



AMERICAN FOREST & PAPER ASSOCIATION

GROWING WITH AMERICA SINCE 1861



Dr. Michael Shelby
CERHR Director
NIEHS
79 T.W. Alexander Drive
Building 4401, Room 103
P.O. Box 12233
MD EC-32
Research Triangle Park, NC 27709

September 7, 2001

Re: Written Comments on Draft Methanol Expert Panel Report

Dear Dr. Shelby:

The American Forest & Paper Association (AF&PA) submits the following comments in response to the Center for Evaluation of Risks to Human Reproduction's July 16, 2001 request for comments on the Draft Methanol Expert Panel Report (the "Draft Report"), 66 Fed. Reg. 37,047. AF&PA is the national trade association of the forest, paper, and wood products industry. AF&PA represents more than 300 member companies and related trade associations involved in growing, harvesting, and processing wood and wood fiber; manufacturing pulp, paper, and paperboard from both virgin and recycled fiber; and producing solid wood products.

AF&PA previously responded, on October 2, 2000, to CERHR's August 17, 2000 request for data to be reviewed by the Methanol Expert Panel. AF&PA has a substantial interest in the assessment of risks presented by exposure to methanol, because naturally occurring methanol is released during the manufacture of wood products and wood pulp. AF&PA also intends to make oral comments during the public comment session planned for October 15, 2001.

AF&PA submitted extensive analysis of the potential for exposure to methanol emissions to create risks to human health and the environment, including adverse effects on human reproduction and development, in connection with its March 8, 1996 petition to EPA to remove methanol from the list of "hazardous air pollutants" under the Clean Air Act. Those materials were also provided to CERHR in AF&PA's October 2, 2000 submission. AF&PA strongly believes that this information demonstrates sufficient data are available to conclude there is no significant risk of adverse effects on human reproduction and development from exposure to methanol via air pollution. AF&PA believes that the information it previously provided contains

1111 Nineteenth Street, NW, Suite 800 _ Washington, DC 20036 _ 202 463-2700 Fax: 202 463-2785 _
www.afandpa.org

America's Forest & Paper People® - Improving Tomorrow's Environment Today®

important analyses that the Expert Panel should consider, and in fact the Draft Report reflects consideration of some, but not all, of the points contained in AF&PA's analysis. The following comments suggest several areas in which AF&PA believes the Draft Report should be modified to make additional or more accurate reference to the materials AF&PA has provided.

These comments rely substantially on reports, attached to this letter, prepared by two highly qualified experts who not only have reviewed portions of the Draft Report but also have done extensive analysis of the study by Burbacher, *et al.* on effects of methanol inhalation on non-human primates, which is referred to extensively in the Draft Report. In addition, Dr. Starr has conducted or supervised extensive analyses of the risks presented by exposure to ambient concentrations of methanol, in conjunction with AF&PA's above-mentioned petition to remove methanol from the hazardous air pollutants list.

Thomas B. Starr holds a doctorate in physics, with post-doctoral work in environmental studies. He is currently a consultant in risk assessment and an adjunct associate professor in the Department of Environmental Sciences and Engineering at University of North Carolina at Chapel Hill School of Public Health. Dr. Starr has spent more than 30 years in the field of toxicology and risk assessment, including a decade as Senior Scientist and Director of the Program on Risk Assessment at the University of Wisconsin-Madison. He is a former president of the Society for Risk Analysis and the Risk Assessment Specialty Section of the Society of Toxicology, and he has served on advisory boards to EPA, Duke University, and the State of North Carolina. He has over 70 publications on human and environmental health effects of exposure to pollutants and other toxic substances.

David G. Hoel has a Ph.D. in Statistics from the University of North Carolina at Chapel Hill. He is currently Distinguished University Professor at the Medical University of South Carolina. Previously he had a long association with the National Institute of Environmental Health Sciences, including serving as its Acting Director and the Director of the Division of Biometry and Risk Assessment. Dr. Hoel has served on numerous National Academy of Sciences committees and other U.S. government advisory committees and serves on the editorial board of numerous publications, including the *Journal of Statistical Computation and Simulation*, the *Journal of Communications in Statistics*, and the *Journal of Environmental Pathology, Toxicology and Oncology*.

Discussion of Developmental Toxicity in Light of Burbacher Study

The Draft Report's discussion of the Burbacher study, sponsored by the Health Effects Institute (HEI), on methanol disposition and reproductive toxicity in adult females and offspring developmental effects following maternal inhalation exposure (References 41 and 93 of the Draft Report) fails adequately to consider the limitations and shortcomings of the statistical analysis of the study. AF&PA believes that the Burbacher study does have substantial value for the Expert Panel's task, but that value is that the Burbacher study is a comprehensive assessment of the reproductive and developmental toxicity of maternal methanol inhalation which shows no meaningful adverse effects for exposures as high as 1800 ppm.

The Burbacher study itself, and especially the included HEI peer-review commentary, present findings that at most suggest areas for further research, rather than confirming any adverse effects on mothers or their offspring from exposure to up to 1800 ppm of methanol. A large number of tests were performed, and yet the analysis of variances showed no statistically significant difference between the control group and the exposed groups in any of these measures of reproductive and developmental toxicity. Only when the researchers performed *post hoc* “linear contrast” comparisons between various groups did any differences emerge. The HEI peer-review panel and AF&PA’s experts all conclude that these analyses could easily identify apparently differences between controls and exposed animals merely by chance, given the small number of animals, the multitude of tests, and the variability of individual responses.

The statistical analyses in the Burbacher Study present the possibility of misconstruing random fluctuations as effects of methanol exposure. The information that might be used to corroborate statistically identified differences in fact tends to disprove the hypothesized effects. As the HEI peer-review commentary notes and AF&PA’s experts state even more strongly, the lack of clear, monotonic dose-response relationships, despite clear differences in blood methanol concentrations; the lack of consistency among cohorts, sexes, and tests; and the difficulty of explaining apparent effects in a 200 ppm group, where maternal blood methanol was only slightly elevated above background; all undercut any assertion that the study demonstrates an effect of methanol on reproductive or developmental health.

The attached report and memorandum from Dr. Hoel detail the shortcomings in the statistical analyses and conclusions of the Burbacher study. Dr. Hoel demonstrates that the data generated by Burbacher “provide a good example of how a large number of statistical tests can produce a few inconsistent, but entirely expected, positive results even when the experiment is truly negative. Based on the sheer number of statistical tests that were employed by Burbacher *et al.* and their failure to adequately control the experiment-wide false positive error rate, we are forced to conclude that there is no convincing evidence for an effect of methanol exposure on the behavioral measures evaluated in these primates.” Contrary to the assertion in the draft Report that “it is not clear, what would be the most appropriate adjustment to make” for the fact that multiple comparisons were made (Draft Report at page 64), Dr. Hoel shows that there are accepted statistical techniques for doing so. In any event, there is no basis for ascribing developmental effects to methanol exposure based on the Burbacher study, given that the few effects reported were not only inconsistent and in some cases contradictory, but were undoubtedly within the range of false positives that would be expected in the statistical analyses Burbacher performed.

As AF&PA’s experts and the HEI peer-review committee noted, the Burbacher study describes a very extensive examination of potential reproductive and developmental toxicity effects. The fact that this study did not produce any clear indications of such effects, even at concentrations almost three orders of magnitude greater than the maximum fenceline concentrations predicted to result from methanol emissions, is confirmation that adequate data *do* exist to support the conclusion that there is no *reasonable* anticipation of adverse health effects from methanol emissions.

Possible Effect on Gestation Length in Burbacher Study

The Draft Report describes, without reaching any definitive conclusions, the statement in the Burbacher study that maternal exposure to airborne methanol resulted in reduced gestation length. The analyses of the Burbacher study which AF&PA provided to EPA on July 3, 2000 and September 1, 2000 demonstrate that the Burbacher study does not provide sufficient basis to conclude that methanol exposure had any effect on gestation length. First, gestation length for all of the exposed cohorts was within the normal range for *Macaca fascicularis*. Second, the observed reduced gestation length was not accompanied by any other signs of pre-maturity, such as reduced birth weight or reduced head circumference. Third, no dose-response relationship was observed. Fourth, and perhaps most importantly, the observation of reduced gestation length was largely dictated by the fact that one offspring in the control group had an abnormally long gestation length, accompanied by signs of post-maturity. As explained in AF&PA's submissions to EPA and in the attached report by Dr. Hoel, this one outlier in the control group (whose duration of pregnancy of 178 days was more than two standard deviations beyond the observed mean of 167 days for the control group) "leads to an invalid inference that the exposed groups' pregnancy durations were significantly shortened by methanol exposure." Excluding that outlier from the control group results in a conclusion that there was no significant difference in the gestation lengths between the control groups and the exposed groups.

Finally, Burbacher's observations of reduced gestation length were also influenced by the fact that there were a relatively large number of Cesarean section births (five) in the exposed groups, but none in the control group. Dr. Alice Tarantal, an expert primatologist with a particular expertise in prenatal and neonatal care of primates and especially *Macaca fascicularis*, in her report that AF&PA submitted to EPA on July 3, 2000 (Reference 94 in the Draft Report) and in materials presented to EPA in September 1 (not referenced in the Draft Report), observed that spontaneous vaginal bleeding, which apparently triggered the decision to perform Cesarean sections in four of five cases, is not a reliable indicator of maternal or fetal distress, and therefore the high incidence of Cesarean sections in the exposed groups is "most likely spurious."

