TCE Meeting Presentation Supporting Material for L. Rhomberg
Paul Dugard
to:
Marc Rigas
05/03/2010 10:08 PM
Please respond to Paul Dugard
Show Details

Dear Dr Rigas:

Please find attached the supporting material for Dr Rhomberg's presentation on May 10. This was originally part of HSIA's submission of comments during the public review of the IRIS draft and it has been abstracted for the convenience of panel members.

Thank you.

Paul Dugard

Public Comments on US EPA

Draft Trichloroethylene Assessment

Docket ID No. EPA-HQ-ORD-0791

Prepared by

Lorenz R. Rhomberg, Ph.D., FATS

February 1, 2010

Gradient

1 February 2010

Submitted *via* Email to ORD.Docket.epa.gov

Re: Docket ID No. EPA–HQ–ORD–2009–0791; Public Comments on US EPA Draft Trichloroethylene Assessment

In accordance with the announcement in 74FR56834, I am pleased to present the following public comments on the United States Environmental Protection Agency (US EPA) "Draft Toxicological Review of Trichloroethylene: In Support of the Summary Information in the Integrated Risk Information System (IRIS)" [EPA/635/R-09/011A, dated October 2009]. I am writing as a Principal at Gradco LLC d/b/a Gradient, an environmental sciences consulting firm headquartered in Cambridge, MA. My comments are my own, but the effort to compile them was supported by the DuPont Corporate Remediation Group. These comments are intended for Docket ID No. EPA–HQ–ORD–2009–0791.

I write as a former US EPA employee who has worked on earlier assessments of this chemical while at the agency, and as a former academic who did dose-response analysis on trichloroethylene (TCE) animal bioassay data under contract to US EPA. I have also presented public comments on earlier assessment documents on behalf of a variety of clients.

The Draft TCE Reassessment (as I will call the document in question), is a complex and far-reaching one, with many novel analytical aspects and treating many relevant scientific issues. Although many of these aspects bear comment, I have chosen to focus my comments on one particular area: the methodology used in quantitative risk assessment for non-cancer effects of TCE.

Summary of Main Comments

Overall, the Draft TCE Reassessment reflects considerable effort and thoughtfulness on the part of US EPA in that it considered and incorporated a wide range of data in support of the proposed RfC and RfD for TCE. I respect the level of effort that was required to interpret and process this large body of data and the need to, in the end, propose one RfC and one RfD. However, although in the end a choice needs to be made, it is important, particularly in cases where a very large body of complicated data exists, that the process of making those choices involves a careful weight-of-evidence approach, so that the consequences of the various alternatives to those choices can be carefully worked through and considered as part of the overall analysis. Moreover, it is important that the process of evaluating

the various alternatives be transparent to the reader, and to the risk manager who ultimately will use the resulting criteria.

The four overarching comments I have on the assessment are: (1) it is important to carefully consider and clearly communicate how human variability was characterized in the assessment; (2) it is important to have a high level of confidence in the pharmacokinetic (PK) model assumption that humans generate higher levels of dichlorovinyl cysteine (DCVC) than rats; (3) a sufficient weight-of-evidence approach is needed for ascribing and communicating confidence across the large array of candidate RfDs/RfCs; and (4) clear communication of the entire process is needed.

The following bullets summarize the specific comments, which are discussed in more detail below.

- The allowance for inter-human PK variability double counts and misconstrues the nature of the dose-response curve.
- The reassessment document should specifically call out and individually characterize elements of its analysis that make particularly big differences to the RfC and RfD determination.
- The assessment should discuss proportionality between applied and internal dose, the circumstances under which that proportionality is seen as changing, and the impact of the changes on the quantitative risk analysis.
- It is not clear that DCVC constitutes an appropriate basis for an internal dose metric for kidney non-cancer toxicity.
- The document's conclusion that humans have high flux through the conjugative pathway is at odds with previous assessments, and is not well supported by evidence; yet, this assumption markedly lowers RfC/D values compared to those using traditional applied-dose approaches.
- Reliable estimates of the extent of variability among humans in DCVC activation have not been established, yet this factor is very influential in lowering the RfC/D.
- A weight-of-evidence approach should have been applied, and made transparent, to ascribe a level of confidence for each of the potential RfCs/RfDs derived in the assessment.
- There is considerable uncertainty in the proposed RfC of 0.001 ppm for TCE, particularly related to potential uncertainty in the Physiologically Based Pharmacokinetic (PBPK) modeling of the DCVC dose metric in humans, and the relationship of that dose metric with increased kidney weight.

Specific Comments

Comment #1 - The allowance for inter-human PK variability double counts and misconstrues the nature of the dose-response curve.

