PEER REVIEW SUMMARY REPORT

External Peer Review Meeting on the *Toxicological Review of Beryllium and Compounds (CAS No. 7440-41-7)*

Prepared for:
Amanda Persad, Ph.D., D.A.B.T.
U.S. Environmental Protection Agency
National Center for Environmental Assessment
109 T.W. Alexander Dr.,
Research Triangle Park, NC 27709

Prepared by:
Versar, Inc.
Contract No. EP-C-07-025
Task Order 27

Peer Reviewers:
Patrick N. Breysse, Ph.D., MHS
Herman J. Gibb, Ph.D., MPH
David Kriebel, Sc.D.
Lee S. Newman, M.D., M.A.
Kyle Steenland, Ph.D.

August 4, 2008

The Integrated Risk Information System (IRIS) is an EPA data base containing Agency consensus scientific positions on potential adverse human health effects that may result from chronic (or lifetime) exposure, or in select cases less-than-lifetime exposures, to chemicals in the environment. IRIS currently provides health effects information on over 500 chemical substances.

IRIS contains chemical-specific summaries of qualitative and quantitative health information in support of two steps of the risk assessment process, i.e., hazard identification and dose-response evaluation. IRIS information includes a reference dose (RfD) for non-cancer health effects resulting from oral exposure, a reference concentration (RfC) for non-cancer health effects resulting from inhalation exposure, and an assessment of carcinogenicity for both oral and inhalation exposures. Combined with specific situational exposure assessment information, the health hazard information in IRIS may be used as a source in evaluating potential public health risks from environmental contaminants.

The IRIS program, within EPA's National Center for Environmental Assessment (NCEA), has previously developed a Toxicological Review of Beryllium and Compounds, an assessment which is currently available on the IRIS database. At that time the cancer assessment remained open pending the publication of new carcinogenicity data. Though the new draft contains a chronic oral reference dose, a chronic inhalation reference concentration, and a quantitative cancer assessment, only the qualitative cancer assessment is slated for external peer review.

Peer Reviewers:

Patrick N. Breysse, Ph.D., MHS
The Johns Hopkins University
Department of Environmental Health Sciences
Baltimore, MD 21205

Herman J. Gibb, Ph.D., MPH (Chair)
Sciences International, Inc.
Alexandria, VA 22314

David Kriebel, Sc.D.
University of Massachusetts Lowell
School of Health and Environment
Lowell, MA 01854

Lee S. Newman, M.D., M.A.
University of Colorado at Denver School of Medicine
Denver, CO 80262

Kyle Steenland, Ph.D.
Emory University
Rollins School of Public Health
Atlanta, GA 30322

II. CHARGE TO THE REVIEWERS

The U.S. Environmental Protection Agency (EPA) is seeking an external peer review of the scientific basis supporting the human health assessment of beryllium that will appear on the Agency's online database, the Integrated Risk Information System (IRIS). There is currently an assessment on the IRIS database for the health effects associated with beryllium exposure.

The assessment for beryllium currently on the IRIS database was completed in 1998. However, the development of an Inhalation Unit Risk (IUR) for cancer was deferred until publication of a NIOSH epidemiologic study, subsequently published as Sanderson et al. (2001). This draft assessment contains the updated cancer assessment. The derivations of the Reference Dose (RfD) and Reference Concentration (RfC) are unchanged and are not subjects of this review. The text associated with the updated cancer assessment, which is the subject of this review, is highlighted. Note that no changes have been made to the IUR for cancer, however, and the existing IUR has been retained

Peer review of the updated sections of this assessment is being sought to ensure that all available data relevant to the qualitative descriptor of the cancer assessment of beryllium have been appropriately and objectively evaluated. Below is a set of charge questions that address scientific issues in the cancer assessment of beryllium. Please provide detailed explanations for responses to the charge questions.

General Charge Questions:

1. Are the updated sections of the Toxicological Review logical, clear and concise? Has EPA accurately, clearly and objectively represented and synthesized the scientific evidence for cancer hazard?

2. Please identify any additional studies that should be considered in the assessment of the cancer health effects of beryllium.

3. Please discuss research that you think would be likely to reduce uncertainty in future assessments of beryllium.

Chemical-Specific Charge Questions:

1. Under the EPA's 2005 *Guidelines for carcinogen risk assessment* (www.epa.gov/iris/backgr-d.htm) (Section 2.5), beryllium via inhalation exposure is classified along a continuum between *likely to be carcinogenic to humans* and *carcinogenic to humans*. Please comment on the scientific justification for the cancer weight of the evidence characterization for this exposure route. Has the scientific justification for the weight of evidence characterization been sufficiently, transparently and objectively described?

2. EPA has determined that the literature published since the 1998 IRIS assessment is inadequate to support a reassessment of the cancer inhalation unit risk (please refer to Appendix C of draft assessment). Please comment on EPA's rationale for not deriving an updated inhalation unit risk. Please identify any currently available studies or methodologies that could be used to derive an inhalation unit risk (IUR). Please comment on EPA's decision to retain the current IUR values.

3. Given that EPA was not able to update the inhalation unit risk factor and that NIOSH is in the process of updating its cohort analysis (both extending the follow up time by 13 years and adding two facilities with more recent exposure levels) that should prove valuable in updating the IUR, do you recommend placing the update of this Toxicological Review on hold until these data have been incorporated?

III. GENERAL IMPRESSIONS

Patrick N. Breysse

General impressions and specific answers provided in this summary are limited to the portions of the EPA draft that deal with carcinogenicity of beryllium. In general, the document presents a reasonable summary of the primary epidemiology literature. While I am less familiar with the literature relating to animal and genotoxicity studies, my sense is that this literature is also reasonably summarized.

One important limitation not emphasized in this review is that the lack of detailed exposure data will likely result in non-differential exposure misclassification that may bias results towards the null. This is particularly relevant with respect to the study by Sanderson et. al. One likely explanation for the lack of a clear dose response in this study is exposure misclassification, especially at higher exposures. Exposure misclassification is likely to be a bigger issue in this study than the smoking cofounder issue. The analysis of smoking confounding by Sanderson et al. leads me to discount the importance of smoking as an important confounder across this literature, since much of the research conducted is in the same or similar workplaces.

The document does not provide adequate justification for concluding that beryllium falls along a continuum between likely to be carcinogenic to humans and carcinogenic to humans. This is a confusing and poorly justified classification. In my opinion, a conclusion that beryllium is carcinogenic to humans is more justifiable based on the consistent epidemiologic literature including the findings from the recent Sanderson et al. study, and the animal data documenting that beryllium can cause cancer in multiple animal species.

Herman J. Gibb

The logic for the argument that beryllium lies along a continuum between likely to be carcinogenic to humans and carcinogenic to humans is not supported by the text. For example, the document states that as the level of elevated risk is small with wide CIs, the level of confidence that the elevated risk observed is most likely due to beryllium exposure is low (page 83). It states that there is no underlying statistical rationale for log-transforming the data in the Sanderson et al. study (page 93), and that Levy found no elevated odds ratios for any of the non-transformed exposure metrics. The authors make no comment on the lack of dose response in the Sanderson et al. study. On page 82, the document suggests that too much emphasis may be placed on a single cohort rather than examining the range of SMRs in the different cohorts in the Ward et al. study, thereby lowering the confidence in the observed consistent elevated risk. On pages 80 and 101, the document claims that a sound causal association between human exposure and lung cancer has yet to be elucidated. The meaning of that statement is not particularly clear, but it is apparent that it does not support classifying beryllium as being on a continuum from being a likely carcinogen to a human carcinogen.

David Kriebel

I have been asked to review those sections of the document which are relevant to the qualitative cancer assessment for beryllium and compounds. These sections deal primarily with: 1) a review of the human epidemiologic evidence for lung cancer effects of beryllium; and 2) the evaluation of the evidence for carcinogenicity, described in a weight of evidence narrative.