Thus, in AF&PA's July 3, 2000 submissions to EPA, Dr. Tarantal, Dr. Hoel, and Anthony Scialli, M.D. of Georgetown University School of Medicine all concluded that the Burbacher study does not provide evidence of methanol reproductive toxicity. For the reasons stated above and in those submissions, AF&PA agrees with the conclusion stated by two of the panel members, that "the reduced length was not clinically relevant" and "the data do not indicate that methanol is a female reproductive toxicant in female macaques." (Draft Report at 65.)

Downplaying of Aspartame and Stanton Studies

The Draft Report describes extensive studies by Reynolds *et al.* (Reference 97 in the Draft Report) and Suomi (Reference 98), in which aspartame was fed to five groups of monkeys, each group consisting of four monkeys. Because aspartame is hydrolyzed to methanol in the gut, the Rogers and Suomi studies represent exposure of infant monkeys to methanol ranging up to 250-270 mg/kg bw/day. That methanol exposure did not have an effect on growth or numerous

developmental milestones, including various measures of learning performance and hearing ability.

The Draft Report somewhat discounts the value of these studies because “the statistical power of the hypothesis tests is unclear,” because “[t]he studies did not find any effects at the dosage used,” and “the only useful information to come from them is that the highest dose APPEARS to be tolerated.” (Draft Report page 66.) As Dr. Hoel points out, the comments about the statistical power of the testing apply equally to the Burbacher study; in any event, the statistical power calculations could be carried out at this point for both studies. Most important, as Dr. Hoel notes, is that the findings of Reynolds and Suomi are consistent with the findings (properly interpreted) of Burbacher; namely, that both prenatal and neonatal exposure to methanol in doses likely to substantially exceed human exposures do not have an effect on growth or neurobehavioral development. It is perplexing that the Draft Report seems to regard this important finding as a “weakness” or “flaw” of the Reynolds and Suomi studies.

Similarly, the Draft Report seems to downplay the significance of a very relevant study by Stanton *et al.* on the post-natal effects of *in utero* exposure of rats to high concentrations of airborne methanol. That study showed no indication of effects of methanol exposure on a battery of neurobehavioral tests measuring sensory, motor, and cognitive functioning. The Draft Report concludes: “The overarching weakness of the study is that effects were not found and that the group size, (n=6-7 with litter as the unit of measure) was too small for the tests employed to have statistical power to pick up deficits with known developmental neurotoxicants.” (Draft Report at page 60.) Once again, it appears as if the study is being discounted because it did not find neurobehavioral development effects, rather than noted as tending to confirm the other studies that showed no adverse effects on neurobehavioral development. Also, as AF&PA has noted previously, the group size was also quite small in the Burbacher study, with some comparisons between groups involving two or three individuals; that limitation should be treated similarly in the Expert Panel’s evaluation of these two studies (or the Expert Panel should explain why it is not).

Pharmacokinetic Modeling

The attached report from Dr. Starr points out a number of ways in which the Draft Report’s discussion of PBPK models by Perkins *et al.* and Horton *et al.* are inaccurate or incomplete. AF&PA urges the Expert Panel to consider Dr. Starr’s comments carefully, as (1) there is a substantial amount of pharmacokinetic information available on methanol and (2) that information can be critical in the evaluation of potential health risks from methanol exposure.

Importantly, Dr. Starr points out that both the Perkins model and, to an even greater extent, the Horton model do a good job of predicting human blood methanol concentrations resulting from exposure to airborne methanol at the concentrations likely to be relevant for the Expert Panel’s risk assessment. Data are available on changes in blood methanol concentration as a result of human exposure to known concentrations of airborne methanol, in several studies. Those experimental results agree quite well with the PBPK models’ predictions of resulting blood methanol concentrations. This information is particularly important (1) for comparing

potential effects on humans to experimental effects in rats and mice (humans being substantially less sensitive than rats and even less sensitive compared to mice) and (2) for assessing whether changes in human blood methanol that would result from environmental exposures to methanol are biologically relevant (in comparison to endogenous generation and retention of methanol through human metabolism).

Importance of Evaluating Risks in the Context of Natural Methanol Generation

The Draft Report recognizes briefly that humans generate methanol as a result of the normal metabolism of various foods, fruits, and beverages, and that humans therefore have a baseline blood methanol concentration of approximately 1.8 mg/l. The Draft Report also mentions, but without any analysis at all, that baseline human blood methanol concentrations vary substantially. In fact, the studies cited in the Draft Report, and analyzed in AF&PA's previous comments, show that humans naturally experience substantial variation, both individually and between individuals, in their blood methanol concentrations.

The Draft Report is entirely lacking, however, in an assessment of the implications of these facts, especially in conjunction with the validated PBPK models discussed above. It would be irrational and unscientific to conclude, based on studies of methanol effects on other species and application of multiple safety factors (and, in the case of the Burbacher study ascribing adverse effects based on random false positives), that exposure to low concentrations of methanol presents risks to human health, if such exposures are insignificant in relation to the methanol exposure we all have as result of our normal biological processes.

As Dr. Starr explains in his report, through PBPK modeling one can show that exposure to the predicted worst-case fenceline concentration of methanol resulting from industrial methanol emissions would increase blood methanol concentration by only 0.07 mg/l. This compares to an average baseline blood methanol concentration of 1.8 mg/l and, in the Batterman study (Reference 25 in the Draft Report), an interindividual standard deviation from that baseline of 0.7 mg/l. In other words, individuals exposed to the highest ambient concentration of methanol would be expected to have blood methanol concentrations only about 4% higher than the mean and, even more significantly, far less than the mean plus the standard deviation. (Average blood methanol would increase by only about 10% of the standard deviation.)

It would not make any sense to ascribe toxic effects to methanol exposures that result in aggregate blood methanol concentrations (baseline concentration plus the increment resulting from methanol exposure) that are still well within the normal range of unexposed individuals. AF&PA urges the Expert Panel to consider this issue carefully in its assessment: if extrapolation from animal studies and application of multiple safety factors would result in a conclusion that exposure to airborne methanol in the tens of parts per million (which would still result in total blood methanol that is in the normal range for unexposed individuals) presents a significant risk to human health, something is wrong with the analysis. The only justifiable conclusion is that either the studies were faulty, or the methods for extrapolating from those studies were faulty or excessively conservative. As Dr. Starr points out, if the maximum expected airborne exposure of 3.7 mg/m³ presents a significant risk, then this implies that 46% of the population is at risk even

without any anthropogenic methanol exposure, due to their own metabolic processes. The public would not be served by such an irrationally alarmist conclusion.

Use of Benchmark Doses

The Draft Report presents “benchmark dose” estimates for the developmental toxicity effects reported in Rogers, *et al.* (Draft Report at page 57.) As explained in Dr. Starr’s attached report, it would be inappropriate for the Expert Panel to rely on benchmark dose estimates from the Rogers study in reaching conclusions about methanol developmental toxicity. First, the published report of the Rogers study fails to provide critical information needed to independently replicate (or modify) their model-fitting process and subsequent benchmark dose derivations. Second, the lower-bound benchmark dose estimates presented in the Draft Report are in fact the lower 95% confidence bound estimates of the dose that would pose a 5% added risk. Use of this measure could introduce a substantial, unnecessary additional element of conservatism into the risk assessment process. This lower-bound benchmark dose estimate is necessarily substantially lower than the corresponding NOAEL derived from the same data.

Thus, the benchmark dose estimates contained in the Draft Report would not be an appropriate basis for a reference concentration or a reference dose. Moreover, use of such a benchmark dose is unnecessary since the Rogers study provides a NOAEL dose that would be a justifiable basis for deriving a referenced concentration.

Inaccurate Statement of Maximum Predicted Environmental Exposure

The Draft Report, at page 3, suggests that dispersion modeling of predicted ambient exposure to methanol conducted by AF&PA showed an estimated maximum predicted 24-hour methanol concentration from all sources identified as emitting 500 tons or more of methanol per year of 7.58 mg/m³. In fact, while initial modeling of the source with the highest predicted ambient impact (occurring near the source’s property line) produced a predicted maximum 24-hour concentration of 7.58, AF&PA subsequently did additional modeling, reflecting the lower emission rates that resulted from changes in the facility between 1995 and 1999. That later, more representative modeling predicted a worst-case 24-hour fence-line concentration of 3.65 mg/m³ (2.8 ppm). That revised modeling was presented to EPA in AF&PA’s submission of July 3, 2000, and was included in AF&PA’s October 2, 2000 submission to CERHR. Thus, the best information available suggests that worst-case 24-hour exposures would be over 2 _ times less than presented in the Draft Report.

AF&PA hopes that these comments will be useful to the Methanol Expert Panel as it completes its important work. We urge the Panel to complete the Methanol Expert Panel Report promptly, as there is significant interest in this subject, and moreover we believe that the Draft Report contains some inaccurate or incomplete statements that should be corrected promptly.

Please contact the undersigned with any questions at (202) 463-2587, fax (202) 463-2423, or john_festa@afandpa.org.