There are two questions about the allowance for human variability in metabolic activation. The first, addressed elsewhere in these comments, is whether the extent of variability has been reliably estimated. The second, addressed here, is how allowance for variability has been entered in to the RfD/C calculations. It would appear that allowance for human variability has been double-counted because inter-individual variability is built in to the tolerance distribution-based dose-response curve.

The method employed in the document is to set a point of departure (PoD) on the animal-based dose-response curve, using central estimates of "standard rat" internal doses as the dose measure. That is, inter-individual PK variation among rats, even though it exists, was not estimated and not considered in the dose-response curve estimation. For non-cancer endpoints, the dose-response curve is interpreted as a tolerance distribution – as the cumulative distribution of individual sensitivity variation. The reason that some animals respond at a given (externally applied) dose and others do not is that some have their individual tolerances exceeded while others do not, and higher doses exceed the individual tolerances of a greater fraction of the variable population, thereby yielding higher disease incidences.

Some of this variation is in PK, and so to some extent, the rats that respond do so because they are more vulnerable owing to their individual PK variation that makes them have a higher proportionality of internal to external dose. The contribution of this effect is captured in the fitted dose-response curve, which also reflects variation in sensitivity for other, non-PK reasons, but the contributions of PK variation are already incorporated, and are not readily split out without some attempt to characterize PK variation among individual rats.

The rat dose-response curve is then used to determine a PoD by finding a dose that yields a low predicted response, say 1%. Because the dose scale is measured in *average* internal dose among the rats, the dose associated with a 1% response level is the average internal dose for rats such that 1% of them are expected to have their individual tolerances exceeded. For the sake of argument, if we hypothetically say that there is absolutely no inter-rat variation in PK, then all

the rats in a hypothetical experiment at the 1% response dose will have the same internal dose, and which rats respond and which do not will be ruled entirely by variation in pharmacodynamic (PD) sensitivity to this fixed internal dose. But, if one instead hypothesizes that variation in sensitivity is entirely ruled by *PK* variation (with no contribution of PD variability) then the 1% of rats responding are that same 1% that are most sensitive owing to their PK variation – that is, they are the 99th percentile of the internal dose distribution.

The reality is somewhere in between, with both PK and PD variability contributing to variation in ability to tolerate the dose. But without characterization of PK variation among individual rats, we have no way to split the components out (though there is the conventional split between PK and PD that we apply to Uncertainty Factors).

Staying with the hypothetical case that sensitivity variation is all in PK, then the only reason to make further allowance for *human* PK variation is if variation in PK among humans is greater than variation among rats, and even then the correction should only be for the degree to which it is greater – that is, the ratio of the 99th percentile in humans *versus* the 99th percentile in rats rather than the ratio of the 99th to the 50th percentile in humans.

The hypothetical case of pure PK dependence of sensitivity variation is made to clarify the argument, but in the real case of contributions from both PK and PD, the principle illustrated still applies. There is some mix of influence of PK- and PD-based sensitivity among the responding rats, and the effect of this is captured in the fitted dose-response curve, for which the dose variable is the average internal dose. That internal dose is likely higher on average among the 1% of rats responding, because of the contribution of PK to their sensitivity; but, since this is unmeasured, all the analysis can say is that when a group of rats is dosed at a given external level, the average internal dose among them has some level estimated by the rat PBPK model. In view of the (unknown) contribution of PK to sensitivity and the (unknown) degree to which PK varies among rats, there is some (unknown) degree to which some rats have higher-than-average internal doses and thereby have an increased response probability (which is dictated by PD sensitivity to internal dose levels).

When the rat PoD is extrapolated to a human PoD based on average PK in the two species, it implicitly assumes that the mix of PK and PD, and the extent of inter-individual variation in PK, are the same in humans as in the rats. If one then makes a correction for the difference between the 50th percentile of PK in humans

and the 99th percentile (as the draft reassessment does) it essentially implicitly assumes that all of the variation in sensitivity reflected in the dose-response curve is attributable to PK alone.

If one assumes that the mix of PK and PD influence is similar across species, then the correct correction is the ratio of 99th percentiles across species, but since the 99th percentile in rats is not estimated, this cannot be calculated. If one cannot assume that the mix of PK and PD is the same, then it is doubly impossible to calculate a correction.

The method that has been employed in the draft reassessment seems to implicitly assume that all of the dose-response in rats is attributable to PD (and this drives the PoD down as far as possible in internal-dose terms) and that all of the dose-response in humans is attributable to PK (and this drives the sensitive human allowance down as far as possible). The net result is to yield an RfC that is overcorrected for human inter-individual variation to a degree that is not possible to know with the analyses available.

