A Project of the
Alliance for Risk Assessment

Guidance for Contaminated Sites:

Trichloroethylene (TCE) Risk Assessment Case Study

April 15, 2013

allianceforrisk.org

Executive Summary

The overall purpose of this effort was to develop a range in noncancer risks, similar to the range used for cancer risks in management of waste sites. This range would enable noncancer risk to be evaluated during site decisions in a way that reflects the uncertainty of the noncancer benchmark. Based on readily available information from U.S. Environmental Protection Agency (EPA) and elsewhere, we adopted a general approach to develop a range for noncancer risk values, such as Reference Concentrations (RfCs). The recent evaluation of EPA's RfC for trichloroethylene (TCE) was used as a case study. The confidence in this range can be considered in the determination of risk management choices.

For TCE, this range was judged to be 3 $\mu g/m^3$ to 20 $\mu g/m^3$. The results of the NTP study-based RfC were used to determine the floor and midpoint of this uncertainty range. The highly controversial results from the Johnson et al. (2003) study-based RfC, while associated with low confidence, were nevertheless used to determine the ceiling level of this uncertainty range. This 3 $\mu g/m^3$ to 20 $\mu g/m^3$ range was entirely within the wider individual uncertainty range from the Keil et al. (2009) study; therefore, this latter study was considered to be confirmatory.

Toxicologists are not able to distinguish the absence of health risk between any two or more values within this overall uncertainty range of 3 $\mu g/m^3$ to 20 $\mu g/m^3$. That is to say, toxicologists cannot differentiate the "safety" of 4 $\mu g/m^3$ versus 17 $\mu g/m^3$. Because of this, managers may use different values within the multi-endpoint uncertainty range of 3 $\mu g/m^3$ to 20 $\mu g/m^3$, along with site-specific exposure assessments, and other site considerations to make different management decisions on a case-by-case basis.

For example, when a site-specific exposure assessment defines a range of exposures that are primarily below the multi-endpoint uncertainty range of 3 $\mu g/m^3$ to 20 $\mu g/m^3$, then the probability of inducing any noncancer effects in the exposure population is lower and the priority for any management action is reduced (see ES Figure 1a). In this case, a risk manager may decide to take no action, or to delay action pending further information.

In contrast, when the exposure assessment defines a range in exposures that is primarily above the multi-endpoint uncertainty range of 3 $\mu g/m^3$ to 20 $\mu g/m^3$, then the probability of inducing noncancer effects in the exposure population is higher and the priority for risk management action is increased (see ES Figure 1c). In this case, a risk manager may decide to take action, or to ask for specific information that would refine the estimates of health risk and/or exposure.

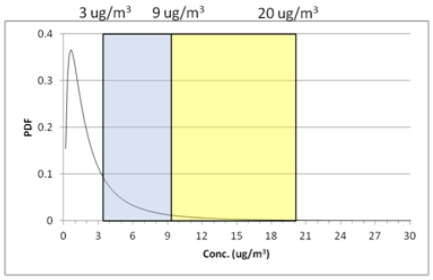
When the exposure assessment defines a range in exposures that are primarily in the multi-endpoint uncertainty range of 3 $\mu g/m^3$ to 20 $\mu g/m^3$, then risk managers can use the intermediate value in this uncertainty range (*i.e.*, 9 $\mu g/m^3$) and other site considerations, to gauge whether a management action is needed or if further information should be sought (see ES Figure 1b).

The multi-endpoint uncertainty range can also be used to develop a comparable range of hazard quotient estimates for single-chemical exposures to TCE alone, or hazard index estimates for exposures to mixtures of TCE and other chemicals (*e.g.*, solvent impurities such as 1,1,1-trichloroethane, 1,2-dichloroethane, 1,1-dichloroethene; or degradation products, such as *cis*-1,2-dichloroethene and vinyl chloride), when only a point estimate of exposure is available for comparison. This adaptation is consistent with EPA's definition of the RfD/RfC, and is akin to

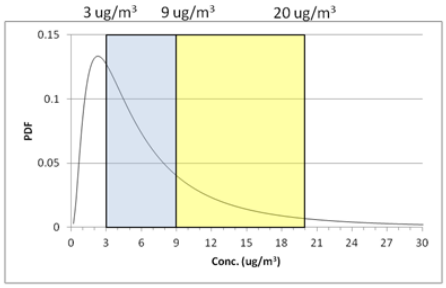
the range developed for cancer risk (i.e., an excess lifetime cancer risk of 10^{-6} to 10^{-4}). When considering exposures to mixtures, risk managers should be cognizant of the target organ(s), mode(s) of action and mechanism(s) of action of the various chemicals in the mixture, based on critical effects and other effects that may be elicited at environmentally relevant concentrations. Additionally, the role of co-exposures and interactions of chemicals may be considered and/or ruled out in developing the range in the hazard index for actions that are recommended by risk managers.

The multi-endpoint uncertainty range is composed of floor, midpoint and ceiling values for which the appropriate averaging time corresponds to different exposure durations (*i.e.*, developmental or chronic exposure periods). Therefore, this range can be applied to both long- and short-term exposures, with the associated differences in exposure averaging times. For shorter-term exposures, the results from the Johnson *et al.* study (2003) might also be used to describe the best averaging time, but if so, this averaging time should be based on the average time of cardiac development in humans during fetal growth, approximately 24 days. The use of a 24-day average time for cardiac development in humans is consistent with the fact that the dosing in the Johnson *et al.* study (2003) occurred during the whole time of cardiac development in the rat.