Sincerely,

John L. Festa, Ph.D.
Senior Scientist

Attachments

**Technical Comments on the Draft
NPT-CERHR Expert Panel Report on
Reproductive and Developmental Toxicity of Methanol**

**prepared by
Thomas B. Starr, Ph.D. & Principal
TBS Associates
Raleigh NC**

**on behalf of the
American Forest & Paper Association
Washington DC**

7 September 2001

1. Introduction

The efforts of the CERHR Expert Panel to critically review the scientific literature regarding methanol are to be commended. It is sincerely hoped that the Panel will develop a coherent and integrated assessment of the potential human risks that might be posed by methanol exposure. Such an assessment would be a valuable resource for this important chemical.

At the present time, however, the Expert Panel Draft document falls well short of providing a coherent synthesis of the methanol toxicity literature. The Summaries, Conclusions and Critical Data Needs sections of the Draft are completely unwritten, and the literature review chapters of the Draft are, for the most part, simple recitations of individual study findings as they have been reported in the literature. No attempt seems to have yet been made to integrate these observations into a coherent and comprehensive picture of potential methanol toxicity. If the Panel's efforts are to be truly useful in the risk assessment process, such an integration of the assembled body of data must be the Panel's primary focus.

The following comments pertain to two important areas covered in the Expert Panel Draft document. The first relates to issues concerning the pharmacokinetics of methanol and the use of pharmacokinetic models to predict blood methanol levels as an aid in understanding significant interspecies differences that exist with regard to methanol toxicity. The second relates to the potential use of benchmark dose methodology in establishing "points of departure" for reference concentration calculations, which we believe to be ill-advised at the present time. In these comments, please note that citations appearing in the Draft document References section have been identified by the reference numbers employed in the Draft document. All other citations are fully identified in the References section of these comments.

2. Issues Concerning Pharmacokinetics

Comments on the Draft Discussion of the Perkins *et al.* Models

Section 2.1.6 of the Draft reviews physiologically-based pharmacokinetic models of methanol uptake and disposition. The discussion of the "semiphysiologic" models developed by Perkins and colleagues (49, 50) is the most extensive, and calls attention to a number of important facts. Most notable is the fact that methanol metabolism is a saturable, enzyme-mediated process, i.e., a process with a finite capacity, that generates a nonlinear relationship between administered methanol doses and the corresponding levels of methanol in blood, at least at sufficiently high doses. Furthermore, the onset of saturation for methanol metabolism is species-specific. For example, Perkins *et al.* showed that mice developed the highest blood methanol levels, followed by rats, and then humans, when all three species were identically exposed to a sufficiently high airborne methanol concentration. At 5,000 ppm methanol, for example, the blood methanol level in mice is predicted by the Perkins *et al.* model (50) to be 13-18-fold higher than that for humans, while the rat would be 5-fold higher than that for humans.

Perkins *et al.* (50) emphasized the significance of these interspecies differences in methanol metabolism in accounting for interspecies differences in methanol-induced toxicity, stating that "These results demonstrate the importance in the risk assessment for methanol of basing

extrapolations from rodents to humans on actual blood concentrations rather than on methanol vapor exposure concentrations.” This point needs much stronger emphasis in the CERHR Draft. Blood methanol levels provide a unifying “delivered dose” concept that should facilitate integration of the data collected in different species under different exposure conditions into a coherent body of information that is directly relevant and useful when addressing the potential human risks posed by dietary, occupational, and environmental methanol exposures.

Among the Expert Panel conclusions regarding the Perkins *et al.* work is a statement on page 30 of the Draft document that “The caveat that needs to be applied is that the exposure levels were high. This can lead to anomalies in absorption.” These statements are misleading for two reasons.

First, Perkins *et al.* fit their pharmacokinetic model to a variety of data sets including time course data for human volunteers exposed for 8 hours to airborne methanol concentrations of 77, 156.5 and 229 ppm. Since the OSHA TLV is 200 ppm, these concentrations cannot be considered to be too high to be directly relevant to realistic human exposure situations.

Second, while it is correct that Perkins *et al.* reported that methanol absorption, as represented by the absorption factor Φ , is reduced in rats exposed to increasing inhalation concentrations up to 20,000 ppm, this appears to be an extremely high dose, and possibly rat-specific, phenomenon, as well. In their Table 2, which portrays species-specific absorption factors estimated at various airborne methanol concentrations, the only important difference occurs between rats exposed to 2,500 ppm and 10,000 ppm methanol respectively. Perkins *et al.* state that there is a “lack of effect of exposure concentration on Φ (the absorption factor) in humans” They state further that “ Φ in mice also was independent of exposure concentration.” There are thus no “anomalies” in absorption at methanol concentrations of 2,500 ppm or lower in any species.

In addition, a recent paper by Fisher *et al.* (2000) has explored the “wash-in-wash-out” phenomenon in greater detail with a physiologically-based pharmacokinetic model of methanol uptake in monkeys. They concluded that this phenomenon “behaves as a simple linear process over a wide range of exposure concentrations and for various exposure durations... .” Thus, a constant absorption factor provides an adequate characterization of methanol uptake in the respiratory tract except in extreme circumstances. It follows that the nonlinear relationships between blood methanol levels and airborne methanol concentrations are not due to “anomalies in absorption.” They are attributable solely to the fact that methanol metabolism is a saturable process that cannot keep up with methanol intake as the airborne methanol concentration increases.

Finally, the Expert Panel’s conclusions regarding the Perkins *et al.* models include the surprising statement on p 30 that “The models need to be validated at lower exposure doses before they can be applied to humans.” On p 45 of the Draft, the Perkins *et al.* models are also described as remaining “unvalidated for the extrapolation of dose effects to levels associated with current or anticipated exposures of the general and worker population.”

Given that Perkins *et al.* fit their model to human blood methanol time course data during exposure to airborne methanol concentrations ranging from 77 ppm up to 5,000 ppm, this

statement makes no sense. Furthermore, Starr (1999) has shown that both the Perkins *et al.* model and the Horton *et al.* model (42) (discussed in more detail below) predicted initial blood methanol accumulation rates and steady state blood methanol levels for monkeys exposed to 200 ppm that were in excellent agreement with the measured values reported by Burbacher *et al.* (86). This exercise provides a truly independent validation of both pharmacokinetic models, albeit using a non-human primate, at the OSHA TLV of 200 ppm.

Finally, as is discussed in more detail below, the Perkins *et al.* “semi-physiologic” one compartment model predictions for mice, rats, monkeys, and humans exposed to various airborne methanol concentrations ranging from 7 to 1,300 ppm are in essential agreement with those from the far more detailed physiologically based pharmacokinetic model of Horton *et al.*, a model in which the Expert Panel has greater confidence.

Comments of the Draft Discussion of the Horton *et al.* Model

The Horton *et al.* model (42) is given remarkably short shrift in the Draft. It is discussed in just two short paragraphs on p 30. In the first sentence of the first paragraph, this physiologically-based pharmacokinetic model is described as a “two-compartment PBPK model which does not include a fractional absorption parameter (ϕ).” The Horton *et al.* model is in fact a 4-compartment model as is correctly stated in the second sentence of the same paragraph.

The Draft is also correct in stating that this model does not account explicitly for retention of methanol in the mucus lining of the respiratory tract, a characteristic represented in the Perkins *et al.* model by an absorption fraction whose value is less than unity. A direct consequence of this “deficiency” is that, with all other factors, including airborne methanol concentration, being equal, the Horton *et al.* model can be expected to predict somewhat higher blood methanol concentrations than would the Perkins *et al.* model, and this has been observed (see comments on ENVIRON analysis below). This property was also predicted by Fisher *et al.* (2000) in their rigorous analysis of the “wash-in-wash-out” phenomenon: “the absorbed dose of methanol vapors would be overestimated, if the “wash-in-wash-out” phenomenon were not to be considered.”

Thus, the Horton *et al.* model can be expected to provide more conservative estimates of risk from methanol exposure than would the Perkins *et al.* model, at least at methanol concentrations sufficiently low to permit the Perkins *et al.* model to achieve steady-state.

The critical advantages that the Horton *et al.* model has over the that of Perkins *et al.* are its two pathways for methanol metabolism, by either the catalase or alcohol dehydrogenase enzyme systems, and its explicit accounting for methanol elimination via excretion in urine and expired air. These additional components prevent the Horton *et al.* model from “blowing up” unrealistically at extremely high airborne methanol concentrations. In contrast, the Perkins *et al.* model predicts that under such conditions, blood methanol levels will continue to increase from exposure onset without bound, which is simply not possible. As a consequence of its physiologic structure, the Horton *et al.* model also predicts directly measurable quantities such as methanol concentrations in expired air, urine and specific tissues such as liver and kidney as well as that in blood.

Comments on the Draft Discussion of the ENVIRON Analysis and the Need for Integration of Pharmacokinetics and Toxicity Data Across Species and Exposures

The Draft discussion of ENVIRON's comparative analysis of the Perkins *et al.* and Horton *et al.* models (55) on pp 30 and 45 is also surprisingly brief. The ENVIRON analysis is exemplary of the kind of integration of data across species and studies that is sorely lacking in the Expert Panel's Draft document. For example, in Table 4 of the ENVIRON analysis, predictions of initial rates of methanol accumulation in blood and steady state blood methanol concentrations were tabulated for mice, rats, monkeys, and humans identically exposed to airborne methanol concentration ranging from 7 to 2,600 mg/m³.