Comment #2 - The reassessment document should specifically call out and individually characterize elements of its analysis that make particularly big differences to the RfC and RfD determination.

The analyses in the draft reassessment document are complex and in many cases novel. Findings and conclusions from various chapters come to bear as they are applied to the RfD/C determination. It is difficult for a reader – and more importantly, for a risk manager using the document as a reference – to trace the sources of and reasoning behind analytical findings that are subsequently used in calculations and to gauge the impact of specific judgments or conclusions from earlier chapters on the final RfD/C determinations. Moreover, the impacts of specific choices or judgments on the final calculation – and the differences that would arise if different choices were made – need to be isolated and documented. Only in this way can a risk manager understand where the changes from earlier assessments are coming from, what they are based on, how reliable is the basis for that change, and how different the analysis would be under other arguably appropriate alternatives.

Transparency means more than just showing all the calculations in large appendices; there is a critical need for effective communication about the impact of

choices and judgments that are made, about the basis for those judgments, and about the impacts of those judgments vis-à-vis possible alternatives on the final outcome.

For example, it should be made clear that the chief impact on changing the RfD/C from what they would be under default procedures (and from how they were previously characterized) is the invocation of much greater flux through the conjugative metabolic pathway in humans than had previously been estimated. As discussed further elsewhere in these comments, this result is the chief reason that an internal-dose basis for an RfC based on kidney toxicity comes out much lower than if the RfC were based on other endpoints or on applied dose, though this conclusion is not obvious without deep reading of the document and detailed tracing of the calculations. There are reasons to question whether this finding of high human flux through the conjugative pathway is correct (as discussed elsewhere), but any discussion of that question and any documentation of the basis for that conclusion is far removed from its application in a later chapter. The discussion of what pathway, and what measure of that pathway's activity, is best used as an internal dose metric for kidney toxicity is in yet another place, and these conclusions can also be questioned. But again, that discussion (to the extent it exists anywhere) is far removed from its place of application.

A truly transparent analysis would (1) make it clear why the use of the particular internal dose metric changes the outcome; (2) describe how much it changes the outcome; (3) explain why the estimate of the flux in humans is now much greater than it had been in previous analyses; (4) articulate the main pros and cons of the basis for this increased estimate, the judgment as to its reliability, and the basis for that judgment; and (5) characterize the impact of having made other reasonable choices. It should do this in a concise and consolidated way as a commentary on the RfC calculation, available for a risk manager to use in understanding the basis and reliability of the number, with specific references (not just chapter numbers) to places in the document where the details are discussed.

Comment #3 - The assessment should discuss proportionality between applied and internal dose, the circumstances under which that proportionality is seen as changing, and the impact of the changes on the quantitative risk analysis.

If the internal dose measure at the level of a target organ or tissue is strictly proportional to the externally applied dose or to the exposure level, then analyses

based on internal or external doses should be identical. It is only when this proportionality changes that doing an internal-dose basis to the calculations changes; and this should be examined, not just as a "bottom line," but in a way that allows ascribing impacts to different parts of the calculation. The components to be considered are as follows:

- Change in proportionality of external to internal dose over the dose range of the bioassay experiments – these will affect the shape of the dose-response curve and its interpretation, since in the case of external dose, proportionality changes are incorporated into the estimated curvature;
- Change in proportionality of external to internal dose in low-dose extrapolations – this includes extrapolation down to a PoD and also any extrapolations below the PoD, since it will differ in outcome and interpretation if these are done on an internal- or external-dose basis;
- Change in proportionality of external to internal dose between routes of exposure – this influences the impact of using an internal-dose basis for route extrapolation, in particular the oral-to-inhalation extrapolation that is heavily used in the document, and it is needed to judge whether route-extrapolated endpoints are in accord with what was observed in bioassays by the route extrapolated to (for instance, whether the kidney toxicity in rats extrapolated from oral bioassays to inhalation are in accord with what was seen in rats in inhalation bioassays at internal doses comparable to those inferred in the extrapolation);
- Change in proportionality of external to internal dose over time patterns of exposure – animal bioassays often have five days/wk dosing, and human exposures are often intermittent and varying. To the extent that proportionality changes with the rate of dose delivery, this may affect calculations or the applicability of extrapolations;
- Change in proportionality of external to internal dose between experimental animals and humans – the difference between the modal animals and modal humans affects the cross-species extrapolation. As noted earlier, this is a large contributor to the low RfC calculated for TCE using internal dose, and the validity of the result hangs on the reliability of the estimate of relative human *vs*. rat metabolic activation, and on the choice of pathway to use as an internal dose measure; and
- Change in proportionality of external to internal dose among people (or animals) in a variable human population – both animals and humans will have some variation in the proportionality of internal to external dose, though the document only tries to estimate the extent of this in humans. A key question, however (as discussed below) is whether and by how much humans exceed the test animals in inter-individual variation in PK.