My general impression of the review of the epidemiologic evidence is that it is insufficiently developed to provide a full picture of the overall strength of the available evidence. Several study quality issues like control of confounding and the quality of exposure data are presented in an overly simplistic way. For example, the problem of residual confounding by smoking, while of concern in several studies, was settled to the satisfaction of both the IARC and NTP committees, but is presented here as if it were a serious limitation of this body of evidence.

The second major theme of the material I was asked to review is the assignment of a weight of evidence descriptor and an accompanying weight of evidence narrative for the carcinogenicity of beryllium. I have reviewed the instructions provided by EPA on how these should be written (pages 2-49 to 2-58 in Guidelines for Carcinogen Risk Assessment EPA/630/P-03/001F, March 2005). I find the beryllium review deviates from these guidelines in several key respects. Most importantly, the draft proposes a weight of evidence descriptor for beryllium carcinogenicity by the inhalation route which states: beryllium falls along a continuum between *likely to be carcinogenic to humans* and *carcinogenic to humans*. This is an unnecessarily confusing descriptor, and I believe gives the impression that the evidence is even less certain than if the descriptor was likely to be carcinogenic. The authors seem to have misunderstood the instructions on descriptors. It is clear from the instructions that EPA intended for one descriptor or another to *always* be chosen, while acknowledging that each descriptor covers a range of levels of evidence along a continuum.

Each of these two main points is covered in more detail below.

Lee S. Newman

My review, as instructed, focused on beryllium carcinogenesis. This is a well-crafted draft document, in that regard. While I find the report to be generally clear and objective in its representation of the literature, it falls short of expectation in its conclusions, as discussed below. An additional reanalysis that was published shortly after completion of this draft should be incorporated into the next draft. While it would be desirable to see the results of additional epidemiologic studies in order to further reduce any lingering uncertainty, the weight of evidence supports the conclusion that beryllium is carcinogenic to humans. Action should not be postponed based on the rationale that better studies are being conducted and will be available in the indefinite future.

Parenthetically, although my review of this draft concerns the carcinogenicity of beryllium, one cannot help but observe that major sections regarding the non-carcinogenic effects of beryllium are out-of-date, notably including the evidence regarding exposure levels related to chronic beryllium disease risk. This may bear indirectly on the relationship between chronic beryllium disease x cancer risk discussed in this draft.

Kyle Steenland

I believe the information presented is in general accurate and clearly presented. I do not believe the conclusions are sound, as indicated below.

IV. RESPONSE TO CHARGE

(A) General Charge Questions

1. Are the updated sections of the Toxicological Review logical, clear and concise? Has EPA accurately, clearly and objectively represented and synthesized the scientific evidence for cancer hazard?

Patrick N. Breysse

There have been a number of published discussions detailing methodological concerns about the Sanderson et al. case-control study. The reanalysis by Schubaner-Berigan et al. (2008), first published on-line in (2007), addresses many of these limitations and should be included in this review. Since this paper was originally published on-line in 2007, its publication falls within the preview of this review.

Herman J. Gibb

The updated section of the Toxicological Review is clear and concise. What the document has not done is provide the logic and synthesis to arrive at the conclusion that the evidence is on a continuum between likely to be carcinogenic to humans and a human carcinogen.

Aside from this lack of justification, it is unclear why the Agency has chosen to place beryllium on a continuum between likely to be carcinogenic to humans and carcinogenic to humans. The distinction between the two descriptors is so slight that it makes the reader wonder why the cause for such ambiguity.

David Kriebel

See my comments above, and developed in more detail below. I do not find the weight of evidence narrative logical or clear.

Lee S. Newman

Yes, with a few exceptions. Notably, at the time that EPA drafted this document, one important paper had been submitted for publication but not yet published (Schubauer-Berigan et al. Occup Environ Med 2008; 65:379-383). This publication bears significantly on some of the EPA's synthesis, especially as it relates to the interpretation of the Sanderson Am J Industr Med 2001 paper, the EPA internal analysis of that data set, and possibly the EPA's conclusions regarding their ability to use the Sanderson data set to perform an IUR. That study's continuous analysis and approach to covariates and adjustment for smoking, as well as its examination of the effect of methods that avoid taking the logarithm of zero are informative. It further clarifies and reinforces the weight of evidence in the literature that beryllium is associated with lung cancer risk in humans. The section regarding mode of action could be improved by inclusion of some of the

literature that demonstrates beryllium's effects on inflammatory pathways that are also linked to mechanisms of carcinogenesis. Some examples are provided below, including evidence of the potent oxidative stress induced of beryllium.

Kyle Steenland

While the text is reasonably concise and clear (although upon occasion abbreviations are used with insufficient explanation or definition), I disagree with the synthesis.

First, I believe that EPA is mistaken in its conclusion that beryllium falls somewhere on a continuum between 'probably carcinogenic' and 'definitely carcinogenic.' While EPA guidelines may allow for such an ambiguous conclusion, common sense dictates that a public health agency should make a clear decision between these categories. It should be noted that both IARC (1993) and NTP (1999) concluded that beryllium was a definite human carcinogen, based on much the same data available to EPA now. I agree with the IARC and NTP evaluations. This is a substance for which two human cohorts (beryllium registry patients, (Steenland and Ward 1992 JCNI) and a large worker cohorts at seven plants (Ward et al. 1992 AJIM)) both indicate excess lung cancer mortality (SMRs of 2.0 and 1.5 (long latency group), respectively), and for which there is supporting animal data for inhalation.

Second, I believe a risk assessment should be conducted (i.e., an inhalation IUR should be calculated) using average exposure lagged 10 years in the recent re-analysis of Sanderson et al. (2001) by Schubauer-Berigan et al. (2008). These new 2008 analyses control for date of birth, which is confounder not controlled in the original analyses by Sanderson et al. (2001).

Although EPA may not be accustomed to using average instead of cumulative exposure, there are considerable data to indicate (as the EPA itself notes) that for beryllium, average exposure may be the important metric, accompanied by sufficient latency. Supporting evidence is provided by the data on sensitization, which indicates that average (or maximum, but this is more difficult to estimate and average will work well enough) exposure – rather than cumulative – predicts sensitization (which appears to almost inevitably result in CBD if subjects are followed long enough). The biological processes leading the sensitization may parallel those required for lung cancer due to beryllium, although this is not known. More importantly, the epidemiologic data for lung cancer from Steenland and Ward (1992, JNCI), from Ward et al. (1992 AJIM), and from Schubauer-Berigan (2008) et al. all suggest that duration of exposure is less relevant compared to average exposure, especially when accompanied by sufficient latency.

It is feasible to calculate URLs based on average rather than cumulative exposure. I.e., what is the level of excess lung cancer mortality predicted over a lifetime by an average exposure of 'x' ug/m^3, regardless of the duration of that exposure?

It would be preferable to calculate such a new URL rather than rely on the out of date Mancuso study of 30 years ago, with its accompanying more crude than necessary single

exposure estimate (exposure at a single estimated level vs. the non-exposed general population).

The stated EPA objections to using the Sanderson et al. data (now the Schubauer-Berigan data) do not make sense to me. Internal comparisons are perfectly acceptable, indeed preferable in so much as they minimize confounding, to external SMR comparisons.