The low end of this exposure range was described as "just slightly larger than any dispersion modeling-predicted 24-hour average concentration at or beyond the fencelines of existing methanol-emitting facilities (PES 1996)." It is thus conservatively representative of a "worst case" continuous environmental exposure to methanol. The high end of this range was chosen to be the LOAEL concentration for the most sensitive toxic endpoint reported in the toxicology literature, namely, cervical rib malformations in the offspring of mice exposed to methanol via inhalation during gestation, as observed by Rogers *et al.* (83).

Entries in this table from the ENVIRON analysis clearly reveal the consequences of the saturable metabolism of inhaled methanol and the marked interspecies differences in projected steady-state blood methanol levels. For example, while 2,600 mg/m³ is only slightly more than 31-fold greater than ENVIRON's recommended human Safe Exposure Level of 83 mg/m³, the steady-state blood methanol level in mice exposed to 2,600 mg/m³ via inhalation is predicted by the Horton *et al.* model to be more than 1,000-fold higher than the steady-state blood methanol level in humans exposed to 83 mg/m³. This indicates clearly why mice are significantly more susceptible than humans to the potential developmental toxicity of methanol exposure: mouse blood methanol levels rise much more rapidly and nonlinearly than those of humans as the airborne methanol concentration to which these species are exposed increases. It also demonstrates that humans have a greater margin of protection from potential adverse effects of methanol exposure than would be expected based solely upon consideration of the airborne methanol concentration (Starr *et al.* 2000).

One can also use the results presented in Table 2 and 4 of the ENVIRON analysis (55) to perform additional independent validations of the model predictions against actual human data. For example, Lee *et al.* (23) reported that human serum methanol levels rose from a pre-exposure mean of 1.82 mg/L to 6.97 mg/L, an increment of 5.15 mg/L following exposure to a methanol concentration of 200 ppm (260 mg/m³) for 6 hours. Since the Perkins *et al.* and Horton *et al.* models are essentially linear below 1,000 ppm for humans, one can simply multiply the model predictions for exposure to 83 mg/m³ as presented in Tables 2 and 4 of ENVIRON (55) by the ratio $260/83 = 3.13$ to obtain the corresponding blood level predictions resulting from exposure to 260 mg/m³. For the Perkins *et al.* model, the predicted steady-state increment in blood methanol level is $3.13 * 2.3 = 7.2$ mg/L, while for the Horton *et al.* model, this predicted increment is $3.13 * 1.6 = 5.0$ mg/L. These predictions are thus in excellent agreement with

completely independent human observations.

Similar use can be made of the report by Osterloh *et al.* (28), wherein a 4 hour exposure of human volunteers to 200 ppm methanol was shown to produce a mean blood methanol increment of $6.5 - 1.8 = 4.7$ mg/L. Again, the same model predictions of 7.2 and 5.0 mg/L for the Perkins *et al.* and Horton *et al.* models, respectively, are in excellent agreement with the human data.

Finally, the Expert Panel is strongly encourage to supplement Tables 7.2B- 7.2E with model-predicted steady-state blood methanol and formate concentrations arising from continuous exposure to the various reported methanol concentrations. Such predictions would represent upper bounds on the blood methanol levels that could be achieved if the reported exposures were to continue indefinitely. It would also appear to be equally useful to generate the same kinds of projections for the exposure levels presented in the toxicity Tables 7.3A-7.3Q and 7.4A-7.4D. At present, there is virtually no integration of these data across species and exposure levels in the Draft document.

Comments on the Importance of Human Baseline Blood Methanol Levels and Interindividual Variability in Providing a “Reality Check” on Extrapolation

The fact that methanol is an endogenous biochemical that arises from the normal metabolism of various foods, fruits, and beverages is briefly acknowledged in the Draft document on pp 3 and 7. In addition, several reports in which human baseline blood methanol levels were characterized are discussed at some length in Section 2.1.1.1 of the Draft document. For example, it is noted therein that Lee *et al.* reported a mean pre-exposure blood methanol level for three human subjects of 1.82 ± 1.21 mg/L (23). Similarly, in the Batterman *et al.* study (24), baseline blood methanol levels averaged 1.8 ± 0.7 mg/L in twenty subjects. However, the potential implications of these endogenously generated blood methanol levels for risk assessment were not addressed.

To place the mean baseline blood methanol level in perspective, it is useful to determine the airborne methanol concentration that would be required to generate and sustain such a blood level in the absence of any natural sources. This is readily accomplished with use of the Horton *et al.* PBPK model. Indeed, the ENVIRON analysis (55) has already shown that a steady-state increment in human blood methanol of 1.6 mg/L would arise from continuous exposure to an airborne concentration of 83 mg/m³, or approximately 64 ppm methanol. Because the Horton *et al.* model is virtually linear at such low exposure levels, a steady-state human blood methanol level of 1.8 mg/L would therefore require continuous exposure to about 93 mg/m³ methanol ($1.8/1.6 \times 83$), absent any endogenous sources.

This level provides one very useful measure against which the impacts of environmental releases can be gauged. If the largest such release generates fenceline concentrations no greater than about 3.7 mg/m³ (Starr, 2000), and this concentration is only about 4% of the airborne concentration required to sustain the normal human baseline blood methanol level, there cannot be any serious concerns regarding the potential for toxicity as a consequence of that incremental methanol exposure.

Additional insight and reassurance is gained when one compares the increment in human blood

methanol levels predicted to arise from exposure to such fenceline concentrations with the reported standard deviation of human baseline blood methanol levels. Using the Horton *et al.* model, Starr (2000) showed that continuous exposure to 3.7 mg/m³ methanol leads to a predicted steady-state increment of only 0.07 mg/L in human blood methanol levels. As noted above and in the Draft on p 13, Batterman *et al.* (24) estimated the interindividual standard deviation of this baseline level to be 0.7 mg/L, so the predicted increment of 0.07 mg/L represents only about 10% of the interindividual standard deviation in baseline levels. This increment is thus a trivially small signal compared to the scale of interindividual variability in the baseline blood methanol level.

Indeed, as Starr (2000) noted, “even without airborne exposure, and assuming a normal distribution of individual baseline values about the mean, approximately 46% of the population would be expected to have blood methanol levels greater than the mean baseline *plus the increment*, i.e., greater than about 1.87 mg/L. Thus, if an increment of 0.07 mg/L in human blood methanol is to be regarded as potentially harmful, one must accept the untenable implication that nearly half of the population is already at risk simply from dietary and metabolic contributions to blood methanol, without any additional contribution from airborne methanol exposure.” Starr (2000) concluded, “Common sense dictates that downward extrapolations from the very high levels that cause frank toxicity must not be continued downward to levels so small as to be trivial when viewed from a “bottom up” and “additive to background” perspective.” The Expert Panel Draft needs to address this important issue.

2. Issues Concerning the Use of Benchmark Doses Rather than NOAELs and LOAELs

In its review of the study by Rogers *et al.* (83), the Expert Panel Draft document acknowledges that the authors identified a developmental toxicity LOAEL of 2,000 ppm and a corresponding NOAEL of 1,000 ppm using cervical rib malformations as the most sensitive endpoint. In addition, the Draft presents central and lower bound benchmark dose estimates derived by Rogers *et al.* for this and related endpoints in Table 3-1 on p 56. There is some confusion created by the terminology used in this table and the related discussion. While the central estimates of the doses estimated to pose a 5% added risk are described as maximum likelihood estimates (MLEs), the lower 95% confidence bound estimates are referred to simply as benchmark doses (BMD_{05s}). This gives the distinct impression that only the lower bound estimates are actually benchmark dose estimates, but this is not correct.

Both the maximum likelihood and lower bound estimates are benchmark doses. The former is a central estimate while the latter is a lower bound (i.e., biased) estimate. An easy way to eliminate the confusion is to refer to the central estimates as BMD_{05s} and the lower bound estimates as BMDL_{05s}. This is consistent with terminology that is now in fairly widespread use in the general toxicology literature.

A more important issue concerns whether any of the benchmark dose analyses of Rogers *et al.* should be preferred over the corresponding NOAELs and LOAELs as potential bases for establishing a reference concentration for the developmental effects of methanol. A significant problem with the Rogers *et al.* analyses is that the documentation of their model-fitting process is quite poor. For example, they state that they fit a generalization of the log-logistic model

proposed by Kupper et al. (1986) to the incidence data for cervical rib, but the exact model form they employed is not identified explicitly in their publication. While it appears to be the six parameter LOG model described by Equation 2 in Allen *et al.* (1994), which allows for possible effects of litter size on the incidence of cervical rib in addition to those related to dose, this is not known with certainty. Furthermore, the specific statistical procedure employed by Rogers et al. in developing lower confidence bound estimates of benchmark doses also was not described in their publication.

In addition, somewhat surprisingly, Rogers *et al.* provided estimates only of the dose-related parameters of the LOG model in their published report, omitting the parameters related to possible effects of litter size. Furthermore, they describe the benchmark dose estimates presented in their Table 6 as being appropriate "for the average litter size in the experiment," which itself is not provided. The Rogers *et al.* publication thus fails repeatedly to provide critical information that would permit an independent replication (or modification) of their model fitting process and their subsequent benchmark dose derivations.