It is not that these questions are not discussed in the document, but many of the discussions are buried in details. Making these questions explicit, and evaluating their *individual* roles in altering the RfD/C calculations, is important to understanding how and why PK calculations and assumptions are affecting the outcome. The robustness of the calculations can then be judged according to the robustness of the basis for invoking changes in proportionality of external to internal dose.

Comment #4 - It is not clear that DCVC constitutes an appropriate basis for an internal dose metric for kidney non-cancer toxicity.

The kidney is seen as a sensitive target, and low RfC values drive the consideration of an overall RfC. The incorporation of internal doses makes the calculated RfC much lower than it would be if based on administered doses. It is therefore critically important that the internal-dose basis of kidney toxicity characterization be correct and reliable. The changes in non-cancer toxicity standards implied by the analyses in the Draft Reassessment hinge largely on assumptions about the PK of internal doses in kidney in rats and humans; and, if these assumptions are wrong, the basis for lowering the RfC is lost.

This being said, there are many questions about the PK assumptions that have been employed. First is the choice of DCVC as the basis for the dose metric. Just because DCVC is used for kidney cancer evaluation does not mean that the same dose measure is appropriate for non-cancer toxicity. Indeed, Lash *et al*. (2000) describe formic acid as a potential mode of action (MOA) for kidney damage for TCE, distinguishing the case of cancer and non-cancer kidney effects, stating, "Hence, although formic acid formation may contribute to TCE-induced renal damage, this is not likely to be a significant MOA in TCE-induced kidney *carcinogenesis*" (emphasis added). While the beta-lyase pathway may play a predominant role in kidney carcinogenesis, the possible roles of other chemical actors (formic acid and trichloroethanol) are not adequately addressed. The PBPK modeling effort focuses solely on the products of the beta-lyase pathway and apparently ignores these other possibilities. The conclusions are accordingly dependent on this being the correct dose metric. If alternative pathways could be addressed *via* the model, this could either provide some support for US EPA's position that they are not relevant or it could show that a different dose metric is warranted. The current argument, *i.e.*, that there are differences in kidney

histopathology between TCE- and trichloroethanol-treated rats, and that this indicates a different MOA, is not particularly compelling.

Comment #5 - The document's conclusion that humans have high flux through the conjugative pathway is at odds with previous assessments, and is not well supported by evidence; yet, this assumption markedly lowers RfC/D values compared to those using traditional applied-dose approaches.

The consensus of scientific opinion had been that humans have low flux through the conjugative pathway, which would lead to low internal doses to the kidney. It was also the consensus that it is difficult to pin down the extent of flux through this pathway for experimental reasons. The draft reassessment document indicates that the human flux through the conjugation pathway can be concluded to be much greater than in rats. In view of the importance of this judgment to the eventual RfD/C, it must be clearly explained why this altered conclusion is warranted.

As stated on page 3-128, the PBPK model reports one to two orders of magnitude more glutathione (GSH) conjugation and DCVC bioactivation in humans relative to rats. US EPA acknowledges that the 95% confidence intervals of the predicted population means for the two species overlap but there is little discussion of how this result is inconsistent with much of the previous data on TCE metabolism and TCE health effects in both humans and animals. For example, Lash *et al.* (2000) state that metabolic studies of PCE and Compound A indicate greater flux through the beta-lyase pathway in rats compared to humans (*i.e.*, several fold higher in rodents). It would be unusual if TCE were somehow different from these structurally similar compounds such that the flux in humans was many times higher than in rats. Along similar lines, Lash *et al.* (2007) state that the flux of tetrachloroethylene (PCE) through the GSH pathway is approximately fivefold faster in rodents than that of TCE. They also indicate that the reactive intermediates derived via the beta-lyase pathway from PCE are more reactive than those derived from TCE. This would suggest that PCE should be a much stronger kidney toxicant than TCE in the rat; yet, to our knowledge, neither chemical could be regarded as a very potent nephrotoxicant. For example, in the National Toxicology Program (NTP) and National Institute of Health (NIH) oral bioassays (NTP, 1990; NIH, 1977) toxic nephrosis was observed in rats treated with either chemical and at similar doses. In human studies, neither chemical is consistently shown to be a potent nephrotoxicant (if anything, studies

such as that by Henschler *et al*. (1995) would suggest TCE is more potent). This line of reasoning argues against the primary role of the beta-lyase pathway in PCE/TCE nephrotoxicity, and should be discussed in the document.

