respectively) defend the reevaluation as appropriate and accurate.

Point 1. Plausibility of the estimates in relation to the observed incidence of aplastic anemia and other fatal blood disorders.

Utterback and Rinsky: "However, a careful examination of their report clearly shows the annual average benzene exposure concentrations in excess of 250 parts per million (ppm) for one job title and ~150 ppm for several others. Many of these estimated exposures reportedly lasted for years. The predicted benzene exposures for the two most highly exposed categories extended over a decade." "Sustained exposure to benzene at the levels derived by Paustenbach et al. have been known for almost a century to cause severe blood dyscrasia, including pancytopenia and aplastic anemia [LeNoir, 1897; Snyder et al., 1993]. The benzene exposure concentrations predicted by Paustenbach et al. would likely have caused an epidemic of fatal nonmalignant blood disorder in this cohort of workers [WHO, 1993]."

Paustenbach: "Here, and throughout the paper, Utterback and Rinsky seem to misunderstand that we did not claim that airborne concentration of benzene averaged 250 ppm, but rather that the absorbed dose (due to dermal uptake and inhalation during long workshifts and periods of short-term high exposure) was equivalent to inhaling 250 ppm for 8 hrs. As clearly indicated in ... Tables 7 and 8 of Paustenbach et al. (1992), most of our estimates of the airborne benzene concentrations were at or below the prevailing TLVs. The one exception is the quencher."

Evaluation: There are two important questions interlaced here. One, would the estimated exposures (as defined by Paustenbach) cause fatal non-malignant blood diseases? Two, how many fatal non-malignant blood diseases were observed among Pliofilm workers?

With respect to the first question, Utterback and Rinsky are comparing the Paustenbach exposure estimates to early studies of reported deaths from high exposures to benzene. A review of benzene-induced aplastic anemia (Smith, 1996) suggests that inhalation exposures to >100 ppm are associated with an increased incidence of aplastic anemia of about 1/100, which falls to a rate of about 1/10,000 at exposure levels of 10 to 20 ppm. The comparisons of aplastic anemia are usually with airborne concentrations only. However, the historical data is based primarily on observations of shoe and leather workers, where dermal exposures would also likely be present. Paustenbach is correct that most of the average air concentrations they reported were below 100 ppm, with very high peak exposures noted. However, the significance of the impact of the exposure duration is unclear. If the Pliofilm workers worked longer days than the historical cohorts, then there might be an expectation of greater risk of developing aplastic anemia. If we remove the dermal absorption component of Paustenbach's estimates (i.e., reduce them by 20 percent (Paustenbach et al., 1992)), we are left with eight job codes with average air concentration exposures of >100 ppm for at least six years, with one job assigned 8 h-TWA air exposures ranging from roughly 140 to 210 ppm over 29 years (based on Table 12 of Paustenbach et al., 1992). An additional 11 job codes were estimated to have average air exposures between 50 and 100 ppm (dermal component removed).

Utterback and Rinsky claim that only one worker in the Pliofilm Cohort died from aplastic anemia or other fatal blood disorder. Paustenbach claims that there was evidence of significant occurrence of blood disorders, including at least three deaths, in the Pliofilm Cohort. The basis for this difference appears to stem from different interpretations of a 1942 Department of Labor report on the rubber industry and whether the workers referred to in the report were employed at the Pliofilm plant or at other facilities.

Paustenbach also cited studies by Kipen et al. (1989) and Cody et al. (1993) to suggest widespread job-related blood disorders. These two studies compared red and white blood cell levels among Pliofilm workers to exposures based on the Crump and Allen (1984) estimates. The studies suggested depressed white blood cell counts among workers in the 1940s, indicative of high exposures. A subsequent analysis of the data by Ward et al. (1996) noted that pre-employment blood levels during the 1940s were lower than pre-employment levels in later decades, suggesting that early exposures were not as high as predicted by Cody et al. (1993). Analysis by Ward et al. (1996), which controlled for temporal differences in pre-employment blood counts, reported a strong relationship between exposures as estimated by Rinsky et al. (1987).