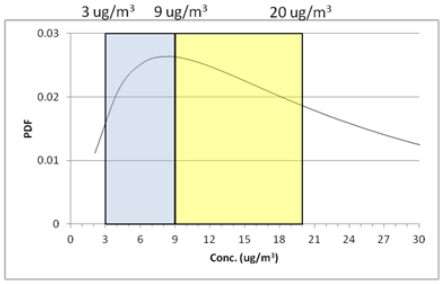
During the course of this study, it became apparent that EPA's use of cardiac anomalies for developing an RfC is highly controversial and not universally shared among government agencies and experts bodies. It may be that an analysis of this specific effect is warranted by appropriate experts, but at the very least risk managers need to understand that this endpoint cannot be used with confidence.



ES Figure 1a. Exposure distribution of indoor air concentrations primarily below the 3 µg/m^3 to 20 µg/m^3 hazard range. Relatively small proportion of exposures is higher than 3 µg/m^3. Nominal actions or no further action may be warranted for risk management.



ES Figure 1b. Exposure distribution of indoor air concentrations falling within the 3 µg/m^3 to 20 µg/m^3 hazard range. Relatively small proportion of exposures is higher than 9 µg/m^3. Limited action may be warranted for risk management



ES Figure 1c. Exposure distribution of indoor air concentrations primarily above the 3 µg/m^3 to 20 µg/m^3 hazard range. Actions to reduce exposures may be warranted for risk management.

Introduction

Risk managers responsible for making choices about acceptable oral and inhalation exposures at hazardous waste clean-up sites may benefit from an in-depth understanding of the imprecision and uncertainty in the process used to develop non-cancer toxicity criteria, specifically, non-cancer Reference Doses (RfDs) and Reference Concentrations (RfCs). This understanding will allow a higher level of confidence in decisions regarding site screening and closure, with respect to the evaluation of the non-cancer effects of chemicals present at the site. Although screening and closure exposure levels based upon the non-cancer endpoint (*i.e.*, "non-cancer driven") are now common, this was generally not the case in the past. Formerly, cancer risk was typically the driver in site decisions for any chemical evaluated with respect to both cancer risk and non-cancer hazard, particularly when risk management decisions emphasized the lower end of the excess lifetime cancer risk range. As risk managers became increasingly familiar with how to effectively manage the cancer risks associated with site chemicals, it is now routine to make screening and closure exposure level decisions using the widely-accepted 100 fold, cancer risk range (*i.e.*, 10^{-4} to 10^{-6}).

However, recent experience indicates that risk managers must contend with situations where risk management decisions at sites are driven by non-cancer endpoints. Recently, the RfD and RfC values for some chemicals (*e.g.*, trichloroethylene, or TCE) have been revised to lower levels (*i.e.*, indicative of greater toxicity), thereby becoming the subject of much attention. For

example, the RfD and RfC values for TCE published in September 2011 (U.S. EPA, 2011a) are lower than the values previously in use. The lower RfD and RfC values resulted from the evaluation of critical effects not previously quantified, and the use of dose modeling that enabled the estimation of effects resulting from low-dose exposures (*i.e.*, at levels between the lowest observed adverse effect level (LOAEL) and the no observed adverse effect level (NOAEL).

The hazard quotient (HQ) is a quantitative indicator of the magnitude of hazard associated with an exposure to a chemical by a receptor, based on non-cancer effects. The HQ may be defined as the ratio of the exposure dose to the RfD, or the ratio of the exposure concentration to the RfC. Multiple exposures (by multiple routes of exposure and/or chemicals) are characterized by a hazard index (HI). When one or more non-cancer endpoints drive risk management decisions for chemical exposures, risk managers generally apply a point value (*i.e.*, an HQ = 1) as the criterion of safety; risk managers do not generally apply a range of acceptable HQ values (*i.e.*, a "hazard range"). This conventional practice is in contrast to risk management decisions driven by the cancer endpoint, wherein such decisions are made on the basis of the cancer risk range described above. Although U.S. EPA (Integrated Risk Information System) has defined the RfD/RfC as possessing "uncertainty spanning perhaps an order of magnitude", risk managers have generally not implemented decisions based upon this implicit uncertainty. Consequently, non-cancer hazards have frequently been evaluated and regulated with sort of an "off/on switch" that is at, or very near, the RfD/RfC, wherein the hazard quotient of one (1) is interpreted as a "bright line" for risk management decision-making.

Risk management tools that incorporate the uncertainty associated with each RfD/RfC value have not been widely implemented. The statement "uncertainty spanning perhaps an order of magnitude" is sometimes interpreted as a precautionary statement addressing the implicit health-protectiveness of the RfD/RfC value. This interpretation is perhaps an intuitive one, because the derivation process is considered so uncertain. This perspective heightens the health concerns associated with exposures exceeding a hazard quotient of one. However, the process by which uncertainty is accounted for in the derivation of the RfD/RfC is *implicitly* precautionary; the application of an uncertainty factor provides a margin of safety in response to each source of uncertainty. Each uncertainty factor (greater than one) serves to reduce the RfD/RfC by increasing the gap between the RfD/RfC and the protective value (*i.e.*, the "safe dose" based on experimental observation, dose adjustment and toxicokinetic extrapolation). Most determinations of the RfD/RfC have more than one source of uncertainty, and consequently exhibit multiple successive precautionary adjustments (*i.e.*, the application of multiple uncertainty factors). The use of multiple sequential uncertainty factors significantly increases the margin of safety and the implicit protectiveness of the RfD/RfC.