The EPA struggled with using weighted-least squares of categorical points in the Sanderson et al. data when considering using these data for possible risk assessment, possibly omitting the final quartile which shows a decline in RR vs quartiles 2 and 3. The EPA also noted difficulty in using the log transform continuous model from Sanderson et al., due to the supra-linearity at low doses. Both of these issues are relevant, but also are solvable without great difficulty by using different models which account for the plateauing of the exposure-response at higher exposures. One such model is the two-piece linear curve (i.e., a spline with two points). It should be noted that occupational carcinogen data frequently exhibit a tailing off of RRs at highest exposures, which may be due to a variety of factors including a depletion of susceptibles, a saturation of relevant pathways at higher doses, greater measurement error at high doses leading to RR attenuation, or the healthy worker survivor effect, among other things. See Stayner and Steenland (Scan J Wk Env Hlth, 2003) for a full treatment of this issue. In this kind of situation (tailing off of dose-response at high exposures), a two-piece linear curve provides a good solution for risk assessment. For log linear models, it provides a nearly linear curve in the low dose regions, and it has the advantage of not ignoring the data in the high dose region (unlike the weighted regression approach using categorical data, while throwing out the uppermost category). It avoids the supra-linearity of the log transform model. The point of inflection may be chosen empirically by trying many points sequentially and using the best model likelihood. For an example of the 2 piece linear model for risk assessment using dioxin, see Steenland et al. (AJE, 2002), and for a general discussion of dose-response issues and splines see Steenland and Deddens (Epidemiology, 2004).

Difficulties in reproducing the NIOSH categorical cutpoints seem to me a non-issue, since one can closely work with NIOSH to make sure data analyses are conducted appropriately and NIOSH work can be duplicated.

Use of average instead of cumulative exposure for risk assessment may have a precedent in the CBD/sensitization data, in which a LOAEL is (appropriately) adopted from Kreiss et al. (1996). However, I do not have the Kreiss et al. study in hand, and EPA does not describe their method of calculating the LOAEL (this should be added, I am guessing that it was calculated based on average and not cumulative exposure). It is not clear to me why a URL could not also be calculated for sensitization, and compared to the URL for lung cancer. This would seem useful.

Finally, it should be noted that the incidence density sampling design used in Schubauer-Berigan et al. (and Sanderson et al.) is the standard and unbiased method for analyzing cohort data, notwithstanding claims to the contrary by Deubner et al. and Levy et al. Additional matching criteria in which controls are required to have an age at death similar

to their matched case will result in biased relative risk estimates, as is clear from first principles (Lubin and Gail, Biometrics, 1984).

Additional comments were requested following the peer review panel meeting of July 16, 2008. I was requested to provide more detail on 1) the likelihood of confounding by smoking, 2) problems associated with using a model with a log transformation of exposure, and 3) the possible preferential inclusion of patients with lung cancer in the study of the berylliosis case registry. Comments on these issues follow below.

1) Smoking is unlikely to be a confounder in the beryllium studies of Steenland et al., Ward et al. and Sanderson et al. All three of these studies made indirect adjustments for smoking, as described below. Regarding the first two of these studies, in which workers were compared to the general population (external comparisons) via SMRs, on theoretical grounds it has been shown (Axelson 1978, SJWEH) that only very large differences in smoking habits between exposed and nonexposed can account for lung rate ratios (e.g. SMRs) on the order of 1.3. Empirically Blair et al. (1985 J Occup Med) and Siemiatycki et al. (1988 J Occup Med) have shown that lung cancer rate ratios adjusted for smoking change little from unadjusted rate ratios. For internal comparisons of workers to workers (e.g., Sandeerson et al.), who would be expected to share smoking habits, little confounding by smoking would be expected. In external comparisons, where workers who smoke more are compared to the U.S. population which smokes less, confounding by smoking might be expected to account for an SMR of 1.10-1.20, using realistic differences in smoking (Blair et al 1985 J Occup Med, Axelson and Steenland, AJIM, 1988).

In the Ward et al. study, smoking data were available on 1466 cohort members (16% of cohort) in 1968, and age-specific smoking habits were compared to smoking habits in the U.S. population in either 1965 or 1970, with results being averaged. Using the technique of indirect adjustment suggested by Axelson (1978, SJWEH), and Axelson and Steenland (AJIM, 1988), Ward et al. calculated that the effect of smoking alone, absent of any occupational risk, would account for an SMR for their cohort of 1.13. Given that the SMR for those with long latency was 1.46, such confounding would be relatively minor and the SMR would still be significantly elevated, at about 1.3.

In the Sanderson et al. internal comparison of high exposed to low exposed workers in the Reading plant, again smoking data were available from the same 1968 survey. Sanderson et al. compared the duration, average, and cumulative estimated exposure to beryllium among the different smoking categories for the 386 workers in the survey (about 9% of the cohort), adjusted for age, and stratifying on blue collar vs white collar (professional) workers. The latter represented about 10% of the cohort and smoked less and had much lower exposures. Among blue-collar workers, there were no differences in exposure between smoking groups, indicating that smoking could not by definition be a confounder in this group. Furthermore, analyses of the blue-collar workers alone yielded the same results as the full cohort. These results suggests that smoking had little or no confounding effect in the Sanderson et al. analyses.

In Steenland and Ward's study of berylliosis patients (using an external referent group, the US population), data on smoking habits as of 1965 were available for 223 (32%) of the cohort, and these were compared to the smoking habits of the US population as of 1965. The cohort members smoked slightly less than the US population, suggesting the smoking would be a negative – rather than positive – confounder in this study. The finding that these workers smoked less than the US population was unusual, and may have been due to the effects of beryllium lung disease, such that patients were less likely to smoke.

2) Regarding the log transform model, as indicated above, this model fits well to data in which the dose-response tails off (attenuates) at high exposures, in that the log curve can turn downward at high exposures. Such is the nature of the dose-response described by Sanderson et al. (2001) and Shubauer-Berigan et al. (2008). However, at low exposures this log curve has a very high slope (i.e., is supra-linear) which is inappropriate for risk assessment in the low dose region, as it will predict a very steep increase in risk with a very small increase in dose. However, as indicated above, other models beside the published log-transform models used in Schubauer-Berigan et al. (2008) can be used which will also fit well data with an attenuated dose-response at high exposures. One such model is the 2-piece linear model, in which a nearly linear curve covers the low dose region, rising to an inflection point at which a second nearly linear piece covers the rest of the data. In the case of dose-response attenuation such as that which occurs with beryllium and lung cancer, the second piece has a markedly lower slope than the first piece.

3) Selection bias, in which lung cancer cases were preferentially included in the cohort of berylliosis patients studied by Steenland and Ward (1992, JNCI), is unlikely in this study for several reasons, all of which were discussed in the original publication. First, as mentioned by EPA in the IRIS document, the Registry records indicate only five cohort members had lung cancer when they entered the registry, and none of these five had lung cancer. Second, the lung cancers which did occur in this cohort generally occurred many years after entry into the Registry, and did not differ in time of followup (21 years) from other cohort members (20 years). Only 3 of the 28 lung cancers died within 5 years of beginning followup; had lung cancer cases been preferentially included in the Registry, one would expect them to have died soon, as lung cancer has a very poor survival rate. These facts argue against the idea that lung cancer cases were preferentially included in the Registry.

References for above, other than beryllium studies already cited elsewhere:

Axelson O (1978) Aspects on confounding in occupational health epidemiology, Scan J Work Environ Health 4: 85-89.

Axelson O and Steenland K (1998) Indirect methods of assessing tobacco use in occupational studies, Am J Ind Med, 13,1: 105-118.

Blair A, Hoar SK, Walrath J (1985) Comparison of crude and smoking-adjusted standardized mortality ratios.J Occup Med. Dec;27(12):881-4.

Lubin JH, Gail MH (1984) Biased selection of controls for case-control analyses of cohort studies. Biometrics. 40:63-75.

Siemiatycki J, Wacholder S, Dewar R, Cardis E, Greenwood C, Richardson L (1988) Degree of confounding bias related to smoking, ethnic group, and socioeconomic status in estimates of the associations between occupation and cancer, .J Occup Med. 30:617-25.

Steenland K and Deddens J (2004) A practical guide to exposure-response analyses and risk assessment in occupational epidemiology, Epidemiology 15: 63-70.

Stayner L, Steenland K, Dosemeci M, Hertz-Picciotto I (2003) Attenuation of exposure-response curves in occupational cohort studies at high exposure levels, Scan J Wk Env Hlth 29: 317-324.