The basis for finding such large human flux through the conjugative pathway is also questionable. The result comes from the hierarchical Bayesian analysis of the PBPK model. The US EPA PBPK model yields good fits to the rat and human urinary DCVC excretion data and also to S-dichlorovinyl glutathione (DCVG) measured in human blood. We would suggest caution, however, in assuming that just because the model, as formulated and parameterized, fits the available DCVC/DCVG data, that highly quantitative predictions can then be made concerning the mean and variation of the various model parameters. This is particularly of concern given the huge changes resulting from the Bayesian updating of the DCVC bioactivation constants (*i.e.*, from 0.14 to 0.0087 in the rat and from 0.0021 to 0.023 in the human). The basis for the prior is not clear, but what is evident is that something other than direct experimental characterization is driving the updated DCVC bioactivation result, and some direct confirmation that such large flux actually occurs would seem critical to using this result in so influential a manner.

Given the disparity between the model results and prior general scientific opinion about rat vs. human differences in GSH conjugation towards TCE, it would be valuable to use the model to predict what possible DCVC target organ doses would be for some of the key epidemiology studies. The reported prevalence of kidney damage could then be compared across studies for logical consistency with estimated DCVC concentrations. This would serve as a useful "reality check" for a model that is making novel claims regarding chemical toxicity. In any case, a clear and convincing case must be made as to why the previous scientific consensus about human DCVC activation and its estimation is being overturned.

Comment #6 - Reliable estimates of the extent of variability among humans in DCVC activation have not been established, yet this factor is very influential in lowering the RfC/D.

It is not only the high estimate of the *average* amount of human DCVC activation via flux through the conjugative pathway that results in markedly lowered reference values, it is also the calculation of the impact of estimated *variability* among humans in this rate. Elsewhere in these comments it is argued

that the method for considering the impact of inter-human variability is flawed; but, in addition, there is the question of how reliably its extent has been estimated. In the previous comment it was noted that the soundness of the basis for estimating a much-changed average DCVC activation is unclear in view of widely acknowledged experimental difficulties and the evident influence of the Bayesian updating procedure. This concern applies even more to the characterization of variation among individuals, and great care must be taken to avoid attributing to genuine inter-individual variability differences that are really just due to experimental error, which can have marked effects for measurements on single individuals.

US EPA notes that the variability in the renal GSH conjugation and bioactivation of DCVC is substantial due to the data set of Lash *et al.* (1999, as cited in the assessment). The Lash *et al.* data set, consisting of eight males and eight females in the 100-ppm dose group and five individuals (three males, two females) in the 50-ppm dose group is indeed very limited for characterizing such an important parameter in the model. The stability of any variance estimate drawn from such a small sample size (when developing a model meant to characterize the whole human population) should be viewed as tentative. This has fairly important implications when attempting to use the PBPK model for RfC calculations in ways meant to protect large fractions (*i.e.*, 99%) of the human population. It would also be helpful to show the model predictions as compared to Lash *et al*.'s results for the 50-ppm dose group (Figure 3-10 only shows the 100-ppm group) to get a better sense of the model's predictive ability at lower exposure concentrations.

Comment #7 – A weight-of-evidence approach should have been applied, and made transparent, in order to ascribe a level of confidence for each of the potential RfCs/RfDs that were derived in this assessment.

The Draft TCE Reassessment takes the approach of conducting dose-response analysis on virtually every available data set on every non-cancer endpoint and turning each into a candidate RfC or RfD; it then winnows these down based largely on criteria of goodness-of-fit of dose-response models and of giving low ("most sensitive") RfD/RfC values. There is some advantage in this approach in that it allows one to look at the array of values that potentially reflect each endpoint. What such an approach deemphasizes, however, is the judgment about the representativeness of data sets in characterizing a potential human

toxicity and about the consistency among studies that investigate the same endpoint. Not all effects appearing in animals have an equally compelling case as potential effects in humans. A well behaved data set that can be closely fitted by conventional models does not necessarily represent the best estimate of that effect among several studies that examine it, and the analysis showing elevated responses at the lowest level is not necessarily the most reliable.

A weight-of-evidence approach should have been applied more rigorously and transparently, in which each endpoint or group of related endpoints is examined for consistency among studies and potential human relevance (Lewandowski and Rhomberg, 2005). An unreplicable effect, even if fitted well by dose-response analysis, does not provide a meaningful guide for human risk evaluation. It may be that a study that is not the most amenable to dose-response curve fitting or other quantitative analysis is nonetheless judged the best at fairly representing the body of studies on an endpoint for purposes of projecting potential human risk. It is appropriate to use the ability to fit good dose-response curves as a key criterion, but it should not be the only criterion. That is, the choice of studies on which to base RfD/RfC values is a judgment based on biological insights as well as on statistical curve fitting, and it must strike a balance between biological meaningfulness and representativeness on the one hand and well-behaved curve fitting on the other when these two aspects are in some degree of conflict. It must be made clear that all of the relevant studies for each endpoint were carefully considered and that the choices for data sets to represent particular endpoints are justifiable on both statistical and toxicological criteria.