It is not surprising that many risk mangers view the evaluation of non-cancer hazards as a "bright line". Throughout most of the 1980s and 1990s, the NOAEL (the presumptive "safe dose" below the "effect threshold") for the critical effect served as the "point of departure" (POD) for the derivation of the RfD/RfC by the application of one or more uncertainty factors (each uncertainty factor generally equal to three or ten). In the past 15 years, however, the RfD/RfC has increasingly been based upon sophisticated models, which define internal concentrations from exposure dosing, account for toxicokinetic differences between the test animal and human populations, perform route-to-route internal dose extrapolations, and interpolate effects levels between the LOAEL and NOAEL in the experimental study. These probability-based elements

are now commonly incorporated into the RfD/RfC derivation process, thereby uncoupling the POD from the threshold dose (or, at least, obscuring the association between them).

In this paper, the authors propose the development of a "hazard range" as a risk management tool to facilitate informed decision-making about exposures and likely effects to humans (including sensitive sub-populations). This hazard range accounts for the imprecision and uncertainty implicit in the derivation of the RfD/RfC value. The lower end of the range, commonly referred to as the "floor", is the RfD/RfC. A distinguishable upper end of the range is established, generally associated with a specific toxic effect level. Thus, the lower and upper ends of the hazard range can be clearly established from information available on the derivation of the RfD/RfC. Risk management decisions about exposure levels wholly above or below the range may be straightforward. However, risk management decisions about exposure levels wholly or partially within the range are more complex; therefore, it is helpful to have a "midpoint" within the range to aid in these decisions.

The use of many uncertainty factors to derive the RfD/RfC, and the inherent imprecision of the RfD/RfC value, each contribute to how the midpoint is established. These considerations include the level of confidence in the dose-response curve, and the completeness and representativeness of the animal study(ies) used as the basis of the RfD/RfC derivation. Thus, a comprehensive analysis of the imprecision and uncertainty used in the derivation process is necessary and results in a hazard range with a protective floor, a midpoint, and an effects level ceiling.

How a responsible risk manager addresses exposure levels within that range is dependent on both the imprecision and uncertainty analysis, and two additional risk management evaluations. The first evaluation takes a holistic look at the margin of safety applied in response to each source of uncertainty in the RfD/RfC derivation process. This risk management evaluation is often more semi-qualitative, relying on an understanding of how three-fold or ten-fold adjustments (*i.e.*, equivalent to uncertainty factors of one-half or one full order of magnitude) were applied and how these adjustments impact the level of confidence in the "health effects level" and the RfD/RfC derived from it. The second risk management evaluation addresses the sum of the animal and human evidence linking the toxic effect to the exposure. Here consideration of the strength of epidemiology evidence, repeatability of the animal results and the general quality of the studies and extrapolations used are examined for important strengths and weaknesses that may affect level of confidence in decisions.

Responsible risk management considers both health protection and practical implementation; an enhanced understanding of the inherent imprecision, uncertainty and margin of safety built into the RfD/RfC facilitates and enables responsible risk management. In this paper, the authors have proposed a method for determining a hazard range to inform such an understanding. The establishment of a hazard range for the RfC for TCE has been used as an example.

Risk management concerns resulting from widespread TCE exposure are considerable. Its use as a common degreaser and its prevalence as a breakdown product of the dry cleaner agent tetrachloroethylene (TCE) have resulted in widespread exposure. The volatility of TCE creates an indoor air inhalation concern via volatilization from soil or groundwater through the subsurface. Common disposal practices with older dry-cleaning facilities, automotive repair shops and machining operations have resulted in groundwater and soil contamination, and potential inhalation exposure to indoor air at many commercial and industrial settings.

Additionally, an understanding of non-cancer hazard is important because common household products have historically contained TCE. This results in indoor air "background" in about half the residential structures (Dawson and McAlary, 2009). The indoor air background levels may approximate the RfC in many structures and may confound or complicate determinations of whether subsurface contamination is the source of TCE in indoor air through vapor intrusion. These important variables regarding potential and actual human exposures amplifies the importance of a clearer understanding of the hazards associated with environmental concentrations above the RfC.

The analysis of TCE is complicated by the fact that TCE has both long term and short term exposure concerns. The chronic RfC was determined using an average of two critical effects studies (Keil *et al*., 2009; Johnson *et al*., 2003), with support from a third studies (NTP, 1988), all of which were incorporated into a range of candidate RfC values that are reasonably close in magnitude. Chronic exposure concerns exist for developmental, immune system and kidney toxic effects. Based on U.S. EPA's use of one study (which examined exposure across the 21-day fetal developmental period in the rat and identified developmental fetal heart malformations), the RfC incorporates effects based on short term exposures. Thus, we examine the hazard associated with both short term and chronic exposures to environmental concentrations above the RfC.

To provide guidance on the practical understanding of the uncertainty and other non-cancer issues associated with TCE, three charge questions were formulated and addressed for this Alliance for Risk Assessment (*ARA*) project (http://www.allianceforrisk.org/Projects/TCE.html):

1. Develop additional risk assessment guidance on how to interpret the non-cancer endpoint when it is used for deciding clean-up standards or acceptable exposure levels when closing sites.

2. Clarify the issues surrounding the potential developmental cardiac malformations for use in understanding clean-up standards and short term exposure levels.

3. Explore the margin of safety measures used to set the TCE RfC and evaluate if these measures are consistent with the baseline principles developed for determining the RfC.

The following 3 sections of text give a general, and sometimes overlapping, response to each of these questions. Section 4 is a synthesis of significant information from the first 3 sections for use in risk management decisions.