Steenland K, Deddens J, Piacitelli L (2001) Risk assessment for 2,3,7,8-*p*-dioxin (TCDD) based on an epidemiologic study, Am J Epidemiol 154: 451-458.

2. Please identify any additional studies that should be considered in the assessment of the cancer health effects of beryllium.

Patrick N. Breysse

Include the following:

Schaubauer-Berigan et al. (2007). Adjustment for temporal confounders in a reanalysis of a case-control study of beryllium and lung cancer. JOEM 65:379 (2008).

Herman J. Gibb

Schubauer-Berigan MK, Deddens JA, Steenland K, Sanderson WT, Petersen MR. 2008. Adjustment for temporal confounders in a reanalysis of a case-control study of beryllium and lung cancer. Occup Environ Med 65:379-383.

The following are not studies but articles which at least merit consideration in the revision:

Deubner DC, Lockey JL, Kotin P, Powers MB, Miller F,. Rogers AE, Trichopoulos D. 2001. Re: Lung cancer case-control study of beryllium workers. Sanderson WT, Ward EM, Steenland K, Petersen MR. 40(3): 284-285.

Sanderson WT, Ward EM, Steenland K. 2001. Re: Response to criticisms of "lung cancer case-control study of beryllium workers" Am. J. Ind. Med. 2001. 40(3): 286-288.

David Kriebel

Schubauer-Berigan (Schubauer-Berigan, Deddens et al. 2008) reanalyzed Sanderson, and investigated potential confounding and effect modification by birth year (which was highly correlated with age at hire). The authors also assessed the sensitivity of the exposure-risk association to a small but potentially important methodologic choice; using a small value as a start when taking the logarithm of exposure metrics to avoid the impossibility of logging zero values. The Schubauer-Berigan reanalysis found clear evidence of an association between beryllium exposure and lung cancer risk, although the exposure metric and time lag which revealed the strongest evidence were different than in the original study. Sanderson reported an association between cumulative exposure with a latency of 20 years, while Schubauer-Berigan found the beryllium-lung cancer association when using average exposure with a 10 year latency. Changing the start value used when logging exposure metrics did not importantly affect the results. This paper was published as an eprint online in September 2007, and should be included in this review.

Researchers hired by the beryllium industry have raised methodologic objections to the Sanderson and Schubauer-Berigan methods of risk estimation which rely on logistic regression models and incidence density sampling for cases and controls (Deubner, Roth et al. 2007; Levy, Roth et al. 2007). I do not agree with this critique, and am confident that the analytic methods used by the NIOSH investigators are appropriate and unbiased.

Other methodologists have taken a similar position (Langholz and Richardson 2008). As Dr. Steenland notes in his comments, additional matching criteria forcing controls to have a similar distribution of age at death as the age at death distribution of the cases will result in biased relative risk estimates (Lubin and Gail 1984).

Brown et al. (Brown, Schonbeck et al. 2004) conducted a nested case control study of lung cancer in a cohort of plutonium-exposed workers at Rocky Flats. The main focus of the study was on risk from plutonium, but an attempt was also made to assess risks associated with asbestos, hexavalent chromium, nickel and beryllium. One hundred and twenty cases of primary lung cancer identified from death certificates and tumor registry were matched to 720 controls. There was evidence for increased lung cancer risk with increasing plutonium dose. Beryllium exposure was estimated by a job exposure matrix, but no details were provided in the paper. The authors reported that cumulative exposure to beryllium was not significantly associated with lung cancer risk, but no details or results were presented. This paper provides only limited evidence bearing on the question of beryllium carcinogenicity because no quantitative results were presented. For completeness however, it should be included in the review.

Following the peer review meeting discussion on July 15, I am adding here summaries of the literature on three methodologic issues which EPA should have considered in more detail in its review. These issues are: 1) confounding by smoking in occupational epidemiology; 2) the use of log transformed exposure data in occupational epidemiology; and 3) the tendency for exposure-response curves for occupational carcinogens to fall off at high exposures.

<u>Confounding by smoking.</u>

Confounding is understood by epidemiologists to be perhaps the single most important limitation to causal inference in observational studies, and the statistical control of confounding to be among the most important tools of the epidemiologist's trade. When data are unavailable for a known potential confounder however, our standard tools are not effective, and the researcher (or reader) often resorts to professional judgment in evaluating how seriously compromised the study results may be. One common mistake that is made in these instances is the failure to appreciate that weak confounding cannot explain strong apparent exposure-disease associations (Checkoway, Pearce et al. 2004). But such a general principle, while correct, is not very useful without guidance on how to assess strength of potential confounding, in the absence of data. This is one example of a kind of a what if question that can often be usefully informed by sensitivity analyses conducted as a complement to standard epidemiologic methods. A particular type of sensitivity analysis, sometimes called indirect adjustment or external adjustment, has been used to assess potential confounding by smoking in the absence of individual confounder data, in occupational cohort studies.

<u>Important literature on this method includes:</u>

Kriebel, D., A. Zeka, et al. Quantitative evaluation of the effects of uncontrolled confounding by alcohol and tobacco in occupational cancer studies. Int J Epidemiol 2004; 33:1-6.

Steenland, K. and S. Greenland. Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. Am J Epidemiol 2004; 160(4): 384-92.

Steenland K, Beaumont J, Halperin W. Methods of control for smoking in occupational cohort mortality studies. Scand J Work Environ Health 1984; 10:143-9.

Axelson O, Steenland K. Indirect methods of assessing the effects of tobacco use in occupational studies. Am J Ind Med 1988; 13:105-18.

Axelson O. Confounding from smoking in occupational epidemiology. Br J Ind Med 1989; 46:505-7.

Blair A, Stewart P, Lubin JH, Forastiere F.Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures. Am J Ind Med. 2007; 50(3):199-207.

Blair A, Steenland K, Shy C, O'Berg M, Halperin W, Thomas T. Control of smoking in occupational epidemiologic studies: methods and needs. Am J Ind Med 1988; 13:3-4.

Blair A, Stewart WF, Stewart PA, et al. A philosophy for dealing with hypothesized uncontrolled confounding in epidemiological investigations. Med Lav 1995; 86:106-10.

Axelson O. Aspects on confounding in occupational health epidemiology. Scand J Work Environ Health 1978; 4:85-89.

't Mannetje A, Kogevinas M, Chang Claude J, et al. Smoking as a confounder in case-control studies of occupational bladder cancer in women. Am J Ind Med 1999; 36:75-82.

Hooiveld M, Spee T, Burstyn I, Kromhout H, Heederik D. Lung cancer mortality in a Dutch cohort of asphalt workers: Evaluation of possible confounding by smoking. Am J Ind Med 2003; 43:79-87.

Siemiatycki J, Wacholder S, Dewar R, Cardis E, Greenwood C, Richardson L. Degree of confounding bias related to smoking, ethnic group, and socioeconomic status in estimates of the associations between occupation and cancer. J Occup Med 1988; 30:617-625.

Flanders WD, Khoury MJ. Indirect assessment of confounding: graphic description and limits on effect of adjusting for covariates. Epidemiology 1990; 1:239-46.

Bross IDJ. Pertinency of an extraneous variable. J Chron Dis 1967; 20:487-95.

Schlesselman JJ. Assessing effects of confounding variables. Am J Epidemiol 1978; 108:3-8.

Gail MH, Wacholder S, Lubin JH. Indirect corrections for confounding under multiplicative and additive risk models. Am J Ind Med 1988; 13:119-130.

Checkoway, H., N. Pearce, et al. (2004). Research Methods in Occupational Epidemiology. Oxford, Oxford University Press.

Log transforming exposure data

There was discussion during the reviewers' meeting on July 15 about the meaning of a log transform of an exposure metric – either cumulative or average exposure for example, in multivariate risk models. There are two related reasons for log transforming an exposure metric.