Furthermore, the weight-of-evidence evaluation could have been incorporated into a quantitative level of confidence for each proposed RfC/RfD, so that there is a means for communicating the level of confidence represented by the weight-of-evidence for each endpoint. Given the large number of RfCs/RfDs derived in this assessment, and the implicit utility of these values by risk managers and decision makers, the level of confidence in each RfC/RfD relative to those chosen as the final proposed values is something that needs to be carefully considered through a weight-of-evidence evaluation for each endpoint, and then clearly communicated. However, it appears that more weight was placed on the most sensitive endpoint, and the goodness-of-fit for the dose-response data, in the absence of sufficient consideration of the weight of evidence in support of that endpoint as the critical effect.

This is particularly relevant in derivation of the proposed RfC (discussed more below), where the drivers of the lowest RfCs appear to be from six studies,

only one of which is an unpublished rat inhalation study where kidney effects were observed (Woolhiser *et al*., 2006, as cited in the Draft TCE Reassessment). The remaining studies were rodent oral studies from which route-to-route extrapolations were conducted to derive the RfC, a process that should affect their weight in the overall analysis of kidney effects since they include the added uncertainties inherent in the extrapolation, especially in the PK-based extrapolation, which must assume that the models are accurate and dependable for different routes, and that the measure of internal dose chosen is correct and comparable in toxicity even under the different time-patterns of tissue exposure inherent in the oral and inhalation exposures.

Additional uncertainties should be noted in a weight-of-evidence evaluation of kidney toxicity. As described above, there is uncertainty in extrapolation from rodents to humans in the DCVC bioactivation portion of the PBPK model that is the basis of the proposed RfC for kidney effects. There is additional uncertainty regarding whether the kidney effect endpoint from the Woolhiser *et al*. (2006) rat inhalation study (increased kidney weight) is in fact related to DCVC bioactivation. If an associated level of confidence, based on the weight of evidence, had been derived and presented for the RfC based on the Woolhiser *et al*. (2006) study (and for each proposed RfC and RfD), the reader, and risk managers and decision makers could evaluate the level of confidence in the proposed toxicity values against other potential RfCs/RfDs that may reflect what appear to be less sensitive endpoints, but perhaps with a higher associated level of confidence based on the weight of evidence.

Comment #8 – There is considerable uncertainty in the proposed RfC of 0.001 ppm for TCE, particularly related to potential uncertainty in the PBPK modeling of the DCVC dose metric in humans, and the relationship of that dose metric with increased kidney weight.

There is uncertainty in choosing the p-cRfCs in the lower end of the candidate RfC range.

The Draft TCE Reassessment presents a range of p-cRfCs (based on PBPK modeled internal dose metrics) and c-RfCs (based on applied dose) in Table 5-19. The values in this table reflect the lowest RfCs for the various effect domains. As discussed in section 5.1.5.2, US EPA suggests that although the range of lowest candidate RfCs within each health effect spans 3,000-fold (from 0.0003 to 0.9 ppm), there are

> [S]ix p-cRfCs from both oral and inhalation studies [that] are in the relatively narrow range of 0.0003 – 0.003 ppm at the low end of the overall range.

It further suggests, in the context of discussing the advantage of deriving multiple RfCs from multiple studies, that

> [W]hen multiple candidate values happen to fall within a narrow range at the low end of the overall range … that it leads to a more robust RfC (less sensitive to limitations of individual studies) and that it provides the important characterization that the RfC exposure level is similar for multiple noncancer effects rather than being based on a sole explicit critical effect.

Although more studies resulting in similar RfCs will provide, to some extent, more confidence in that range of RfCs, the confidence associated with each of those RfCs in that range should also be carefully considered using a weight-of-evidence approach (as discussed in the comment above), and in comparison to other proposed RfCs that may not be as low, but for which there may be more confidence based on the weight of evidence for those endpoints.