Section 1:

Additional Risk Assessment Guidance On How To Interpret The Non-Cancer Endpoint

Trichloroethylene Non-Cancer Exposure

The evaluation of low-level inhalation exposures to TCE has become a subject of increased concern for regulators and risk managers for contaminated site managers over the past year. This concern has been largely prompted by the final TCE assessment conducted by the U.S.

Environmental Protection Agency (U.S. EPA, 2011a) through its Integrated Risk Information System (IRIS) database. A Reference Concentration (RfC) for TCE of 0.002 mg/m^3 (*i.e.*, 2.0 µg/m^3) was posted on IRIS on September 28, 2011.

Prior to the final TCE assessment regulators either relied upon other sources for an estimate of the safe level for chronic or lifetime human exposures (*e.g.*, the provisional reference concentration, RfC, values of 40 µg/m^3 from EPA's National Center for Environmental Assessment [U.S. EPA, 2001] and the Chronic Inhalation Reference Exposure Level [REL] of 600 µg/m^3 from California EPA's Office of Environmental Health Hazard Assessment [OEHHA, 2007]), or did not quantitatively evaluate inhalation exposures with respect to non-cancer human health endpoints. Consequently, regulatory and public health agencies often evaluated TCE exposures on the basis of the cancer endpoint only. Even when TCE exposures were evaluated on the basis of both non-cancer and cancer endpoints, medium-specific concentrations[1] were typically based upon the cancer endpoint, particularly if the acceptable excess lifetime cancer risk (ELCR) was established at the lower end of the acceptable cancer risk range (*i.e.*, 1 x 10^{-4} to 1 x 10^{-6}).

A comparison of the California EPA REL (600 µg/m^3) and the IRIS RfC (2 µg/m^3) shows that the former value is 300 times greater than the latter. Alternatively stated, the estimate of the safe level of chronic inhalation exposures from IRIS is two and one-half orders of magnitude lower than the analogous estimate from OEHHA. An evaluation of TCE exposures using the final IRIS RfC *in lieu* of the OEHHA REL would result in an estimate of non-cancer hazard (*i.e.*, the hazard quotient, or HQ) that is 300 times greater.

As a result of this change in the RfC value, regulatory and public health agencies using an acceptable ELCR greater than 1 x 10^{-6} for the development of medium-specific concentrations (*e.g.*, ambient air, sub-slab vapor or soil gas) will generally find the non-cancer endpoint (*i.e.*, the RfC) driving these values; at an ELCR of 1 x 10^{-6} (*i.e.*, a lower level of acceptable cancer risk), it is likely that the screening level or RAO will be based upon the inhalation unit risk. For example, the U.S. EPA RSL for residential air is 0.43 µg/m^3, based on the cancer endpoint and a target ELCR of 1 x 10^{-6}; a value that is lower than the RfC of 2 µg/m^3. A residential air screening level based on the cancer endpoint and a target ELCR of 1 x 10^{-5} would be 4.3 µg/m^3; a value greater than the RfC of 2 µg/m^3. Therefore, the acceptable ELCR that regulatory and public health agencies adopt will determine which endpoint drives the assessment, particularly when as single numerical point value of ELCR is identified as the target ELCR for the calculation of a medium-specific concentration.

The selection of a numerical ELCR within the acceptable risk goal range may be established by statute or regulation, or may be selected on a case-by-case basis by risk managers. Risk managers with regulatory and public health agencies are generally experienced with the selection of one or more target ELCR values within the acceptable risk goal range. If, for example, the calculation of a medium-specific RAO based on an ELCR of 1 x 10^{-6} proved to be problematic (*e.g.*, the proposed concentration was below laboratory quantitation limits, near or below background levels, or too impractical to achieve), the risk manager could elect to base the RAO on a higher target ELCR within the acceptable risk goal range.

[1] For example, screening levels concentrations above which further investigation is justified; or remedial action objectives [RAOs], concentrations above which a remedial action is justified.

However, non-cancer assessments have typically been conducted with a fixed target HQ of one. As a consequence, risk managers have not exercised the same flexibility with respect to the establishment of medium-specific concentrations based upon the non-cancer endpoint (*i.e.*, a target HQ of one). This has proven problematic with respect to the setting of screening levels, exposure limits or RAOs for TCE. Risk managers have been compelled to establish medium-specific concentrations based on an HQ of one, or even more conservatively, establish the RfC value itself as the medium-specific concentration. To do otherwise, risk managers would need a comprehensive understanding of the sources of uncertainty associated with the derivation of the medium-specific concentration, and consequently, the relevance of the hazards posed from exposures to concentrations exceeding the RfC.

Three sources of uncertainty associated with implementation of a medium-specific concentration (*e.g.*, an indoor air exposure limit for TCE) may be identified:

1. the uncertainty associated with the derivation of the RfC for TCE;
2. the uncertainty associated with the hazard estimate based on the RfC for TCE;
3. the uncertainty associated with the determination of the exposure concentration; and
4. the uncertainty associated with the determination of the exposure concentration with specific regard to the RfC for TCE.

Each of these sources of uncertainty is discussed below.

Uncertainty Associated with the Derivation of the RfC

Understanding the risk of exposures to TCE at concentrations above the RfC will require a comprehensive understanding of the science and science policy used to set the RfC. This science and science policy evaluation is best understood if one divides the issues into two interdependent but distinct perspectives. The first perspective is toxicology-based, and examines the process and methods used to extrapolate from the animal results to safe human exposure levels. The second perspective focuses more on understanding the actual assessment of risk, how risk is measured and the uncertainty and precision with which numerical expressions of risk can be practically applied.