First, there are often a small number of very large values of the exposure metric, and if not logged, these will have a great deal of influence on the slope. Thus we log to reduce the influence of these few points, and to have a better fit to the full dataset. Second (and related) the relationship with the unlogged data may be non-linear and this is often because of a few very high exposures. These may not be associated with proportionally higher risks, so that there is an asymptote to the exposure-response curve. In these cases, logging may again improve the fit of the line to the data.

One should generally also fit a spline to look at the data untransformed, to see what the shape of the exposure-response relationship is without any filters caused by cutpoints or parametric models (like the linear one). Then, if the spline with the logged data looks more linear than on the native scale, we might log (Steenland and Deddens 2004). Leverage and influence statistics should also be done on the logged and unlogged data.

A biologic rationale is also relevant: we often find it more plausible to hypothesize that risk may rise on some multiplicative scale – logarithmically for example. Indeed, toxicologists often observe linear log-log relations between dose and many biologic responses. Schubauer-Berigan and colleagues (Schubauer-Berigan, Deddens et al. 2007) cite the classic text of Breslow and Day on a similar biologic rationale for the log transformation of exposure data ((Breslow and Day 1980) pp. 227–238) in cancer epidemiology.

If the log transform is used, how should one decide between the risk estimates using logged and unlogged exposure metrics? This should not be done using statistical significance alone (Loomis, Salvan et al. 1999). The correct approach is to use investigations of model adequacy as described above (splines, leverage and influence) and finally, goodness of fit. The better metric is the one that fits the data better. If the improvement in deviance or log likelihood is modest, then it will provide little guidance and one should use more qualitative criteria like those above. If the improvement in fit is

substantial, one should weigh this heavily in choosing between logged and unlogged metrics.

A related problem that has arisen in the beryllium cancer studies: logged metrics cannot include zero values, and so one must add a small quantity to all values, called a start. Unfortunately, the model fit may well be sensitive to the choice of start. If one chooses a start that is very small relative to the rest of the data, there will be a group of data (the former zeros) all with values which are quite distant (to the left) of the mass of the data. They may collectively have considerable influence at an arbitrarily chosen point, and artificially alter the goodness of fit and potentially the slope as well. This calls for sensitivity analyses to assess the impact of alternative starts (Greenland 1996).

Non-monotonic exposure-response curves

It is common in occupational epidemiology to observe an exposure-response curve which turns down at high exposures. Reasons for this pattern are explored in a paper by Stayner and Steenland (Stayner, Steenland et al. 2003). EPA should review this paper and consider the findings of the Schubauer-Berigan reanalysis of Sanderson in this light.

Lee S. Newman

Cancer Epidemiology:

Schubauer-Berigan MK, Deddens JA, Steenland K, Sanderson WT, Petersen MR. Adjustment for temporal confounders in a reanalysis of a case-control study of beryllium and lung cancer. Occup Environ Med 2008;65:379-383.

Genotoxicity:

Fahmy MA, Hassan NH, Farghaly AA, Hassan EE. Studies on the genotoxic effect of beryllium chloride and the possible protective role of selenium/vitamins A,C, and E. Mutat Res 2008;652:103-11.

Mode of Action:

Zhao JQ, Du GZ, Ziong YE, Wen YF, Bhadauria M, Nirala SK. Attenuation of beryllium induced hepatorenal dysfunction and oxidative stress in rodents by combined effect of gallic acid and piperine. Arch Pharm Res. 2007;30:1575-83.

Dobis DR, Sawyer RT, Gillespie MM, Huang J, Newman LS, Maier LA, Day BJ. Modulation of lymphocyte proliferation by antioxidants in chronic beryllium disease. Am J Respir Crit Care Med 2008;177:1002-11.

Coates SS, Lehnert BE, Sharma S, Kindell SM, Gary RK. Beryllium induces premature senescence in human fibroblasts. J Pharmacol Exp Ther 2007;322:70-9. (Other studies on cell senescence and beryllium-induced apoptosis should also be considered.)

Kyle Steenland

None.

3. Please discuss research that you think would be likely to reduce uncertainty in future assessments of beryllium.

Patrick N. Breysse

More epidemiologic studies of these same U.S. cohorts and registries, using the same limited exposure data, (with the exception of the ongoing NIOSH study) are not likely to be fruitful. The ongoing NIOSH study includes a detailed historical exposure reconstruction. The lack of systematic exposure reconstruction is a weakness of much of the existing epidemiology literature. More detailed exposure reconstruction may improve existing studies. The identification of new cohorts of beryllium exposed workers should be explored. One possible cohort is within the Department of Energy.

More information on a carcinogenic mechanism would also be helpful. Although I do not think the absence of a carcinogen mechanism should be used to discount the carcinogenicity of beryllium.

Herman J. Gibb

- An epidemiologic study that adequately addresses smoking.
- Studies on the mechanism of action.

David Kriebel

NIOSH is currently conducting a new retrospective cohort study of the principal U.S. beryllium production facilities, including a detailed exposure reconstruction. This study should provide considerably stronger findings on human lung cancer risk than the existing studies.

Lee S. Newman

Regarding carcinogenicity of beryllium, even greater certainty would be gained if future assessments were able to do more of the following:

- Examine cohorts not previously studied.
- Study cohorts with longer duration of follow up.
- Study larger workforce populations (greater study power).
- Improve exposure characterization, with better exposure measures including more monitoring data, information on beryllium forms, information on job titles, tasks, acute excursions, wider range of exposures, range of duration of exposure.
- Add greater precision in smoking exposure characterization.
- Collect data on exposures to other known carcinogens.
- Measure intermediate carcinogenesis pathway endpoints to better examine biological plausibility evidence in humans in cohorts, as well as experimental studies that examine mode of action hypotheses (e.g. inflammation x cancer risk).

- In the future, link cancer risk to beryllium sensitization and chronic beryllium disease (which are now more precisely defined than in decades prior to the 1980s).
- Additionally, better controlled animal carcinogenesis studies may be beneficial, especially in clarifying further the biological mechanisms of beryllium carcinogenesis.

Kyle Steenland

See above.

(B) Chemical Specific Charge Questions

1. Under the EPA's 2005 *Guidelines for carcinogen risk assessment* (www.epa.gov/iris/backgr-d.htm) (Section 2.5), beryllium via inhalation exposure is classified along a continuum between *likely to be carcinogenic to humans* and *carcinogenic to humans*. Please comment on the scientific justification for the cancer weight of the evidence characterization for this exposure route. Has the scientific justification for the weight of evidence characterization been sufficiently, transparently and objectively described?

Patrick N. Breysse

No. As discussed above, the weight of evidence indicates that beryllium should be classified as a human carcinogen. The rationale for discounting the epidemiologic literature provided in the document is not convincing. The available studies suffer from population overlap because these studies include the only populations of highly exposed beryllium workers in the U.S. There are no other groups to study. Given the consistent epidemiologic and animal literature, the rationale for **not** classifying beryllium as a human carcinogen should be reconsidered.

Herman J. Gibb

The justification for the classification has not been made. EPA has described a number of limitations to the evidence, and it is incumbent on the Agency to describe why such limitations do not detract from the classification. If after reviewing the logic, the Agency concludes that the classification is not justified, then the classification should be changed. The narrative should drive the classification not vice-versa.

The issue of smoking has not been adequately addressed. The relative risks are of such a low magnitude that smoking could have confounded the observed association between beryllium and lung cancer risk. The document provides little description of how the smoking data available in the various studies were collected or how adjustments for smoking were made. On page 41, the document indicates that within the professional and nonprofessional groups in the Sanderson et al. study, there were no statistically significant differences in beryllium exposure levels among current smokers, former smokers, and nonsmokers, thus indicating that smoking was not a major confounder. In the summary on page 44, it makes no mention of these inter-group comparisons, but states that the reason that smoking was not a confounder is because there was no difference in lung cancer risk between the professionals and nonprofessionals. The discussion needs to be made consistent. The document claims on page 44 that the question of whether smoking was a confounder was answered because both professionals and nonprofessionals had elevated risks of lung cancer. There were, however, only 14 professionals. If the risk was elevated in the professional group, the confidence limits around the odds ratio were likely very large including 1.0. Finally, the analyses which purportedly answer the question of whether smoking was a confounder are all indirect analyses (i.e., there are no regression analyses using smoking data).