It is also worth noting that use of lower bound benchmark doses (as were identified in the Expert Panel Draft) as "points of departure" for reference concentration or margin-of-exposure calculations introduces additional conservatism into the risk assessment process without explicit acknowledgment or justification. For example, the work of Allen et al. (1994) indicates that on average, Weibull model BMDL_{10s} for developmental toxicity endpoints are approximately 3-fold lower than the corresponding NOAELs derived from the very same studies. Use of still smaller incremental response rates, such as 5% or 1%, leads to even greater disparities between the estimated BMDLs and the corresponding NOAELs. Such disparities would carry over to calculated reference doses or reference concentrations, implicitly introducing additional conservatism. This conservatism is not scientifically justified, and it could be readily avoided by employment of a central estimate of the "point of departure" or, better still, the corresponding NOAEL.

In summary, while the NOAEL dose of 1,000 ppm in the Rogers et al. study provides an entirely adequate and scientifically justified basis for deriving a reference concentration for methanol, the Rogers *et al.* benchmark dose analyses do not. We strongly encourage the Expert Panel to recommend use of the traditional NOAEL-based methodology in establishing a reference concentration for methanol's developmental effects.

References not cited in the Expert Panel Draft document

- Allen BC, Kavlock RJ, Kimmel CA, Faustman EM. 1994. Dose-response assessment for developmental toxicity. II. Comparison of generic benchmark dose estimates with no observed adverse effect levels. *Fundam Appl Toxicol* 23:487-495.
- Fisher JW, Dorman DC, Medinsky MA, Welsch F, Conolly RB. 2000. Analysis of respiratory exchange of methanol in the lung of the monkey using a physiological model. *Toxicol Sciences* 53:185-193.
- Kupper LL, Portier C, Hogan MD, Yamamoto E. 1986. The impact of litter effects on dose-

response modeling in teratology. *Biometrics* 42:85-98.

PES. 1996. *Methanol. Environmental Emission Inventory and Air Quality Impact Analysis*. Pacific Environmental Services, unpublished report, dated March 1996. Cited in ENVIRON (55).

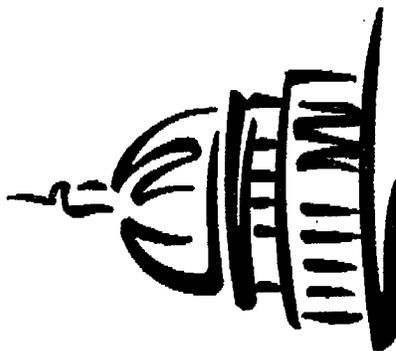
Starr TB. 1999. Technical comments on the Health Effects Institute Report entitled *Reproductive and Offspring Developmental Effects Following Maternal Inhalation Exposure to Methanol in Nonhuman Primates Part 1: Methanol Disposition and Reproductive Toxicity in Adult Females* by Thomas Burbacher, Danny Shen, Kimberly Grant, Lianne Sheppard, Doris Damian, Stephen Ellis, and Noelle Liberato. Unpublished report, 6 pp, dated 12 November 1999. Previously submitted to the Expert Panel.

Starr TB. 2000. Technical comments on determining a Safe Exposure Level (SEL) for methanol: Issues raised by baseline human blood methanol levels. Unpublished report, TBS Associates, 4 pp, dated 29 June 2000. Previously submitted to the Expert Panel.

Starr TB, Hooberman BH, Festa JL. 2000. Blood methanol pharmacokinetics reconciles interspecies responses differences and strengthens risk assessment. Final Program, Society for Risk Analysis 2000 Annual Meeting, p 99. Abstract. Copy attached.

SOCIETY FOR RISK ANALYSIS

SRA 2000
Applications of Risk Analysis in
Industry and Government



The Society for Risk Analysis Annual Meeting will highlight the changing nature of risk, the global and transboundary risk issues, new approaches to risk management and trends in public values and democratic processes to be expected in the coming century and will take place at the Crystal Gateway Marriott in Arlington, Virginia.

FINAL PROGRAM



ANNUAL MEETING
CRYSTAL GATEWAY MARRIOTT
DECEMBER 3-6, 2000

Monday, December 4

1:30 - 3:00 pm

Room: Arlington Ballroom, Salon 1

M11 Risk Management: Siting and Clean-up

Chair: Alvin Greenberg

- 1:30 M11.01 Development of a Set of Risk-Based Comparison Benchmarks as an Alternative to Existing State and Federal Cleanup Goals. C. Williams, R. Freeman, A. Whinnin, K. Frautzen, Kenesha Hengen Braslin, Inc., and Ecology & Environment, Inc.
- 1:50 M11.02 The Organizational Amplification and Attenuation of Risk. S. Emami, Parkton Clinic
- 2:10 M11.03 Managing Research to Evaluate if a Risk is Acceptable: with Application to Nuclear Waste Siting. R. V. Brown, George Mason University
- 2:30 M11.04 Health Risks Associated with Early Transfer of Contaminated Properties - Part of Oakland Case Study. M. J. O'Brien, A. A. Keller, L. Everett, D. H. Bousard, W. E. Keulenberger, University of California at Santa Barbara, Lawrence Berkeley National Laboratory, and University of California, Berkeley

1:30 - 3:00 pm

Room: Arlington Ballroom, Salon 2

M12 PBRF - Poster Platform

Chair: Steve Sivoni

- M12.01 Using Human Dose in Risk Assessment to Protect Public Health. M. E. Anderson, M. L. Longstre, L. S. Edrington, J. A. MacGregor, C. J. Fowler, Colorado State University, 1984, Environmental Toxicology Consulting Services, and NCHS

1:30 - 3:00 pm

Room: Grand Ballroom, Salon A

M13 Risk-Based Methodologies in Decision-Making and Evaluation

Chair: Scott Farrow

- 1:30 M13.01 Design and Application of the Fast Environmental Regulatory Evaluation Tool: FERET. S. Farrow, E. Wang, R. Ponce, R. Zerba, E. Passolunghi, Carnegie Mellon University, and University of Washington
- 1:50 M13.02 The Risk, Success and Failure of New Technology. R. B. Ward, University of Technology, Australia
- 2:10 M13.03 Risk-Based Methodology for the Characterization of Support for Operations Other than War. M. J. Dombrowski, Y. Haimov, J. Lambert, K. Schlessel, M. Subashi, University of Virginia, and The National Ground Intelligence Center
- 2:30 M13.04 Diagnosis Tools for Fate and Exposure Modeling in Life Cycle Toxicity Assessment. O. Joffe, M. Jorgin, D. W. Pennington, Swiss Federal Institute of Technology, Switzerland

1:30 - 3:00 pm

Room: Grand Ballroom, Salon B

M14 Symposium: Effects of Industry Restructuring on High-Hazard Operations

Chair: Vicki Bier

- 1:30 M14.01 Industrial Restructuring and Its Effects on Corporate Risk Management. R. E. Kasperow, J. X. Kasperow, D. Goldring, Clark University
- 1:50 M14.02 The Net Present Value of US Nuclear Power Plants Under License Renewal and Life Extension. G. S. Rohlfess, Stanford University
- 2:10 M14.03 Effects of Electricity Deregulation on Nuclear Power Safety. V. M. Bier, J. K. Jostes, J. D. Geyer, J. A. Tracy, M. P. Webb, University of Wisconsin-Madison, Conquest USA, and Christensen Associates
- 2:30 M14.04 Types of Risk Escalation in Nuclear Power Plant Operations as Consequences of Deregulation and Market Uncertainties. C. Penn, Massachusetts Institute of Technology

1:30 - 3:00 pm

Room: Grand Ballroom, Salons F/G

M15 Symposium: Integrating Economics and Risk Assessment - I

Co-Chairs: Clare Narnod and Christina McLaughlin

- 1:30 M15.01 Integrating Risk and Economic Information to Better Inform Risk Management Decision-Making. J. Blake-Hedges, J. Billy Riker, US Environmental Protection Agency, Washington, DC and Environment Canada
 - 1:50 M15.02 Evaluating the Economic Effectiveness of Pathogen Reduction Technology in Beef Slaughter Plants. S. Malcolm, C. Narnod, M. Offinger, T. Roberts, University of Delaware, and US Department of Agriculture
 - 2:10 M15.03 Neumann-Morgenstern Utility Theory and Pareto-Optimality Criteria Applied to Sustainable Intervention Strategies Using a Risk Assessment Model. E. Ebel, W. D. Schaefer, US Department of Agriculture, Colorado
 - 2:30 M15.04 Determining the Health Benefits of FDA Inspections of Seafood. C. P. McLaughlin, C. Narnod, US Food and Drug Administration, Washington, DC
- Discussants: Sue Fesenc and Roberta Nozales

M12.08 Starr, T.A., Froelicherman, D.H., Fests, J.L., THIS ASSOCIATES, ENVIRONMENTAL CORPORATION, and American Forest & Paper Association. BLOOD METHANOL PHARMACOKINETICS RECOGNIZES INTERSPECIES RESPONSE DIFFERENCES AND STRENGTHENS RISK ASSESSMENT.