Examining the toxicology science and science policy used to develop the RfC is generally beyond the scope of most regulatory risk managers' capabilities. The complexity of the process by which the RfC was derived from experimental data from animal assays (including benchmark dose derivation, allometric scaling and physiologically based pharmacokinetic modeling for equivalent human exposure concentration) precludes most risk managers from being able to make confident decisions regarding the human health effects that may be expected at TCE concentrations above the RfC.

In the absence of a toxicology analysis, risk managers should approach an understanding of the exposure from the risk assessment/management side. If a risk based analysis is made, then most risk managers will find some flexibility for making risk based decisions. First, risk managers

should attempt to characterize and have some practical understanding of the range of imprecision and uncertainty in the RfC estimate. This understanding begins with characterizing how precise, or more accurately, how "imprecise" is the numerical value of the RfC (and any screening level that is calculated on the basis of the RfC) that the risk manager is using to assess risk. Numbers are either exact or an approximation. Approximate numbers have some degree of uncertainty associated with them and precision, generally expressed as significant figures, is commonly used to determine when to truncate numerical values (*e.g.*, the RfC or any screening value calculated on the basis of the RfC) (U.S. EPA, 2012b). In general, when performing calculations with approximate numbers, one uses as many digits as possible in calculations until the final result. U.S. EPA considers the precision of a hazard estimate to be one significant figure (U.S. EPA, 1989, Exhibit 8-4, p. 8-9). When determining the final result the rules of truncating and rounding would apply (U.S. EPA, 2012b). Accordingly, any representation of the hazard index (HI) should be one significant figure. For example, an HI calculated to be 0.94 would be rounded to 0.9, and HI calculated to be 1.6, the HI estimate would be rounded to 2.

Uncertainty Associated with the Estimates of Hazard Based on the RfC

The U.S. EPA (2012b) generally represents the RfC as a value rounded to one significant figure. Similarly, U.S. EPA (2012a) defines the use of the HI [or HQ] as rounded to one significant figure. Thus, the precision of any hazard estimate, including any screening level or RAO derived from a target HQ, is limited to one significant figure. In practical application, the hazards associated with an HQ of 0.95 and an HQ of 1.49, and any value which lies between them, is represented as an HQ of 1 (one); differences in hazard between any two estimates within this range cannot be discerned. U.S. EPA (1993) acknowledges and demonstrates practical application of this imprecision when defining the risk in a Site-Specific Record of Decision from two separate HQs of 0.7 and 1.2; here, U.S. EPA (1993, Appendix B) defines both hazard estimates as essentially equal to an HQ of 1 (one), stating the reason is the imprecision in the reference doses for nickel and copper, respectively.

The uncertainty must also be considered. A major portion of uncertainty involved in calculating a HQ is imprecision, but uncertainty enters into any evaluation when one considers that certainty is true knowledge about the meaning of the data and this is rarely the case when U.S. EPA determines a value for the RfC. U.S. EPA (IRIS, 2012) defines the TCE RfC as "possibly having up to" an order of magnitude uncertainty. There are many ways to characterize an order of magnitude of uncertainty. Guidance is taken from U.S. EPA citing that the order of magnitude is "centered" on the RfC (U.S. EPA, 2002). Further guidance on how to apply the centered order of magnitude is taken from Felter and Dourson (1998), who state many on the EPA RfD/RfC workgroup believe the range of uncertainty amounts to one-half an order of magnitude on either side of the RfC.

Therefore, the uncertainty and imprecision in the RfC can vary from one significant figure up to an order of magnitude. This imprecision and uncertainty must be defined relative to the issues of most concern to the risk manager as demonstrated by the following:

1. U.S. EPA screening levels are not RAOs (Regional Screening Level Table, 2012) but are meant to define exposures under which further investigation is warranted. Therefore

screening levels are not meant to define effects level, *e.g.*, levels at which a toxic effect is expected to occur. Screening levels are meant to define a bright line for further investigation. Defining a bright line for further investigation is much different than defining a bright line for actual health effects. Thus, the risk managers' question becomes: at what exposure level above screening does the risk manager become concerned with a significant or measurable increase in potential for a toxic effect?

2. The use of the RfC term in various equations to determine screening levels or assess the risk of exposure must consider the precision with which such equations or methods can be applied and understood. U.S. EPA's designation for the RfC as rounded to one significant figure (IRIS, 2012) is a clear acknowledgement of the limitations implicit in the precision of the RfC. Therefore, it is not possible to determine the difference in hazard between an HQ of 0.95 and an HQ of 1.49, both of which would round to 1 (one).

3. RAOs are distinct from screening levels. RAOs should therefore be indicative of concentrations that are protective of human health, *i.e.*, concentrations that do not pose a "measurable and quantifiable hazard or risk". The RAOs (like screening levels) should be below effects levels, although RAOs may be at concentrations that are above the screening level. If the risk manager cannot demonstrate that a concentration exceeding a screening poses a measurable and quantifiable hazard or risk, then there would be no need for remedial action.