The lack of an association when the exposure data are not log-transformed is not adequately addressed.

The apparent lack of elevated risk in the cohorts other than the Lorain and Reading cohorts in the Ward et al. study is not adequately addressed.

There appears to be no exposure-response in the Sanderson et al. study. There is some suggestion of an exposure response between the 2nd and 3rd quartiles of exposure when the exposure is lagged 20 years, but no evidence of an exposure response for the 2nd, 3rd, and 4th quartiles. Furthermore, does it make sense to lag exposure 20 years which would mean that exposures as long as 19 years prior to death had no effect on the development of the lung cancer? There was no evidence of an exposure response at all when exposures were lagged 0 or 10 years. If there is some reason for the tailing off of the response at the higher exposures, that needs to be addressed.

David Kriebel

There is not good scientific justification for the weight of evidence characterization in the draft review. First, the weight of evidence descriptor must be one of the limited set provided in the EPA guidelines. Second, the weight of evidence narrative is not adequately developed to support the descriptor. The problem with the descriptor is the use of the on the continuum language for carcinogenicity by the inhalation route.

The first paragraph on page 2-52 of the Guidelines for Carcinogen Risk Assessment makes it clear that EPA did not intend for the weight of evidence to be presented as lying between two descriptors, but rather that one descriptor should *always* be chosen, recognizing that the choice will often be difficult because of the complex nature of the available evidence. The paragraph states: In borderline cases, the narrative explains the case for choosing one descriptor and discusses the arguments for considering but not choosing another. In other words, the authors intended that a single descriptor will always be chosen, while acknowledging that the weight of evidence may at times appear to lie somewhere in between two descriptors. The Guidelines make clear that the text in the weight of evidence narrative should explain why a particular descriptor was chosen, and why the alternative nearby descriptors were not chosen. This has not been done in the draft review.

It is an unacceptable cop out for the draft review to say in essence: we can't decide which it is. If one were to follow the logic of the on the continuum label, it would suggest that there is some level of evidence between likely to be carcinogenic and carcinogenic. What would that level of evidence be – *very* likely to be carcinogenic? The EPA guidelines make clear that no further parsing of the levels of evidence is appropriate, and I agree. The draft review must pick one descriptor.

The IARC monograph concluded that there was *sufficient evidence* in humans of the carcinogenicity of beryllium and beryllium compounds. Their conclusion was based on the cohort studies then available (the Sanderson nested case control study had not yet been published). The IARC review committee noted these key strengths in the human data:

- A large number of lung cancer cases and stable rate ratios (SMRs);
- Consistency in findings among plants;
- Higher risks of lung cancer among workers hired before 1950, when exposures were higher;
- The plant with the highest lung cancer rate also had the highest proportion of acute berylliosis cases in the Beryllium Case Registry;
- Plants with high rates of non-malignant respiratory disease also tended to have high lung cancer rates;
- The high lung cancer rate among cases in the Beryllium Case Registry; and
- A pattern of increasing risks with increasing latency.

The Report on Carcinogens of the National Toxicology Program has listed beryllium and compounds as *known to be a human carcinogen* since the Tenth Report in 2002. The reasoning follows closely to that of IARC; NTP noted consistency of elevation in lung cancer risk among studies, the association with pneumonitis which suggests high levels of exposure, and patterns of increased risk with increased intensity or duration of exposure. They noted that there was no evidence to suggest increased smoking rates among beryllium exposed workers, and therefore discounted the possibility of serious confounding by smoking.

The Sanderson nested case control study and the further analyses of the same data provided by Schubauer-Berigan address several of the weaknesses in the body of evidence used by IARC and NTP, and so further strengthen the evidence for human carcinogenicity by the inhalation route. These analyses now provide stronger evidence for an increasing risk of lung cancer with increasing exposure. Because this was a nested case control study, the problem of choosing a comparison population that limits SMR studies has been avoided. Also, this study was able to assess risks from other occupational exposures, substantially reducing the possibility that the pattern of elevated lung cancer risk could be due to other carcinogens.

I believe that the appropriate descriptor for beryllium carcinogenicity by the inhalation route is: carcinogenic to humans. The weight of evidence narrative is also not adequately developed to support whatever descriptor is chosen. The discussions of potential confounding by smoking and by co-exposures are overly formulaic and do not really deal with the specific studies.

The Summary of Overall Weight of Evidence, Section 4.6.1., beginning on page 80, does not include discussion of the animal evidence on carcinogenicity by the inhalation route. The EPA guidelines make it clear that the animal evidence should be included in the

overall weight of evidence narrative about *human* carcinogenicity (see page 2-50 in the EPA guidelines referenced above).

Lee S. Newman

The weight of evidence with consistent epidemiologic findings across at least two cohort studies and a case-control study, plus resilience of the conclusions following multiple reanalyses of the data, plus a sufficient body of animal and experimental evidence, support the conclusion that beryllium via inhalation exposure should be classified as carcinogenic to humans.

The EPA draft is clearly written and transparent and is to be complemented on the effort. The characterization of the weight of evidence is overly cautious, erring in the direction of understating the evidence for carcinogenicity of beryllium. The draft, as written, places relatively greater weight on a number of relatively small areas of uncertainty in the data or analytic approach, than on the broader, consistent body of evidence in this literature showing beryllium's carcinogenicity. I favor application of the precautionary principle, especially when the weight of evidence is of the general magnitude and consistency seen here.

We as scientists are trained to find flaws in previous research. We endeavor to improve on previous study designs in an effort to reduce uncertainty. Such efforts should not, however, become an excuse for not taking appropriate, preventive measures, even while waiting for the next best study to be completed.

Kyle Steenland

See above.

2. EPA has determined that the literature published since the 1998 IRIS assessment is inadequate to support a reassessment of the cancer inhalation unit risk (please refer to Appendix C of draft assessment). Please comment on EPA's rationale for not deriving an updated inhalation unit risk. Please identify any currently available studies or methodologies that could be used to derive an inhalation unit risk (IUR). Please comment on EPA's decision to retain the current IUR values.

Patrick N. Breysse

EPA should reconsider calculating a new IUR based on the reanalysis of the Sanderson et al. paper by Schubaner-Berigan et al. (2008).

Herman J. Gibb

I recommend using the approach and data (Schubauer-Berigan et al.) suggested by Dr. Steenland to develop an IUR. I think that the approach which he suggests makes better use of available information. I also agree with Dr. Steenland that it would be beneficial to contrast the results of a re-worked IUR with a quantitative assessment for beryllium sensitization.

David Kriebel

I do not agree with EPA's rationale for not deriving an updated IUR. I believe that EPA can and should use the data from the Schubauer-Berigan reanalysis of Sanderson to calculate an updated IUR. Despite the limitations in these more recent data, they are almost certainly an improvement over the decades-old and much cruder study of Mancuso.

I agree with Dr. Steenland that EPA should use the results from Schubauer-Berigan (2008) for average exposure with a 10-year lag (See Table 2B, 9th row labeled: OR BY-adjusted, and Table 3 12th row which shows the same model using the log of average exposure as a continuous variable).