CD-1 mice appear most sensitive to methanol toxicity, which manifests as cervical rib malformations in the offspring of pregnant dams exposed to airborne methanol concentrations of 2,000 ppm or higher 7 h/day on days 6-15 of gestation (Rogers et al., 1993). Across species, toxicity appears to arise only after methanol metabolite salivates at high airborne methanol concentrations, leading to disproportionately higher blood methanol levels. When two pharmacokinetic models of methanol uptake and disposition (Perkins et al., 1995 and Thurton et al., 1992) were analyzed to determine initial rates of blood methanol accumulation and steady state blood methanol concentrations for mice, rats, monkeys, and humans, a remarkable concordance across these species was revealed in the steady state blood methanol concentrations resulting from continuous exposure at a proposed Reference Concentration (RC) of 64 ppm. Initial rates of accumulation were also in close agreement except in the mouse, whose initial rate of accommodation was 6- to 15-fold higher than those for the rat, monkey, and human, indicating that the mouse achieves steady state more rapidly than do the other species. Further, the steady state blood methanol concentration in mice exposed to the proposed RC of 1,000 ppm is 406.8 mg/L, over 254-fold higher than that of humans (1.6 mg/L) exposed to the proposed RC. For comparison, baseline human blood methanol levels (arising from normal metabolism of some dietary constituents) average about 1.8 mg/L (Lee et al., 1992). Thus, accounting for species differences in blood methanol pharmacokinetics at high airborne concentrations raises the margin of exposure at the proposed RC to over 254-fold, and strengthens confidence that no adverse effects will arise from continuous human exposure at that level.

P2.24 Stuedy, Jane, ARJUNIS Geraghty & Willey, Inc., Research Triangle Park, NC, and Jarabek, Janie, U.S. EPA, National Center for Environmental Assessment, Research Triangle Park, NC. BARRIERS AND BRIDGES TO INTEGRATING HEALTH AND ECOLOGICAL RISK ASSESSMENT: A WORKSHOP

Environmental managers must make decisions that are protective of human health and the environment. These decisions should be, but rarely are, based upon information that integrates both health and ecological risk assessments. A workshop was held to bring together scientists practicing in the areas of health and ecological risk assessment to explore similarities and differences in their respective approaches, enhance the dialogue between these two practices, and explore tools that can promote more integrated risk assessments. The workshop was held in October of 2000, and was jointly sponsored by the Research Triangle Chapter of the Society for Risk Analysis (SRA) and the Carolina Chapter of the Society of Environmental Toxicology and Chemistry (SETC). Participants represented state and federal government, academia, industry, and consulting firms.

The workshop began by providing participants with an overview of the "state of the science" in both health risk assessment and ecological risk assessment. This was followed by presentations of case studies that highlight the ecological and human health approaches in two different contexts: chemical-specific versus site-specific risk assessments. The next presentation discussed a new "framework for the integration of health and Ecological Risk Assessment." A series of presentations and discussions followed to allow a more detailed examination of specific topics, with one presenter discussing the ecological risk perspective and the other discussing the health risk perspective. These topics included fate and transport/bioavailability/dose/matrix, scale/level of analysis, indicators/adverse effects, extrapolation, uncertainty, and risk communication. The workshop concluded with a panel discussion based on questions submitted by the attendees. This poster will present observations on the process as well as summarize the key issues for the individual disciplines and challenge of integration.

T2.6.04 Stuebel, C., Kulbowski, A., and Stuebel, P.J.: Minnesota Department of Health. A PUBLIC HEALTH APPROACH TO ASSESSING CUMULATIVE RISKS: MINNESOTA CHILDREN'S PESTICIDE EXPOSURE STUDY

The Minnesota Department of Health conducted a preliminary analysis of health risks from chemical exposures measured in the Minnesota Children's Pesticide Exposure Study. Concurrent measurements of chemicals in multiple media, including air, water, food, beverages, soil, and dust were used to evaluate approaches for assessing cumulative risks. Toxicity values from the US EPA IRIS database, child-appropriate values from the US EPA Exposure Factors Handbook, and intake rates and body weights from the study population were used to estimate health risks. Cumulative risks were calculated by: (a) adding risks across pathways; (b) adding risks for chemicals with a common mechanism of toxicity; and (c) adding risks for chemicals with a common critical effect. Results of this preliminary analysis will be used to identify chemicals and pathways of potential health concern; to identify areas where further data analyses are needed; and to evaluate the limitations and data gaps of this unique data set for conducting cumulative risk assessment. [*Work supported by the Legislative Commission on Minnesota Resources. Data provided by the Minnesota Children's Pesticide Exposure Study, a substudy of the National Human Exposure Assessment Survey.]

T1.6.02 Subramanian, R.P., Golden, S.L., Kral, P., Thurin, J., and Anderson, E.L., Sciences International, Inc. AN EXPLORATORY STUDY OF VARIATIONS IN ENVIRONMENTAL TOBACCO SMOKE EXPOSURE IN THE UNITED STATES

There has been considerable interest in comparing workplace exposure to Environmental Tobacco Smoke (ETS) with exposures at other settings. Jenkins et al. (*J. Exposure Anal. Environ. Epidemiol.*, 6, 473-502, 1996) used personal monitors to measure ETS exposures of 1564 non-smokers in 16 US cities and compared exposure in the workplace with exposure away-from-work. In conjunction with exposure data, extensive diary and questionnaire information were also gathered. We probe these data further to examine 1) correlation between work and away-from-work exposures to ETS, 2) the variability in exposure patterns across cities, and 3) the association of ETS exposure with socio-economic, occupation and lifestyle variables. Our findings are that 1) at the population level, there is a positive association between exposures to ETS at the work and away from work environments, 2) exposure patterns across the 16 cities are highly variable, 3) occupation, education, and household income play a significant role in determining exposure levels, and 4) smoking restrictions at work are associated with lower exposure to ETS both at work as well as away-from-work environments. In addition, age and dietary factors have associations with exposure levels. Generally, the same cities that exhibit lower (or higher) away-from-work exposure levels also show lower (or higher) work exposure levels. Our observations suggest the hypothesis that avoidance characteristics, and socioeconomic and lifestyle factors that affect a population's exposure to ETS may be similar in the away-from-work and work environments. The results of this analysis may be used to better target specific sections of the population to reduce health risks associated with ETS exposure. [Work sponsored by the Center for Indoor Air Research, Lindholm, ND]

W9.02 Susarla, A., Clark University. CAN WHISTLE BLOWING SHANE RISK COMMUNICATION? : AN ANALYSIS OF ORGANIZATIONAL RESPONSES AND THEIR ROLE IN SOCIAL AMPLIFICATION AND ATTENUATION OF RISK IN INDIA

Whistle blowing enriches risk communication and presents fresh perspective in public domain. On the other hand, in response to whistle blowing, hazard control organizations can choose among multiple pathways. What effect does Whistle-blowing have on flow of risk information? Review of literature indicates that outcomes to whistle blowing is guided by numerous conditioning variables (e.g. bureaucracy in organizations), however complete social process of whistle blowing is yet to be revealed. Further, most literature described examples chosen from settings of first-world countries. This paper presents findings of an in-depth inquiry of blowing the whistle and responses by hazard managers in two events that occurred in India. The selected hazards are the episode of pneumonic plague in Surat City in 1994 and Asenic pollution of groundwater in West Bengal. Analyses show that whistleblowers and organizations adopted different strategies in each hazard event. The findings denote that the relationship between whistle-blowing initiatives

Comments on the Draft NTP-CERHR Expert Panel Report
on the Reproductive and Developmental Toxicity of Methanol

David G. Hoel, Ph.D.
Department of Biostatistics
Medical University of South Carolina
Charleston SC

on behalf of the
American Forest & Paper Association

7 September 2001

Introduction

The Committee generally provided a good overview and interpretation of the Burbacher *et al.* study (references 86 and 93 in the CERHR Draft report). A number of biostatistical issues, however, were not adequately addressed. These include: 1) the problem of frequent false positives arising from the multiple statistical comparisons that were undertaken by the investigators; 2) the high sensitivity of inferences regarding effects of methanol exposure on pregnancy duration to a single outlier animal in the control group; and 3) the problem of small numbers and high inter-animal variability, particularly as they impact findings for the Visually Directed Reaching Test and the Fagan Test using social stimuli; and 4) the failure to adequately address the body of non-human primate findings regarding methanol as a coherent whole. These are discussed in more detail below.

1. The Multiple Comparisons Problem ***General Considerations***

On page 64 of the Draft Expert Panel document, there is the recognition that there are a large number of statistical comparisons made in the Burbacher *et al.* analysis without any adjustment of the associated p-value (always taken to be 0.05) for these “multiple comparisons”. One can simply count up the number of Part II Developmental Effects tests as follows.

There are 12 independent physical measurements and 14 behavioral measures each with 5 statistical tests (an overall ANOVA plus 4 contrasts) giving a total of 130 statistical comparisons. There were 5 additional behavioral measurements with 4 statistical contrasts each for an additional 20 tests giving a total of 150 statistical tests. Since testing often was also carried out for both cohorts individually as well as for each sex, this number of 150 statistical tests should be further multiplied by a factor of 4 to yield a total of 600 separate analyses. Clearly, one would expect to see a substantial number of apparently positive (i.e., false positive) results using the Burbacher *et al.* method of analysis even when there is no effect of treatment whatsoever on the measured variables.