Thus, OEHHA finds it difficult to support or refute Utterback and Rinsky's assertion that the Paustenbach exposure estimates are not consistent with the observed incidence of aplastic anemia among the Pliofilm cohort. It is likely, however, that the exposure levels estimated by Paustenbach would have resulted in multiple non-malignant benzene-related deaths, although an estimate of the expected number of deaths would take detailed analysis. Whether multiple non-malignant benzene-related deaths occurred among this cohort of workers is unclear to OEHHA. Our decision was that this information could not be used to determine the plausibility of the exposure estimates.

Point 2. When were engineering controls in place?

As stated above, the primary difference among the three sets of exposure estimates (Rinsky et al., 1981; Crump and Allen, 1984; Paustenbach et al., 1992) is how each of the groups estimated exposures in the early years where few actual measurements were made. A critical question pertaining to the early exposures was when ventilation hoods were installed over the processing areas. Paustenbach et al. appear to suggest that engineering controls were added in 1946, whereas Utterback and Rinsky (1995) suggest that available evidence indicates that these controls were in place well before 1942. This point is particularly important in light of Paustenbach et al.'s adjustments for long work hours during the World War II years.

Utterback and Rinsky (1995): "The Department of Labor conference proceedings [1942] are quite informative about the use of local exhaust ventilation in the rubber hydrochloride plants…the plant physician states, 'At present benzol is used in the manufacture of pliofilm but in an enclosed system. Hooded ventilation with negative pressure is installed over all spreaders.' Control of benzene emissions from the spreaders was provided by 'hooded ventilation with suction above and forced general draft ventilation in the room.' Concentrations of benzene in the vicinity of the RH spreader units [was stated] in the proceedings to be controlled to a level 'in the neighborhood of 20 to 60 parts per million'. The proceedings further mention the 'earlier unsuccessful attempts to remove benzol fumes with downdraft ventilation only." "Remarkably the benzene concentration range and the ventilation controls are never mentioned in the Paustenbach et al. article, even though the concentrate range is the only report of airborne benzene concentrations associated with the rubber hydrochloride industry prior to 1946. (The proceedings of this conference were not cited in the previous exposure assessment by Rinsky et al., 1981, and Crump and Allen, 1984)." "Paustenbach et al. also state: 'In November 1946, following cases of benzene toxicity observed at St. Marys, management apparently embarked on a major program to reduce airborne concentrations of benzene.' The source of this information is attributed to Fluker [1946]. The Industrial Commission of Ohio report by Fluker actually states that 'extensive exhaust equipment has recently been installed for the elimination of benzol vapors

generated by the presses' and that 'tests were made with benzol detectors and the results indicate that concentration have been reduced to a safe level and in most instances range from zero to 10 to 15 parts per million.' There is no mention of the 'major program to reduce airborne concentrations of benzene' that Paustenbach et al. claim to have occurred in 1946. Such an effort apparently occurred prior to 1942 according to the Department of Labor conference proceedings." "Paustenbach et al. also state in reference to the 1946 testing of the ventilation system that 'only a single air sample was collected at the filters before engineering controls were installed, and the concentration was 250 ppm. Samples collected several years later after engineering controls showed levels of 19-50 ppm [Rinsky et al., 1981]. The state of Ohio concluded that the filter-press ventilation system was sufficient to maintain benzene concentration below 100 ppm and typically below 35 ppm [Fluker, 1946]."

Paustenbach: Dr. Paustenbach provided no comment on this point.

Evaluation: It appears that Paustenbach et al. (1992) have likely significantly overestimated the air levels of benzene in the early years of the cohort. Since the statements in the Paustenbach analysis regarding the installation of fume controls lack documentation support, we will agree with the conclusion of Utterback and Rinsky (1995) which have substantial documentation support. This overestimation by Paustenbach et al. (1992) would be magnified several times based on their adjustment procedures, such as adjusting for increased work hours during the war years, detector accuracy, and changes in the threshold limit value (TLV).