There is no compelling reason to prefer cumulative exposure over average exposure as the summary measure of exposure for models of beryllium and lung cancer. The correct metric depends entirely on the disease mechanism, which is not understood for beryllium (and only moderately better understood for the best-studied occupational carcinogens like asbestos and benzene). It is likely that neither average nor cumulative exposure is, in a biologic sense, the correct metric of beryllium dose (Kriebel 2008). If we could know the correct dose metric, we might expect to find that it was correlated with both cumulative and average exposures, as well as with calendar time and even with birth year (exposures often decrease with time as environmental controls are installed). But in such a scenario of correlated time dependent covariates, one cannot rely upon regression models to partition variance accurately between correlated time-dependent covariates like birth year and a summary measure of exposure which is only a rough surrogate for the true but latent dose metric. As noted above, it is not appropriate to use statistical significance to choose between alternative exposure metrics.

I do not agree with EPA's stated objection to using Sanderson et al. for IUR calculations because the study used an internal comparison. Internal comparisons should create no impediments to relative risk estimation, and there is the added advantage of much reduced confounding by avoiding inappropriate comparisons to an external reference population (healthy worker effect is minimized, for example).

I agree with Dr. Steenland's recommendations about how one could proceed with a risk assessment using the results for average exposure in Schubauer-Berigan.

Lee S. Newman

While I appreciate the EPA's efforts to reanalyze the Sanderson data set, and understand their rationale for not relying on categorical exposure-response data, it is my opinion that a weaker case is made for discarding the continuous exposure-response analysis methodology, especially considering the additional recent reanalysis provided by Schubauer-Berigan (2008) e.g. in regard to average exposure. Their rationale is overly conservative and in parts based on speculation (e.g. last paragraph on p. 82 and top of p. 83). As a consequence, industry may be continuing to place beryllium workers at risk for lung cancer mortality.

Kyle Steenland

See above. I disagree with EPA's decision to retain the current IUR values.

3. Given that EPA was not able to update the inhalation unit risk factor and that NIOSH is in the process of updating its cohort analysis (both extending the follow up time by 13 years and adding two facilities with more recent exposure levels) that should prove valuable in updating the IUR, do you recommend placing the update of this Toxicological Review on hold until these data have been incorporated?

Patrick N. Breysse

No. I recommend publishing this document now. When the NIOSH study is published the conclusions in this document should be reconsidered.

Herman J. Gibb

Not unless the information on smoking in the NIOSH update is considerably better than that available in the Sanderson et al. study.

David Kriebel

I do not agree with the proposal to wait for the updated NIOSH cohort study before reviewing the IUR values.

Lee S. Newman

No. The weight of evidence and consistency of findings are sufficient to designate beryllium as carcinogenic to humans. EPA should not postpone. Even while awaiting the publication of a larger retrospective cohort study, EPA should, like IARC and US National Toxicology Program, treat beryllium as a human carcinogen, even while researchers conduct further studies aimed at further reducing uncertainty.

Kyle Steenland

No, I do not. The data are at present sufficient to calculate an IUR. Future NIOSH work (data available unknown) can be used to update the IUR.

V. SPECIFIC OBSERVATIONS

Patrick N. Breysse

1. Change first sentence of last paragraph on page 26 to read: Limitations of this study included potential confounding by cigarette smoking, no quantitative exposure assessment, no latency analysis (other than duration of employment), and lack of clarity in the description of the analytic methods.

2. Discussion of Infante study on page 30: Mention how individuals gain entry into the beryllium case registry (BCR). This is disused in the section that discusses the Steenland and Ward study but should be moved up to the section on the Infante study where the BCR is first introduced.

3. Change sentence on top of page 37 to read: …set a quarterly DWA of 2 $\mu g/m^3$ that was later adopted by ACGIH and OSHA as an 8-hour time weighted average Threshold Limit Value and Permissible Exposure Limit value, respectively.

4. Charge second sentence on top of page 37 to read: AEC sampled the general work air, computing daily average exposures for each job in the plant.

5. Change second sentence, second paragraph on page 37 to read: The DWA was calculated by summing for all job tasks during a workday the product of the average beryllium concentration for each task or area by the time spent by the workers on that task, then dividing that value by the length of the workday.

6. Page 25, bottom of page: The document states… Some of these studies have found positive, statistically significant, associations between beryllium inhalation exposure and lung cancer (Table 4-1). This statement is too dismissive. All of the modern studies have identified increased cancer risk for beryllium. The consistency of findings from different investigators, with different designs, strengths, and weakness, is a remarkable finding that needs to be emphasized.

Herman J. Gibb

- SMR was defined as Standard Mortality Ratio; it should be Standardized Mortality Ratio (See pages 26, 28, 31).

- The statement that a sound causal association between human exposure and cancer has yet to be elucidated (pages 80 and 101) is odd. Causal associations are not something one thinks of as being elucidated. One either believes there is or there isn't a causal association. If it is yet to be elucidated suggests that the authors really don't know if any causal association exists

- ATSDR doesn't draw conclusions on the carcinogenic evidence as is indicated on page 83. ATSDR references the conclusions of IARC, EPA, etc.

- Sanderson et al. is described as a nested case-control study, but on page 37 it indicates that Sanderson et al. found an overall SMR for lung cancer of 1.22. Presumably there was a cohort analysis and a nested case control of the lung cancer cases, but the authors should provide more detail of the cohort study in the narrative.

David Kriebel

Page 28, Table 4-1. Levy et al. (2002 – incorrectly cited as 2007) is not a reanalysis, despite its title. It is an extended critique of the paper, but the authors did not have access to the data, and did not reanalyze them in any systematic way. Thus it is inappropriate to list this paper in a table summarizing the epidemiologic studies of beryllium. Brown (2004), although uninformative because of its small size and poor design, probably should be listed for completeness. Schubauer-Berigan's reanalysis of Sanderson *should* be included in this table as it is a true reanalysis of the original data.

Page 35. Again, Levy 2002 should not be summarized here, as it is not an actual reanalysis, but only a critique funded by the beryllium industry. There could be a short section summarizing the many letters to the editor and other critiques of the various studies, and Levy 2002 could be summarized there. The way it is currently presented lends too much weight to these critical comments.

Page 36. Table 4-4 should be omitted. It is not appropriate to present these findings as independent analyses when they are instead secondary findings not based on the original data, and were paid for by the beryllium industry specifically to critique the Ward (1992) study.

Page 43. Levy (2007) *is* an actual reanalysis of the Sanderson data; however in presenting its findings, it is important to note that this reanalysis was funded by the beryllium industry. Readers should be provided with this information so that they can judge the utility of the findings accordingly.

Page 43. There should be a summary of Schubauer-Berigan (2007), as noted elsewhere in these comments. Schubauer-Berigan (2007) reanalyzed Sanderson, and investigated potential confounding and effect modification by birth year (which was highly correlated with age at hire). They also assessed the sensitivity of the exposure-risk association to a small but potentially important methodologic choice; using a small value as a start when taking the logarithm of exposure metrics to avoid the impossibility of logging zero values. The Schubauer-Berigan reanalysis found clear evidence of an association between beryllium exposure and lung cancer risk, although the exposure metric and time lag which revealed the strongest evidence were different than in the original study. Sanderson reported an association between cumulative exposure with a latency of 20 years, while Schubauer-Berigan found the beryllium-lung cancer association when using average exposure with a 10 year latency. Changing the start value used when logging exposure metrics did not importantly affect the results.

Page 44, 2nd paragraph. This is an insufficiently developed discussion of the strengths and weaknesses of the Ward study. Overall, this was a strong occupational cancer cohort study, with high quality methods, conducted by a group with considerable skill and experience. Cohort construction and data analysis were very competent. The weaknesses are presented in a way to make them seem more important than they are. The investigation of confounding by smoking was not optimal, but used methods that are widely used, and found to be reliable in many other settings. The authors argue convincingly that it is unlikely that the observed excesses of lung cancer could have been explained by increased smoking rates in the plants.

It is unlikely that there was important residual confounding by smoking. The concern about co-exposure to acid mists is also overstated. Evidence for acid mists causing lung cancer is only suggestive according to Siemiatycki and colleagues, in their comprehensive summary and review of the occupational carcinogen evaluations by IARC (Siemiatycki, Richardson et al. 2004). Thus it is unlikely that this exposure could explain the finding of increased lung cancer in those exposed to beryllium. There were no other prevalent lung carcinogens in these work environments, and so it is not appropriate to cite lack of control for these as a weakness.