Thus, due to the imprecision of the methods used to develop the RfC, the risk manager cannot reliably show a measurable risk-based difference between an HQ value of 0.95 and 1.49. If this were true, then the difference in hazard estimates at levels of precision beyond the single significant figure has little meaning. Using the common equation for screening and remedial action (RSLTs, 2012) and calculating exposure levels associated with a target HQ (THQ) of 1 up to 1.49 (which rounds to 1) yields:

$$Screening\ Level\ (\frac{ug}{m^3}) = \frac{THQ\ x\ AT\left(24\frac{hrs}{day}x365\frac{days}{yr}x\ 30\ yrs\right)x\ 1000\ ug/mg}{EF\left(350\frac{days}{yr}\right)x\ ED\ (30\ yrs)x\ ET\left(24\frac{hrs}{day}\right)x\frac{1}{RfC\ (\frac{mg}{m^3})}}$$

where THQ = 1.0, Screening Level = 2.1 µg/m^3, or 2 µg/m^3 when rounded to one significant figure; and

where THQ = 1.49, Screening Level = 3.1 µg/m^3 , or 3 µg/m^3 when rounded to one significant figure.

Thus, it is not possible to discern a discrete difference in the hazard associated with an air concentration of 2 µg/m^{3}and an air concentration of 3 µg/m^3. A similar analysis can be undertaken by calculating the order of magnitude risk range centered on the RfC. While there can be different methods used to assess "half an order of magnitude" around the RfC, the use of the following, after Felter and Dourson (1998), seems the most reasonable and is expressed using the practical methods of assessing exposure in µg/m^3 as:

$10^{0.5} = 3.16$, *rounding to one significant figure, and a factor of 3*

$where\ \frac{RfC}{3} = low\ end\ of\ the\ range, and$

$3\ (RfC) = high\ end\ of\ the\ range$

$The\ range\ of\ the\ TCE\ RfC\ uncertainty\ is\ =\ 0.6\ \frac{\mu g}{m^3}\ to\ 6\ \frac{\mu g}{m^3}$

Thus, the RfC is simply the center that sets the boundaries of the order of magnitude range. U.S. EPA (2002) states how the range should be applied:

"….. [The] reference value is intended to provide an estimate that is centered within an order of magnitude, further emphasizing that the estimate is not a bright line, but has some range of variability that may be considered by risk managers in decision making."

Recognizing there is some flexibility, the risk manager is still in the position of trying to understand what the range of uncertainty around the RfC means in practical reality. Here is where information is used on how the range is "weighted." That is, if the magnitude of the composite uncertainty factor used to determine the RfC is large, then it is likely that the hazard associated with toxic effects above the RfC would shift to the high side of the RfC range. Similarly, use of a composite uncertainty factor that is somewhat marginal would indicate the risk of toxic effects could shift to the central RfC value or the low side of the range.

Toxicology Excellence for Risk Assessment (TERA, 2012) clarifies this concept stating: "Moreover, the precision of the RfD or RfC depends in part on the overall magnitude of the composite uncertainty and modifying factors used in its calculation. The precision at best is probably one significant figure and more generally an order of magnitude, base 10. As the magnitude of this composite factor increases, the estimate becomes even less precise."

Another supporting perspective can be determined using U.S. EPA guidance on how the RfC for TCE was determined (U.S. EPA, 2011a). Here U.S. EPA used a set of three studies at the lowest end of the response spectrum. Three candidate RfC values (*i.e.*, cRfCs) at the lower end of the

responses were grouped in the range of 0.0003-0.0006 ppm and were selected to derive and support the final RfC. The grouping establishes a narrow range of candidate RfC values. U.S. EPA used the lower portion of this range of values to set the RfC. U.S. EPA states the reasoning and justification for this approach (U.S. EPA, 2011a; Section 5.1.5.2, page 5-95):

"One approach to selecting an RfC would be to select the lowest calculated value of 0.0003 ppm for decreased thymus weight in mice. However, as can be seen in Table 5-24, three p-cRfCs are in the relatively narrow range of 0.0003–0.0006 ppm at the low end of the overall range. Given the somewhat imprecise nature of the individual candidate RfC values, and the fact that multiple effects/studies lead to similar candidate RfC values, the approach taken in this assessment is to select an RfC supported by multiple effects/studies. The advantages of this approach, which is only possible when there is a relatively large database of studies/effects and when multiple candidate values happen to fall within a narrow range at the low end of the overall range, are 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."

IRIS agreed with this interpretation and concluded similarly (U.S. EPA IRIS, 2012; Section I.B.2). U.S. EPA did not signal that a developmental endpoint was used to set the RfC. Instead, they cite a group of studies at the low end of the response spectrum and treat them as a group. The U.S. EPA used 0.0004 ppm from the range of 0.0003 to 0.0006 ppm to set the RfC. Given adequate site-specific exposure and source information, it is reasonable to use the higher portion of the range to set remedial objectives. Using the 0.0006 ppm value (*i.e.*, the upper value from the range of candidate RfC values, or 3.26 µg/m^3), and a target HQ of one, a screening level of 3.3 µg/m^3 (or 3 µg/m^3 when represented at one significant figure) may be calculated by using the equation presented above. This estimate is equivalent to the screening level calculated on the basis of a target HQ of 1.49 (discussed above).