Point 3. Adjustments for dermal absorption

Paustenbach adjusted the overall exposure estimates by including exposure via dermal absorption. This has not been routinely done for most epidemiological studies of benzene, but is a potentially important source of exposure. Incorporation of dermal exposures would allow for a more accurate assessment of the relationship between dose of benzene and tumor response. Utterback and Rinsky (1995) took issue with the methods used by Paustenbach et al. (1992) to make this adjustment. Utterback and Rinsky (1995) claimed that Paustenbach et al. utilized improper dermal exposure factors, which overestimated the dermal exposures. The two primary points of contention were (1) the magnitude of the dermal transfer factor for benzene and (2) the surface area of the skin. Paustenbach et al. (1992) utilized a dermal transfer factor of 0.4 $mg/cm^2/h$, which was obtained from a study of human volunteers following application of benzene and occluding the skin (Hanke et al., 1961). Utterback and Rinsky contend that this estimate represents a maximal value obtained under extreme conditions. Utterback and Rinsky proposed that a more reasonable value would be 0.05 $mg/cm^2/h$, based on a study in rhesus monkeys (Maibach and Anjo, 1981). In response, Beatty (Attachment B of WSPA's submission) reported that conversion of the units of the dermal absorption, as reported in Maibach and Anjo, 1981, would result in a dermal transfer factor of approximately 0.48 $mg/cm^2/h$. Moreover, Beatty argues that direct contact with the rubber hydrochloride, or through dampened clothes or shoes, would represent occlusive exposures, thus arguing the higher transfer factor is more appropriate.

Paustenbach et al. (1992) employed a skin surface area factor of 1980 cm^2, which represents the surface area of the hands and forearms up to the elbows. Utterback and Rinsky (1995) suggest that a surface area of 600 cm^2, which represents the surface area of the hands, would be more appropriate.

Utterback and Rinsky (1995): "Differences between our estimates and those from Paustenbach et al. are due almost entirely to differences in skin contact area and uptake rates... The resulting estimates are ~3 % of the values given by Paustenbach et al. ... for the same job classifications."

Paustenbach: "In their analysis, Utterback and Rinsky adopted a dermal absorption rate of 0.05 $mg/cm^2/h$, skin surface area of 600 cm^2 and similar contact times per day as those suggested by Paustenbach et al. We don't find these exposure factors to be more appropriate than ours. We have already explained why 0.4 mg/cm^2-hr seems more reasonable than 0.05. We agree that there

is some uncertainty in our estimate of the likely area of dermal contact. However, contact from the hands to elbows is consistent with many, if not most, exposure assessments of persons in industrial settings." Paustenbach concludes "Lastly, even if one were to adopt their suggestions it would not change our prediction of the role of dermal uptake for workers in the cohort by even 10%. As Table 14 of our manuscript shows the average dose contributed by dermal exposure for the cohort is about 15%-20% to the overall dose. Therefore a 10% change in the dermal update would result in only about 2% change in our estimates."

Evaluation: Paustenbach and Beatty provide good arguments why the Paustenbach et al. dermal exposure estimates are plausible. However, it is difficult to know for sure whether contact factors used by Utterback and Rinsky or Paustenbach et al. are more reasonable. OEHHA does not see how Paustenbach came to the conclusion that use of either set of estimates would have changed the overall estimates by only 2 percent. Dermal exposure estimates by Utterback and Rinsky (1995) were about 3 percent of those estimated by Paustenbach. Thus, for a worker exposed to air concentrations of 100 ppm, for example, adjusted estimates would be about 120 ppm for Paustenbach et al. and 101 ppm for Utterback and Rinsky. In other words, adjustments for dermal absorption under the Utterback and Rinsky proposal would increase the overall estimates about one percent, whereas, adjustments under the Paustenbach et al. proposal would increase the overall estimates about 15 to 20 percent.

In either case, the adjustments of the exposure estimates to account for dermal exposures are relatively small in comparison to the other adjustments made for assessing exposures via inhalation.

Point 4. Were the grab bag samples representative of average exposures?

Utterback and Rinsky (1995): "Paustenbach et al. (1992) contend that historical benzene concentrations that were measured with detector tubes and combustible gas indicators (CGI) are representative of "background" or average concentrations in the RH plant areas. ... There is no claim or implication by Rinsky et al. that the industrial hygiene measurements reflect peak or maximum concentration, only that they overestimated the average concentrations." This conclusion is consistent with general industrial hygiene practice where plant industrial hygienists collect most of their samples for compliance and hazard recognition, i.e., worst-case scenarios [Checkoway et al., 1987; Harris, 1991].... Paustenbach et al. [1992] then adjust the detector tube and CGI samples upward to reflect "peak" exposure concentrations. These adjustments are not trivial, with concentrations as high as 750 ppm and time periods of up to 4 hr in seven of the eight job categories at both St. Marys and Akron. These adjustments are in addition to all others that are noted in this report."