Concern about confounding by smoking and acid mists should be put in context by noting that confounding can only create an apparent (but false) association between an exposure and a disease if the associations between the confounder and the exposure and between the confounder and disease are *both* stronger than the apparent exposure-disease association. In this instance this would require for example that there be a very strong association between beryllium exposure and acid mist as well as a strong association between acid mist and lung cancer. As noted above, the evidence on acid mist and lung cancer is only suggestive, and it is therefore unlikely that this association could be sufficiently strong to explain the observed association between beryllium and lung cancer. In the case of smoking, it is unlikely that the association between beryllium exposure and smoking was sufficiently strong – this would require a substantially higher prevalence of smoking in the beryllium plants than in the general population, a pattern which is not supported by the data.

I find it unhelpful to cite (McMahon 1994) as the apparent source of these points of criticism of the Ward study. McMahon was hired as an expert witness by the beryllium industry, and if this work is to be cited, it should be noted that the author was employed by the industry so that readers can take this into consideration in their weighing of the criticisms.

Page 45, last sentence of first paragraph. I do not understand the point about latency versus duration. It does not seem useful to me.

Page 80. The section on the overall weight of evidence is not well-developed, and I believe its recommendation is not correct, as noted elsewhere.

Page 82, second paragraph. This paragraph is confused and not helpful. The Ward study documented elevated lung cancer rates in five of six different beryllium plants. The one plant with no evidence of increased risk, Lucky Ohio, had only 9 lung cancer deaths. The lack of evidence of excess risk at this plant is therefore not surprising, and does not constitute evidence against the presence of an effect. The last sentence of the paragraph is not clear.

Page 82, third paragraph. Like the last sentence of the preceding paragraph, this paragraph is not clear.

Page 82, last full paragraph. The internal EPA review should be more fully explained, if its findings are important to this review. It is not clear what is meant in the second sentence by ...the data set appears to be inadequate to effectively evaluate chronic long-term exposure to beryllium and lung cancer. Explain why this is so.

Page 82 and 83, paragraph beginning on 82 and continuing on 83. It is possible that there is a detection bias from being in the case registry, although its effect is not likely to be large. Lung cancer even in the 1950s and 1960s was generally well-diagnosed (although generally too late for any meaningful treatment). Thus it is not likely that there was much, if any, increased lung cancer detection among the BCR cases compared to the general population. The review should note that this bias is not likely to be large in size, or if EPA disagrees, there should be an effort made to assess the magnitude of the potential bias. The final sentence of the paragraph is too dismissive of the magnitude of the observed effect – see the SMRs in Table 4-2 on page 31. These findings are remarkably robust. I disagree that the level of confidence in these findings is low. Note also that the IARC committee also placed considerable confidence in the BCR results.

Page 83, 3rd paragraph. It is surprising that this review did not emphasize more fully the reasoning behind the IARC and NTP assessments. The proposed on the continuum categorization contradicts both of these independent reviews, and the reasons for the contradictions should be explained.

Page 93, first full paragraph. It is not clear why one might remove cases from the highest exposure quartile … This discussion of the Sanderson paper is not clear and not helpful. The discussion of the problem of logging exposure data should be amended to include the analyses performed by Schubauer-Berigan (2007).

Page 100, 2nd full paragraph, 2nd sentence. This sentence is not helpful in simply repeating the potential for confounding by smoking and other occupational exposures, when neither is very likely to explain the overall pattern of evidence.

Page 100, final paragraph. As detailed elsewhere, I disagree with the final assessment as along a continuum.

Page 101, first paragraph, 2nd sentence. I disagree that …a sound causal association…has yet to be elucidated. As noted above, I agree with IARC and NTP that the evidence for beryllium causing human lung cancer is sufficient.

Page C-3, first paragraph. The Schubauer-Berigan reanalysis of Sanderson was published in 2007, within the time frame of the literature included in this review.

New References Cited

Breslow, N. and N. Day (1980). The Analysis of Case-Control Studies. Lyon, World Health Organization, IARC.

Brown, S. C., M. F. Schonbeck, et al. (2004). Lung cancer and internal lung doses among plutonium workers at the Rocky Flats Plant: a case-control study. Am J Epidemiol **160**(2): 163-72.

Checkoway, H., N. Pearce, et al. (2004). Research Methods in Occupational Epidemiology. Oxford, Oxford University Press.

Deubner, D., H. Roth, et al. (2007). Empirical evaluation of complex epidemiologic study designs: workplace exposure and cancer. J Occup Environ Med **49**: 953-959.

Greenland, S. (1996). Basic methods for sensitivity analysis of biases Int J Epidemiol **25**: 1107-16.

Kriebel, D. (2008). Reanalysis: lessons great and small. Occup Environ Med **65**(6): 368-70.

Langholz, B. and D. Richardson (2008). Re: Empirical Evaluation of Complex Epidemiologic Study Designs: Workplace Exposure and Cancer. Letter to the editor. J Occ Env Medicine **50**: 1-2.

Levy, P., H. Roth, et al. (2007). Exposure to beryllium and occurrence of lung cancer: a reexamination of findings from a nested case–control study. . J Occup Environ Med **49**: 96-101.

Loomis, D., A. Salvan, et al. (1999). Selecting indices of occupational exposure for epidemiologic studies. Occcupational Hygiene **5**: 55-68.

Lubin, J. H. and M. H. Gail (1984). Biased selection of controls for case-control analyses of cohort studies. Biometrics **40**(1): 63-75.

Schubauer-Berigan, M., J. Deddens, et al. (2007). Re: Exposure to Beryllium and Occurrence of Lung Cancer: A Reexamination of Findings From a Nested Case-Control Study. J Occ & Env Med **49**(7): 708-709.

Schubauer-Berigan, M. K., J. A. Deddens, et al. (2008). Adjustment for temporal confounders in a reanalysis of a case-control study of beryllium and lung cancer. Occup Environ Med **65**(6): 379-83.

Siemiatycki, J., L. Richardson, et al. (2004). Listing occupational carcinogens. Environ Health Perspect **112**(15): 1447-59.

Stayner, L., K. Steenland, et al. (2003). Attenuation of exposure-response curves in occupational cohort studies at high exposure levels. Scand J Work Environ Health **29**(4): 317-24.

Steenland, K. and J. A. Deddens (2004). A practical guide to dose-response analyses and risk assessment in occupational epidemiology. Epidemiology **15**(1): 63-70.

Lee S. Newman

p. 92, l. 1 …Assessments, Is this sentence complete? Was there more to the paragraph?

Kyle Steenland

1. The discussion of consideration of smoking as a possible confounder in Sanderson et al. (it was not) on page 41 is garbled and needs to be re-written. The point is that smoking habits were not related to beryllium exposure level, making smoking by definition not a confounder, in the sub-sample where data on smoking were available. The mention of 'cases and controls' is irrelevant and confusing, and the discussion on professional status is also rather confusing – it was a surrogate of smoking level as I recall, with professionals smoking less, motivating their exclusion in a sub-analysis to see if dose-response trends persisted in a cohort of blue collar workers with presumably similar smoking status. This is clearer on page 44.

2. Paragraph in middle of p. 82 (beginning 'Comparing risk estimates….) makes no sense to me, I can't tell what EPA is getting at.

3. Page 83, statement that the 'elevated risk is small' in Steenland and Ward is not accurate, as the overall SMR was 2.00, a doubling of mortality.

4. I think that the 'ql' on page 97 is the same as the 'unit risk' in the table on page 96. If so, please make terminology consistent. If not, explain.

5. Should a URL be calculated, derivation of unit risk should take into account competing risks via an adjustment for all cause mortality as used in BEER.