In fact, compared with these six RfCs that reflect the lowest range of candidate RfCs, five of which are oral studies (uncertainty in the route-to-route extrapolation is discussed below), there is another range of p-cRfCs on Table 5-19 that might suggest a more robust RfC range. Although without carefully reviewing the weight of evidence for each of these RfCs (and we did not conduct that level of review), we cannot state with certainty that this range would provide a more robust RfC in the end. However, there are six inhalation studies that represent another relatively narrow range of RfCs, from 0.013 to 0.12 ppm, two representing reproductive effects (0.013 and 0.017 ppm), one representing a neurological effect (0.016 ppm), one representing a developmental effect (0.062 ppm), and two representing immunological effects (0.11 and 0.12 ppm). It is not clear that this range of RfCs was considered, as a group, in comparison to the lower range of RfCs, as it was not discussed in the Draft TCE Reassessment. A discussion of the level of confidence in these six RfCs, based on a weight-of-evidence evaluation of the studies for those endpoints, in comparison to the level of confidence in the lower range of RfCs, might suggest that these RfCs as a group are more robust than the six that reflect the lower range of RfCs.

There is uncertainty in the oral to inhalation route-to-route extrapolation.

As discussed, although the six RfCs at the lower end of the range do fall within a narrow range of values, only one of these studies was a rat inhalation study (Woolhiser *et al*., 2006) in which increased kidney weight was the observed

effect. The studies that were the basis of the other five RfCs were rodent oral gavage (NTP, 1976 and NTP, 1988, as cited in the Draft TCE Reassessment) or drinking water (Keil *et al*., 2009 and Johnson *et al*., 2003, as cited in the Draft TCE Reassessment) studies, two of which were based on kidney effects (toxic nephropathy and toxic nephrosis), two of which were based on immunological effects (increased thymus weight and increased anti-dsDNA & anti-ssDNA Abs), and one of which was based on developmental effects (heart malformations). These five RfCs were all derived from a route-to-route extrapolation, and are therefore (on the exposure route basis alone) less certain than the one RfC derived from an inhalation study. In fact, in extrapolating tissue doses from the oral to inhalation route, one could check to see how the extrapolated tissue doses compare with what would be calculated as tissue doses in inhalation rodent studies. To the extent that similar internal doses were indeed examined in the oral and inhalation studies, similar toxicity should have been observed under the logic of the route-to-route extrapolation. In fact, toxicity in the inhalation studies is largely seen at high doses that, at least on an applied-dose basis, are large compared to the doses showing toxicity in oral studies. This casts doubt on the validity of the gavage-to-inhalation extrapolations.

There is uncertainty in the PBPK model reflecting a higher DCVC bioactivation in humans than in rodents.

In addition, the one p-cRfC that was based on an inhalation study (Woolhiser *et al*., 2006) was 400-fold lower than the cRfC derived from the applied dose default methodology from the same study. US EPA discusses how this difference is due to a 30- to 100-fold difference between rats and humans in DCVC bioactivation that is reflected in the PBPK modeling, with humans having a higher level of DCVC bioactivation in the model. As discussed above, there is uncertainty in this difference that needs careful consideration before placing such emphasis on this model as the basis of an inhalation RfC. Given that the Woolhiser *et al*. (2006) study is the only inhalation study in this narrow lower end of the range, this study inherently provides more weight to the proposed RfC than the other four oral studies, and is discussed in more detail below.

There are limitations, and lack of transparency, in using the Woolhiser et al. (2006) as the basis of one of the candidate RfCs.

The Woolhiser *et al*. (2006) study is an unpublished rat inhalation study that was designed to examine immunotoxicity of TCE, but also contained information on kidney weights. Therefore, there is no way for the reader to easily review the results of this study. As discussed in the Draft TCE Reassessment, rats were exposed to 0, 100, 300, and 1,000 ppm TCE for 6 hours/day, 5 days/week, for four

weeks. The authors observed significantly elevated kidney weights at 1,000 ppm TCE exposure. But the Draft TCE Reassessment notes that the "small number of animals and the variation in initial animal weight limit the ability of this study to determine statistically significant increases." Therefore, this study provides weak evidence that inhalation of TCE results in increased kidney weight.

There is uncertainty in the relevance of increased kidney weight as a critical effect for non-cancer effects of TCE.

The observed effect from the Woolhiser *et al*. (2006) study was increased kidney weight relative to body weight. One other rodent inhalation study (Kjellstrand *et al*., 1983) discussed in the Draft TCE Reassessment also observed increased kidney weight from TCE inhalation, and another (Maltoni *et al*., 1988) observed meganucleocytosis. It is not clear that increased kidney weight or meganucleocytosis is directly related to kidney toxicity. Although some older studies seem to suggest that kidney weight increase is related to kidney toxicity (Feron *et al*., 1973), more recent studies (Bailey *et al*., 2004) suggest that the kidney weight to body weight ratio is uncertain, and other methods should be used to confirm weight increases. Barton and Clewell (2000) note that "Although short exposures produced increased kidney weight, it is unclear if this represents a reliable indicator of chronic toxicity (53,54)." As discussed by Hayes (2008), organ weight to body weight changes are typically secondary effects and not necessarily adverse. In addition, there does not appear to be any evidence to suggest that DCVC bioactivation is related to increased kidney weight, at least this is not discussed in the Draft Reassessment.