This is not a new problem, although it is a serious one. It is crucial that experiments be designed and analyzed in such a way as to minimize the chances of falsely concluding that effects are present when they are in fact absent. If this is not done, it becomes impossible to differentiate real effects from those due solely to chance as a consequence of the existence of sampling variability. Statisticians have addressed this problem in detail in many different contexts. Indeed, there are standard and well-accepted methods for the analysis of animal toxicology data. Unfortunately, these methods were either not known to Burbacher *et al.*, or they were ignored by these researchers in their analyses.

There are basically two issues involved with the data analysis of an experiment such as the Burbacher *et al.* study. First there is the issue of multiple comparisons of individual dose groups with a control group for a single experimental outcome. The second is that of the repeated testing of a variety of experimental outcomes as mentioned above.

Single Endpoint Analyses: Treated Groups vs. a Single Control Group

Statistical methods were developed many years ago for the analysis of a single experimental outcome in which a control group is compared with several experimental groups. The National Toxicology Program's Cancer Bioassay Reports reference many of these statistical methods in each Report's Materials and Methods section.

If the data are normally distributed, Williams (Biometrics 27:103-17, 1971 & 28:519-31, 1972) provides a method for the multiple comparison testing of a control group with several different groups ordered by dose. Furthermore, Williams also provides a method for determining the lowest dose group for which there is an identifiable experimental effect.

If the data are not normally distributed, nonparametric methods are required, and the paper of Shirley (Biometrics 33:386-89, 1977) provides the testing methods equivalent to those of Williams for this non-normal situation. If the question is whether there is a significant monotonic trend in the strength of the effect as a function increasing dose, Jonckheere's test (Biometrika 41:133-45, 1954) for non-normal data can be applied.

The above-mentioned techniques effectively adjust for the comparison of multiple dose groups with a single control group while presuming that the ordering of exposure across the groups will produce a similar ordering in the responses of the groups to exposure. Dunn (Technometrics 6:241-52, 1964) and Dunnett (JASA 50:1096-1121, 1955) provide methods for comparing a control group with experimental groups that are not assumed to have an ordered or graded relationship to one another (e.g., not monotonic), as might occur, for example, when qualitatively different drugs are being tested against a single placebo control.

The simplistic statistical methods (i.e. t-test) for pairwise group comparisons used in the Burbacher *et al.* analyses are simply not appropriate. As is discussed further below, use of such methods leads to far more false positives than would be permitted by an adequate experimental design.

Analyses of Multiple Endpoints

The problem of testing multiple, presumably independent, experimental outcomes is comparatively easy to appreciate. If we evaluate 100 different, independent experimental endpoints at a nominal false positive rate of $\alpha=0.05$, then we should expect to observe, on average, $0.05 * 100 = 5$ positive results even when there is no true, underlying effect of exposure on outcome. To control the experiment-wide false positive rate, one usually employs a lower nominal alpha level, such as that provided by the Bonferroni correction ($\alpha=0.05/n$, where n is the number of endpoints being evaluated), or one simply keeps the high overall false positive rate from the multiple endpoints in mind when interpreting the significance of the experimental results. This translates into a need to observe far smaller

p-values for individual comparisons as well as clear-cut dose-responses before an observed effect is taken to be anything more than suggestive. While this latter approach is a commonly used option in routine toxicology testing, formal and strict control of the experiment-wide false positive rate would be preferable.

To better appreciate the scope of this multiple testing problem in the Burbacher *et al.* study, consider Table A which shows the number of statistical tests carried out over the 31 experiments. The number of primary statistical tests (wherein the data were not stratified by gender or cohort) comes to 150. Thus we expected to see $0.05 * 150 = 7.5$ positive results even when there is no underlying effect of treatment, and actually only 4 significant differences (at the $p < 0.05$ level) were observed. Since the positive findings were fairly scattered (see Table B), it appears that there is no true exposure effect. At the very least, the effects that were seen cannot be confidently differentiated from those that would be expected to arise by chance due solely to sampling variability.

Table A
Summary of Statistical Testing
Part II: Developmental Effects, Primary Analyses

	Number of Experiments	Statistical Tests	Expected Positives	Observed Positives
Physical Measures				
ANOVA plus 4 contrasts	12	60	3	0
Behavioral Measures				
ANOVA plus 4 contrasts	14	70	3.5	4*
4 contrasts only	5	20	1	0
Total	31	150	7.5	4

*: A9 (control vs. 600 ppm & control vs. total exposed),
 A13 (control vs. 1800 ppm) & A26 (control vs. 600 ppm).

Note: linear dose response testing reported only for A13 ($p=0.04$), A15 ($p=0.08$)

Table B below shows that a number of secondary analyses were also reported, even though a positive result was not necessarily observed in the primary analysis (e.g., A17). Also, both gender and cohort stratifications were considered. Thus for an experiment without “linear trend testing” we have, besides the ANOVA, four primary contrasts and 16 secondary contrasts (8 cohort and 8 gender) for a total of 20 statistical contrasts. With 14 behavioral measures we then have 280 tested contrasts. With a nominal $p = 0.05$ false positive rate, we expect $0.05 * 280 = 14$ statistically significant contrasts by chance alone, which is very close to the 18 actually reported. With 280 independent contrasts tested, in order to achieve an overall experiment-wide error rate of 5% one must carry-out the individual statistical tests using a p-value of 0.0002 or less, which generally was not achieved.

Table B
Specific Statistically Significant Results:
Primary and Secondary Analyses

Experiment	ANOVA	Sum*	200ppm	600ppm	1800ppm	Linear	Secondary**
A9 ^a	n.s.	0.03	n.s.	0.01	n.s.		none
A13 ^b	n.s.	n.s.	n.s.	n.s.	0.04	0.04	m/f ¹
A15 ^b	n.s.	n.s.	n.s.	n.s.	n.s.	0.08	m/f ²
A17 ^c	n.s.	n.s.	n.s.	n.s.	n.s.		m/f ³
A26 ^d	n.s.	n.s.	n.s.	0.03	n.s.		cohort ⁴

* Sum is control group contrasted with the combined treated groups.

** Secondary Analyses means a new analysis incorporating additional factors such as gender is reported which presumably is motivated by the results of the original analysis.

^a Neonatal behavioral scale: Behavior state factor

^b Visually directed reaching: vs. age (A13), vs. gestational length (A15)

^c Observations of motor milestones in playroom

^d Recognition memory assessment

¹ Males: Control v. 600 ppm p=0.007, Control v. 1800 ppm p=0.03; Females: n.s.

² Males: Control v. 600 ppm p=0.04, Control v. 1800 ppm p=0.04; Females: n.s.

³ Males: Control v. 600 ppm p=0.02; Female: Control v. 200 ppm p=0.01, Control v. 600 ppm p=0.004, Control v. Total exposed p=0.008.

⁴ Cohort 1: n.s.; Cohort 2: Control v. 200 ppm p=0.04, Control v. 600 ppm p=0.0001, Control v. 1800 ppm p=0.03, Control v. Total exposed p=0.002.

Note: n.s. = p > 0.05

When significant results among the numerous endpoints tested are being reported at the p=0.05 level, one needs to observe a significant dose-response trend in an outcome variable and/or replication of the observed effect over both cohorts or over both genders to be convinced that the reported effect is not simply a false positive finding.

In summary, the Burbacher *et al.* data provide a good example of how a large number of statistical tests can produce a few inconsistent, but entirely expected, positive results even when the experiment is truly negative. Based on the sheer number of statistical tests that were employed by Burbacher *et al.* and their failure to adequately control the experiment-wide false positive error rate, we are forced to conclude that there is no convincing evidence for an effect of methanol exposure on the behavioral measures evaluated in these primates.

2. The Sensitivity of Inferences Regarding Pregnancy Duration to a Control Group Outlier

On page 65 of the Draft Expert Panel report, one Panel member discusses the “statistically significant (p=0.03) decrease in the duration of pregnancy”.

The historical range for the duration of pregnancy indicates that the one control animal with “signs of postmaturity” and a duration time of 178 days unduly influenced the

analysis (also see comments by AF Tarantal, reference 94 in the Draft report). This animal's duration value of 178 days is more than 2 standard deviations beyond the observed mean of 167 days for the current control group. Inclusion of this outlier in the analysis of pregnancy duration leads to an invalid inference that the exposed groups' pregnancy durations were significantly shortened by methanol exposure. Excluding this outlier from the control group changes the p value of the test from a marginally significant value of 0.03 to a non-significant value of 0.1. It must also be kept in mind that there was no dose response for pregnancy duration, which lends further support to the view that the control group was high in pregnancy duration by chance alone, and that there was no effect of methanol exposure on duration of pregnancy in the exposed groups.

3. Small Numbers and High Inter-Animal Variability: Comments on Results from Two Specific Neurobehavioral Tests

In addition to the multiple comparison problems discussed in Section 1, the following specific points are made with regard to two of the behavioral measures that were mentioned in the draft Expert Panel Report.

The Visually Directed Reaching Test (p. 63) showed that, among males, all exposure groups had lower performance than controls. Two obvious points needing serious consideration are that 1) the result does not hold for females and 2) if the control male with the 178 day gestation is deleted from the analysis only two males remain which is inadequate for an analysis based upon normal distribution theory. Under these circumstances, very little confidence can be placed in this finding.