Paustenbach: "Of all of the assumptions that went into our estimates of exposure, the one regarding whether the airborne concentrations represent background or "worst case" exposure is the least certain. However, we remain convinced that the measurement of benzene by grab sampling did not represent short-term peak exposures in the Pliofilm process. Despite Utterback and Rinsky's contention, a comparison of the grab and TWA samples offer significant support to this position. ... Perhaps the primary reason we believe that grab samples did not reflect "peak" exposures is that workers testified that they were instructed to remove their personal air samplers while using the respirators, thus indicating that the industrial hygienists and management were interested in the actual exposure of workers rather than the apparent level of exposure." "...even as far back as the 1940s, there has been a reluctance by hygienists within industry to permanently record measurements of these "acute" opportunities for high level exposure because they expected workers to wear respirators during these events and they didn't want to establish a paper trail of information that could be misconstrued as representing actual human exposure. In short, we don't take issue with what has been written by Checkoway et al. (1987) and Harris (1991) regarding the appropriate use and benefits of grab sampling, but we see little evidence that this was the goal of Goodyear at their Pliofilm facilities."

Evaluation: OEHHA tends to agree with the National Institute for Occupational Safety and Health that Paustenbach's assumption regarding the representativeness of the exposure estimates likely leads to an overestimation of exposure, however, there is uncertainty on this point. The adjustment for peak exposures is speculative; no data exist on peak exposures among Pliofilm workers. It also seems somewhat incongruous to suggest use of respirators for specific high exposure tasks, while adjusting the exposure estimates to reflect extremely high, extended exposures.

Point 5. Adjustments for analytical bias in early measurements

Paustenbach et al. (1992) increased the exposure estimates in the early years of the cohort to account for analytical bias in the assays used to estimate benzene concentrations (i.e., the detector tube assay and the combustible gas indicator (CGI) assay). In the case of the CGI assay, Paustenbach et al. (1992) increased relevant early exposures 1.5-fold, since they believed the CGI assays underestimate air concentrations of benzene by 50 percent. Utterback and Rinsky (1995) question the basis of these adjustments, claiming that the publications used to support these adjustments carry significant uncertainties. In response, Paustenbach discussed those uncertainties and noted acceptance of the adjustment by other groups in different situations. Mr. Confer, an industrial hygienist, provided a detailed set of comments in response to the Utterback and Rinsky critique. Mr. Confer discussed at length how these instruments were used in industrial settings in the 1940s and 1950s. He noted that most instruments were never calibrated following receipt from the manufacture and that sensitivity of the detector degraded over time.

Evaluation. The adjustments made to the overall exposure estimates from analytical bias were relatively small (zero to about 15 percent) (from Table 14 of Paustenbach et al., 1992). The arguments presented by Confer and Paustenbach suggest that these early assays likely underestimated exposures, but whether it was by 50 percent or much less than 50 percent cannot be ascertained. Having reviewed these comments, however, OEHHA was stuck by the overall poor quality of the early exposure estimates.

Point 6. Application of the TLV ratio method

Utterback and Rinsky (1995): "Crump and Allen [1984] had previously made upward adjustments in the NIOSH benzene exposure estimates for the RH worker cohort. Their method of adjustment assumes that as the TLVs became more restrictive, benzene exposures in industries were lowered, presumably through work practice and engineering control measures. Therefore, Crump and Allen contend that exposures for historical periods may be reasonably estimated based on proportionate changes in the TLVs for benzene." Utterback and Rinsky then state that when Paustenbach et al. (1992) applied this method to its exposure estimate, their average exposure estimates for some workers were extremely high. Later they state "The RH process that existed in 1975 when the process was terminated was fundamentally the same as it was shortly after it began in the early 1940s. Benzene exposures were monitored during the 1970s by a number of industrial hygienists using currently accepted methods [UNC, 1983; Rinsky et al., 1986]. There have been no reports of substantial engineering improvements after 1942 to limit benzene exposures to a large number of cohort members over the decades of operation at St. Marys and Akron. Rather, as initially reported by Rinsky et al. [1981] there had been some traditional engineering controls, such as a few hoods installed, and perhaps some work practices had been improved."