Summary

Although derivation and consideration of a range of RfCs is a sound approach to deriving an RfC, choosing the lowest range of RfCs (without a sufficient weight-of-evidence evaluation of the RfCs in that range), reflected by only one inhalation study for which the effect of increased kidney weight is questionable, is not strongly supported by the scientific evidence for TCE non-cancer effects. This is based on: (1) the fact that the significance of the observed effect in the Woolhiser study was weak and based on a small sample size; (2) uncertainty in the oral to inhalation route-to-route extrapolation for the five other RfCs in the range; (3) uncertainty in the PBPK model reflecting a higher DCVC bioactivation in humans than in rodents that was used for three of these RfCs; (4) uncertainty in the relevance of increased kidney weight as a critical effect for non-cancer effects of TCE; and finally, (5) the fact that there is another narrow range of six RfCs (from 0.013 to 0.12 ppm) that are all based on inhalation studies and for which, had a level of confidence in those RfCs been presented, might in fact reflect

a more robust set of RfCs, base on a weight-of-evidence analysis of those endpoints.

CONCLUSIONS

TCE is an important chemical, and its assessment by the agency has been in process for decades. It is important that any final reassessment be based on scientific grounds that can be readily understood and gain wide acceptance. Attention to the matters discussed above would help produce such a document.

Thank you for the opportunity to provide comments. I may be reached at 617-395-5552 or lrhomberg@gradientcrop.com or by mail at Gradient (5th Fl), 20 University Road, Cambridge, MA 02138.

Sincerely,

Lorenz R. Rhomberg, Ph.D., FATS

References

Bailey, SA; Zidell, RH; Perry, RW. 2004. "Relationships between organ weight and body/brain weight in the rat: What is the best analytical endpoint?" *Toxicol. Pathol.* 32(4):448-466.

Barton, HA; Clewell, HJ. 2000. "Evaluating noncancer effects of trichloroethylene: Dosimetry, mode of action, and risk assessment." *Environ. Health Perspect.* 108(Suppl. 2):323-334.

Feron, VJ; de Groot, AP; Spanjers, MT; Til, HP. 1973. "An evaluation of the criterion "organ weight" under conditions of growth retardation." *Food Cosmet. Toxicol.* 11(1):85-94.

Hayes, AW; ed. 2008. *Principles and Methods of Toxicology (Fifth Edition)*. Taylor & Francis, Philadelphia, PA, 2270p.

Henschler, D; Vamvakas, S; Lammert, M; Dekant, W; Kraus, B; Thomas, B; Ulm, K. 1995. "Increased incidence of renal cell tumors in a cohort of cardboard workers exposed to trichloroethene." *Arch. Toxicol.* 69(5):291-299.

Kjellstrand, P; Holmquist, B; Mandahl, N; Bjerkemo, M. 1983. "Effects of continuous trichloroethylene inhalation on different strains of mice." *Acta Pharmacol. Toxicol. (Copenh)* 53:369-374.

Lash, LH; Parker, JC; Scott, CS. 2000. "Modes of action of trichloroethylene for kidney tumorigenesis." *Environ. Health Perspect.* 108(Suppl. 2):225-240.

Lash, LH; Putt, DA; Huang, P; Hueni, SE; Parker, JC. 2007. "Modulation of hepatic and renal metabolism and toxicity of trichloroethylene and perchloroethylene by alterations in status of cytochrome P450 and glutathione." *Toxicology* 235(1-2):11-26.

Lewandowski, TA; Rhomberg, LR. 2005. "A proposed methodology for selecting a trichloroethylene inhalation unit risk value for use in risk assessment." *Regul. Toxicol. Pharmacol.* 41:39-54.

Maltoni, C; Lefemine, G; Cotti, G; Perino, G. 1988. "Long-term carcinogenicity bioassays on trichloroethylene administered by inhalation to Sprague-Dawley rats and Swiss and B6C3F1 mice." *Ann. NY Acad. Sci.* 534:316-342.

National Institute of Health (NIH). 1977. "Bioassay of tetrachloroethylene for possible carcinogenicity." Bethesda, MD. National Technical Information Service (NTIS), Springfield, VA. NCI-CG-TR-13; NIH 77-813.

National Toxicology Program (NTP). 1990. "Carcinogenesis studies of trichloroethylene (without epichlorohydrin) (CAS No. 79-01-6) in F344/N rats and B6C3F1 mice (gavage studies)." Research Triangle Park, NC. National Institutes of Health. NTP TR 243; NIH Publication No. 90-1779. 174p., May.