The second test mentioned is the Fagan Test using Social Stimuli. An important consideration here is that the individual animal values are averages of highly variable values from multiple tests, with a substantial number of missing values for some tests and some animals. Secondly, the results of the experimental groups did not differ significantly from one another; the result hinges on the fact that the control group was the only group to prefer novel social stimuli over common ones significantly more than a specific *a priori* criterion of 50%, i.e., no preference at all.

4. The Non-Human Primate are Findings Consistent Across Studies: No Methanol-Related Effects

The Expert Panel draft document reviewed another non-human primate study (Reynolds *et al.* (reference 97) and Suomi (reference 98) discussed on pp. 66-67). A total of 20 animals were used (5 experimental groups with 4 animals per group). As with the Burbacher *et al.* study, the findings in this study were negative. The Panel felt that the study was well done but did not achieve an MTD for methanol. The Panel also noted that the power of the study was not determined. However, the same remarks apply to the Burbacher *et al.* study. The power calculations can always be carried out, after the fact, for both studies. What is important, however, is that the null findings in this study are consistent with the Burbacher *et al.* study's finding of no methanol-related effects.

5. Concluding Remarks

In summary, I would conclude that the Burbacher *et al.* study was well-conducted, but it was not statistically analyzed as it should have been. Based upon the large number of statistical comparisons made in this study, the few scattered treatment-related findings are no more than one would expect by chance; the study is about as negative as a study with so many statistical tests can be. There is no clear-cut association between methanol exposure and the outcomes studied. The negative study by Reynolds *et al.* and Suomi lends further support to this conclusion.

To: John Festa, Tom Starr

From: David Hoel

Date: October 13, 2000

Re: Burbacher Report of Methanol tests on *Macaca fascicularis*

Specific Statistical Issues of the Data Analysis

1) **Duration of Pregnancy:** There were a very large number of pregnancy outcome measures with their corresponding statistical tests given in the report. The results were negative with respect to methanol effects with the exception of duration of pregnancy. What has to be kept in mind is that because of the large number of statistical test carried out one expects to find some statistically significant outcomes without there being a true methanol effect. The report on duration of pregnancy stated that the ANOVA test for differences in dose groups reported a p-value of 0.03 and p-values for paired contrasts of controls versus exposure groups of 0.005, 0.04, 0.02 for the 200ppm, 600ppm and 1800ppm exposure groups, respectively.

The first observation to make with this data on pregnancy duration is that there was one animal in the study (control group #M93327) which was an outlier in that it showed signs of postmaturity and had a duration of pregnancy well beyond the normal range. Removing this animal from the analysis resulted in an ANOVA with a p-value of 0.1 which is non-significant. Secondly, if one were to also remove the one premature animal and/or add a few days to the duration values of the c-section animals the difference between exposure groups becomes even smaller and even less significant.

With the removal of the mature animal from the analysis the individual pairwise comparisons also changed. The only comparison with a reported p-value less than 0.05 is the control v. 200ppm group with a p-value of 0.017. With the multiple comparisons the appropriate p-value to use for achieving statistical significance is 0.013 making this particular comparison non-significant. This adjustment for multiple comparisons will give an experiment wise error rate of 0.05 based on 4 contrasts (c v. 200, 600, 1800 and total exposed).

Finally it should be observed that there is no dose response with this data with respect to duration of pregnancy. The control values appear to be slightly larger than the exposed group values which is likely due to random variations.

2) The Effects of Multiple Comparisons on Specific Tests:

A) Results for "Recognition Memory Assessment using Abstract Patterns" (Table 15) reported statistically significant differences between control animals and exposed animals. The data were based on a mean of 7 abstract task tests that were given in three series of 3, 2 and 2 problems respectively. There are several problems with the summary value used for each animal. First and most importantly is that for some animals there were missing test scores due to tester errors or animal behavior. Since these 7 tests are not

replications and were given over a period of time the calculation of a summary or mean value is not correct. Secondly it is not clear how the summary values were obtained from the individual 7 test scores. Finally the individual test scores are given as percentages and their mean values are treated as normal variates for analysis purposes. With presumably binomial data other analysis methods which recognize the type and structure of the data are more appropriate.

The ANOVA showed no difference between experimental groups ($p=0.18$) but the linear contrasts reported a $p=0.03$ for control v. 600 ppm. As stated previously the multiple comparison issue would require a p -value of 0.013 for an experiment wise error rate of 0.05. Therefore there is no significant difference between controls and experimental animals with regard to this measure. The investigators both here and for another outcome measure considered the subsets of gender or cohort for further comparisons. As one digs deeper the nominal p -value is 0.0026 (20 contrasts tested) for an experiment wise error rate of 0.05. Table A27 reports $p=0.0002$ for control v 600 ppm and $p=0.002$ for control v. total exposed for cohort 2 animals that are significant values. However, no such effect is observed for cohort 1 and there were only 3 control animals in cohort 2.

Now if one carries out an ANOVA test for each of the 7 individual tests the p -values are 0.24, 0.53, 0.83, 0.07, 0.38, 0.53, 0.47. Thus none are individually significant and there is no problem with missing values. If we restrict ourselves to cohort 2 and remove the animals with missing values the mean summary values are 0.66, 0.61, 0.52 and 0.59 for the control, 200ppm, 600ppm and 1800ppm groups, respectively. The problem is that although the control value is larger it is based on only 2 animals and only 3 each in the three exposed groups.

Finally it should be repeated that if one uses the binomial data and not just the proportions the data variability can be properly assessed while the currently used normality methods likely under estimate the true variability and thus overstate the group differences.

B) "The Visually Directed Reaching Test" (Tables 12, A13, A14). As with the previous test multiple comparisons are made with gender instead of cohort being used to show a dose difference in part of the animal study. Based on the adjusted p -value for the multiple comparisons no contrast is statistically significant for this endpoint. It should also be noted that the "mature at birth" animal (M93327) had the lowest value (17 days) and if removed changes the average male control value from 23.7 to 27 leaving only 2 control males.

C) "The Motor Milestones in the Playroom" (Table 13, A18) gave results based on females that when adjusted for multiple comparisons were no longer significant. The results were not seen in the males and the female effect goes in the opposite direction to what one would expect. (i.e. exposure appeared to be beneficial).

D) "Recognition Memory Assessment Social Stimuli -Fagan Test" (Table 16, A28-29). There were not differences between the exposure groups but only the control animals were significantly better than random. There are questions about how one can combine 7 tests to produce a summary value when there are missing observations. If the animals with missing values are deleted then the control mean drops from 0.62 to 0.57 and the outcome likely loses its

statistical significance.

3) Experiment wise Error Rates:

Summary of Statistical Testing Part II: Developmental Effects, Primary Analyses

	Number of Experiment	Statistical Tests	Expected Positives	Observed Positives
Physical Measures				
ANOVA plus 4 contrasts	12	60	3	0
Behavioral Measures				
ANOVA plus 4 contrasts	14	70	3.5	4*
4 contrasts only	5	20	1	0
Total Primary Analysis Tests	31	150	7.5	4

*A9 (600ppm & total), A13 (1800ppm) and A26 (600ppm).

Note: linear dose response testing reported only for A13 (p=0.043, A15 (p=0.08)

Specific Statistically Significant Results: Primary and Secondary Analyses

Experiment	ANOVA	Sum*	200ppm	600ppm	1800ppm	Linear	Secondary**
A19 ^a	n.s.	0.03	n.s.	0.01	n.s.		none
A13 ^b	n.s.	n.s.	n.s.	n.s.	0.04	0.04	m/f
A15 ^b	n.s.	n.s.	n.s.	n.s.	n.s.	0.08	m/f
A17 ^c	n.s.	n.s.	n.s.	n.s.	n.s.		m/f
A26 ^d	n.s.	n.s.	n.s.	0.03	n.s.		cohort

*Sum is control group contrasted with the combined treated groups.

**Secondary Analyses means a new analysis incorporating additional factors such as gender is reported which presumably is motivated by the results of the original analysis.

^aNeonatal behavioral scale: Behavior state factor

^bVisually directed reaching: vs age (A13), vs gestational length (A15)

^cObservations of motor milestones in playroom

^dRecognition memory assessment

Note: n.s. = p > 0.05

The above tables show that a number of secondary analyses were reported even though a positive result was not observed in the primary analysis (i.e. A17). Also both gender and cohort divisions were considered. Thus for a test without “linear trend testing” we have besides the ANOVA, four primary contrasts and 16 secondary contrasts (8 cohort and 8 gender) for a total of 20 contrasts. With 14 behavioral measures we then have 280 tested contrasts. With a nominal 5% type one error we expect 14 statistical significant contrasts which is close to the 18 actually reported. Of the 18 reported effects 3 were in the direction of a beneficial effect of methanol (A17). (It should be noted that A13 and A15 both measure the same outcome but A13 uses age and A15 uses gestational age. Thus we really have 260 contrasts with 13 expected positives and

15 observed positives.) With 280 contrasts tested in order to achieve an experimental wide error rate of 5% one must test at the $p=0.0002$ significance level in the individual tests. Only one individual contrast (see 2A above) achieved this level of significance.

With so much testing on so few animals one needs to observe dose response in an outcome variable and/or replication over cohort or gender when positive results are being reported at these levels of comparison. These data are a good example of a large number of statistical tests producing a few inconsistent but expected positive results for experiments that are basically negative. That is there is no convincing evidence that methanol has an effect on the behavioral measures in these monkeys.