An RAO may be calculated as shown below (*i.e.*, by the same equation as the screening levels were calculated above):

$$Remedial\ Active\ Objective\ (RAO)\ (\frac{\mu g}{m^3}) = \frac{THQ\ x\ AT\ \left(24\ \frac{hrs}{day} x 365 \frac{days}{yr} x\ 30\ yrs\right)\ x\ 1000\ \mu g/mg}{EF\ \left(350 \frac{days}{yr}\right) x\ ED\ (30\ yrs) x\ ET\ \left(24 \frac{hrs}{day}\right)\ x\ \frac{1}{RfC\ (\frac{mg}{m^3})}}$$

as an indicator of an air concentration that is protective of human health, *i.e.*, concentrations that do not pose a "measurable and quantifiable hazard or risk". As stated above, the RAOs (like screening levels) should be below effects levels, although RAOs may be at concentrations that are above the screening level. For purposes of illustration, a value for an RAO may be calculated using the upper-bound value on the range of uncertainty centered on the RfC (*i.e.*, 6 µg/m^3, discussed above) and a target HQ of 1.49 (to account for the implicit precision of the

hazard estimate. This calculation would result in an RAO for residential air of 9.3 $\mu g/m^3$, or 9 $\mu g/m^3$ when rounded to one significant figure. This analysis suggests that a TCE air concentration of 9 $\mu g/m^3$ may not be any less protective of lifetime exposures than the Screening Level of 2.1 $\mu g/m^3$, based on both the uncertainty associated with the RfC and the precision of the target HQ. Therefore, TCE concentrations above the screening level of 2.1 $\mu g/m^3$ merits further investigation; this evaluation suggests that. In a particular circumstance, TCE concentrations as high as 9.3 $\mu g/m^3$may not require remedial action.

Another way to represent the uncertainty that is implicit in the RfC is to evaluate the alternate points of departure that may be used to derive the RfC based on the fetal malformation endpoint (using the $BMDL_{01}$, BMD_{01} as alternate points of departure); and the candidate RfC based on the toxic nephropathy endpoint (using the $BMDL_{05}$, BMD_{05} as alternate points of departure). For each alternate POD, a Screening Level based on the equation above was derived for residential and industrial exposures. Table 1 summarizes the range of TCE screening levels that are calculated based on alternate representations of the fetal cardiac malformation and toxic nephropathy endpoints from section 3, for comparison with indoor air data. The values presented in Table 1 represent point estimates corresponding to a HQ of 1 and centered on the order-of-magnitude uncertainty associated with an RfC, and the imprecision associated with rounding of the HQ value to one significant figure.

Table 1. TCE Indoor Air Screening Levels Based on Alternate Representations of Endpoints

Endpoint	Value (ppm)	Value ($\mu g/m^3$)	Screening Level - Residential ($\mu g/m^3$)	Screening Level - Industrial ($\mu g/m^3$)
RfC, fetal malformation endpoint	0.00037	2	2.1	8.7
BMDL, 1% increased risk on the fetal malformation endpoint	0.004	21	22.3	93.9
BMD, 1% increased risk on the fetal malformation endpoint	0.011	59	61.5	258.1
Candidate RfC, toxic nephropathy	0.00056	3	3.1	13.1
BMDL, 5% increased risk on toxic nephropathy	0.014	75	78.2	328.5
BMD, 5% increased risk on toxic nephropathy	0.02	107	111.7	469.3

In conclusion, what is needed is a clearer toxicological understanding of the range, how it should be weighted and any other insight toxicology experts can provide for risk managers regarding exposure levels and risk of a toxic effect above the RfC. However, any understanding of risk estimates must be interpreted in light of the imprecision and the uncertainty that are a part of the RfC derivation process. The use of significant figure recommendations and uncertainty estimates are parameters that risk managers can consider now, in the absence of any other guidance. At a minimum, without any further toxicology understanding, the differences between exposure levels for TCE between 2 and 3 $\mu g/m^3$ are really not measurable and an increase in calculated risk is really not observable until exposure levels reach at least 3 $\mu g/m^3$.

Uncertainty Associated with the Determination of the Exposure Concentration: General Considerations

Inhalation exposures to TCE are a concern in indoor air, particularly when the TCE concentrations may be largely or partially attributable to vapor intrusion. The indoor air concentrations of TCE may be predicted by modeling or directly measured by air sampling. Air sampling reflects all of the factors in the vapor intrusion pathway (source strength, subsurface characteristics and building conditions), and, of the types of vapor intrusion sampling that can be performed, most effectively characterizes the variability in potential human exposure inherent in the vapor intrusion pathway. Air sampling is a more direct indication of potential exposure and risk to building occupants from VOCs in indoor air arising from vapor intrusion, when compared

to modeled indoor air concentrations predicted from sub-slab vapor, soil gas, or groundwater concentrations. This last point is critical, because the collection of air sampling data often is required by a regulatory agency to assess risk and make a determination of whether or not a vapor intrusion pathway in a building is of regulatory concern and warrants action to reduce exposure.

It is well-recognized that indoor air concentrations and potential exposures vary, whatever measurement period is chosen and whatever measures are undertaken to keep conditions constant. This variation is usually much greater than can be accounted for by sampling and analytical error. Typically, air sampling results peak at a relatively low concentration, with a few results several times the most frequently detected concentration. This has been observed in multiple fields such as indoor radon assessment and industrial hygiene investigations. This pattern of results can also be observed in indoor air sampling data collected from vapor intrusion sites.

Typical indoor air sampling strategies at vapor intrusion sites typically measure indoor air concentrations during a few days throughout a year, usually at times selected based on practical concerns or under assumed worst-case conditions. These worst-case conditions typically occur under winter conditions, where higher indoor air temperatures produce a stack effect that enhances advective transport of soil vapor indoors, and where the outdoor air exchange rate is minimized to reduce heating costs. Sampling may also be performed in the other seasons to provide some understanding of variability in indoor air concentrations. Evaluation of the data typically involves comparison of individual measured concentrations with risk-based screening levels. Reported concentrations higher than the screening level represents a trigger for further investigation or implementation of mitigation measures; depending on the regulator, a single result exceeding a screening level may be sufficient to trigger further action.