Paustenbach: "When one considers all of the points raised in our paper, we believe that our claim that industry was generally responsive to changes in the TLV is accurate. We believe that Utterback and Rinsky have not properly integrated all the pertinent information we presented. For example, we estimated that the airborne concentrations of benzene exceeded the prevailing TLV for only one job category…. We believe that the use of the TLV in both our work and in Crump and Allen's earlier work should best be viewed as a method of interpolation of benzene between the

1940s and the 1960s and 1970s, rather than an extrapolation from the measure levels in the 1960s and 1970s to earlier years." Paustenbach then argues that observations reported by Wilson (1942), Kipen et al. (1989) and others suggest early exposures (in the 1940s) were high enough to result in cases of benzene toxicity. Paustenbach also states "In our view, Utterback and Rinsky's related position that 'The RH process that was terminated was fundamentally the same as when it began in the 1940s. There have been no reports of substantial engineering improvements to limit worker exposures over the decades of operation at St. Mary's and Akron' is completely untenable. Based on the information offered in interviews of the plant superintendent and line workers, there were a number of engineering changes made that significantly reduced the airborne workplace concentrations of benzene in the Pliofilm facility."

Evaluation. It is very difficult to rectify such polar views. Paustenbach contends that engineering controls were improved through the early decades of the manufacturing process, and that evidence of benzene toxicity among Pliofilm workers in the 1940s support the notion of high early-year exposures and lower later-year exposures. On the other hand, Utterback and Rinsky contend that engineering controls were in place before 1942 and were not "substantially" changed over the following years of operation.

It appears that Paustenbach et al. was incorrect about the date that forced air hoods were installed (see discussion of point 2 above). For this reason, OEHHA tends to agree with Utterback and Rinsky that Paustenbach et al.'s exposure estimates are too high for the 1940s. As discussed in the discussion of Point 1 above, there is considerable debate as to whether there was widespread benzene toxicity among the Pliofilm cohort.

Overall evaluation

OEHHA concludes that the exposure estimates of Paustenbach et al. (1992), especially for the early years of the cohort, are likely to be overestimated. Since the primary difference among the three exposure estimates of Pliofilm workers (Rinsky et al., 1981; Crump and Allen; 1984; Paustenbach et al., 1992) is how they handled exposure estimates in the early years of the cohort when few exposure estimates were available, OEHHA feels it would not be justified in utilizing the Paustenbach et al. (1992) exposure estimates as the basis of a health-based toxicity level.

OEHHA notes that cancer potency estimates based on the Paustenbach et al. exposure matrix would fall within the range of potency estimates obtained from the Chinese Worker Cohort (Hayes et al., 1997). Thus, inclusion of cancer potency estimates based on the Paustenbach et al. exposure estimates would not change in any way the overall conclusions reached in the PHG for benzene.

References not cited in the PHG document

Edler L, Kopp-Schneider A (1998). Statistical models for low dose exposure. Mutat Res 405(2):227-236.

Gaylor DW (1992). Relationship between the shape of dose-response curves and background tumor rates. Regul Toxicol Pharmacol 16(1):2-9.

Hoel DG, Portier CJ (1994). Nonlinearity of dose-response functions for carcinogenicity. Environ Health Perspect 102 (Suppl 1):109-113.

Lutz WK (1998). Dose-response relationships in chemical carcinogenesis: superposition of different mechanisms of action, resulting in linear-nonlinear curves, practical thresholds, J-shapes. Mutat Res 405(2):117-124.

Preston DL, Kusumi S, Tomonaga M, Izumi S, Ron E, Kuramoto A, Kamada N, Dohy H, Matsuo T, et al. (1994). Cancer incidence in atomic bomb survivors. Part III. Leukemia, lymphoma and multiple myeloma, 1950-1987. Radiat Res 137(2 Suppl):S68-S97.

Selvin S. (1996). *Statistical Analysis of Epidemiological Data.* Second Edition, Monographs in Epidemiology and Biostatistics, Volume 25. Oxford University Press, New York, pp. 99-100.