Statistical evaluation of indoor air sampling data becomes possible when multiple samples can be collected from a building (or from a group of buildings judged to have similar characteristics). Statistical methods can provide a more refined assessment of potential exposure than a "worst-case" evaluation. Performing statistical analysis on air sampling data requires adopting a model of the distribution of concentrations and associated exposures. Historically, a lognormal distribution is used to summarize air concentration and exposure data. Lognormal distributions are generally leptokurtic, or positively skewed (a long "tail" to the right). Presenting indoor air sampling data in this manner allows the variability in exposures to be combined with dose-response variability for purposes of risk characterization.

Some data are available regarding the expected variability in indoor air concentrations associated with vapor intrusion (VI). U.S. EPA has developed a database of VI sampling results, principally from residential structures and including indoor air sampling data (U.S. EPA, 2012b). The indoor air sampling data can be used to develop an understanding of between-structure variability. For example, U.S. EPA's report states that residential structures at Lowry Air Force Base exhibit almost two orders of magnitude variability in indoor air concentrations measured monthly over a period of six months. Folkes *et al*. (2009) determined temporal variability using up to 10 years of data at 45 unmitigated structures and concluded that concentrations vary by a factor of two to three from an annual average concentration. Folkes *et al*. indicated that seasonal variations explain a portion of the variability but the majority of the variation at the study site appeared to be due to factors such as variation in air exchange rates and differential pressure

caused by wind, barometric pressure, temperature and the use of exhaust fans, HVAC and open windows. This is consistent with the conceptual understanding of factors that can impact variability.

Uncertainty Associated with the Determination of the Exposure Concentration: Specific Implications for Comparisons of Measured Air Concentrations to the RfC for TCE (or to Screening Levels or RAOs Based on the RfC)

The RfC for TCE on IRIS was set by U.S. EPA (2011a), in part, using the critical effect of the Johnson *et al*. (2003) study. This study described a developmental endpoint of fetal heart malformations. The critical effect in this study occurred across a short-term exposure period during gestation from 1-21 days in Sprague-Dawley rats. Because the toxic effect occurs across a short dosing period, some states (*e.g.*, New Jersey, Massachusetts) and U.S. EPA regions (*e.g.*, Region IX, Region X) have developed exposure levels and sampling guidance commensurate with the short exposure (*i.e.*, developmental) time frame. This has raised many questions and concerns regarding how to assess indoor air exposure levels (based on the estimate of reasonable maximum exposure, or RME), given the inherent spatial and temporal variability of indoor air concentrations. If U.S. EPA and the state environmental and public health agencies are going to base compliance and remedial decision-making regarding TCE concentrations in indoor air using a developmental endpoint across a short period of time varying from days to months, then a conceptual understanding and characterization of short term spatial and temporal variability is critical.

Given the range of the expected variability in indoor air concentrations, it now becomes important to design a sampling program that will adequately assess TCE indoor air concentrations in the relevant time frame (*i.e.*, the exposure concentration). Since the exposure period used to determine the fetal heart malformation critical effect was 21 days, it seems reasonable to assess short term variation across a corresponding time window, (*i.e.*, 21 days, for which average concentrations of a month or less may be reasonably representative). The use of indoor air samples taken in the main living areas, and spatial averaging may be appropriate for assessing spatial variability in indoor air concentrations. However, accounting for temporal variability may pose a much more difficult problem; temporal variability across a 21-30 day time window is superimposed upon other cycles of temporal variability (*e.g.*, diurnal, seasonal, and annual variability). Alternatively stated, the indoor air exposure concentration that accounts for temporal variability across 21-30 days in April may not be representative of the indoor air exposure concentrations for the same receptor population in the same structure across 21-30 days in July, October or January. Thus, determination of the indoor air exposure concentration must account for spatial variability and multiple cycles of temporal variability (*i.e.*, diurnal, seasonal, annual) superimposed upon the relevant 21 day window.

The common method of assessing variability has been to focus indoor air sampling at times that are expected to be representative of worst-case conditions. Many regulatory programs assume worst case conditions occur during cold weather periods when HVAC systems provide a negative pressurization of indoor air compared to ambient pressures, and this pressure differential then facilitates the transport of vapor from the subsurface to the indoor air. Other regulatory programs do not assume cold weather conditions will provide worst case conditions

and require sampling during at least two seasons. Generally, this type of sampling entails a single 24 hour "snapshot" sample which does little to assess the temporal variability associated with seasonal or annual cycles, unless one assumes that the one-day sampling event was positively biased (*i.e.*, conducted under expected worst case conditions). The problem with determining any "variability" in indoor air is that it is difficult to determine when all of the factors impacting variability will combine to provide a scenario that may be construed to be "worst case conditions". Thus, indoor air exposure concentrations to represent the 21-day exposure period ideally should be based upon multiple sampling events which are representative of variations in temperature, water table elevation, and building-specific parameters such as pressure differential and ventilation rate. Adequately accounting for this variability may likely require that indoor air sampling be conducted on two or more 21-30 day events (*e.g.*, winter, summer) throughout a single year. As discussed previously, statistical evaluation of indoor air sampling data becomes possible when multiple samples can be collected from a building (or from a group of buildings judged to have similar characteristics).

Given the sources of spatial and temporal variability discussed above, there are some possible approaches to determine exposure levels across a 21-30 day window within a structure include:

1. Collect multiple indoor air 24-hour TO-15 type composite samples randomly selecting the 24-hour periods within a 21-30 day time window. Then use statistical software to determine an exposure point concentration for that window.[2]

2. Use passive diffusion samplers for a 21-30 day time window that provide the "average" exposure concentration over that window.

3. Use real-time sampling such as an onsite GC or GC/MS recording readings at some frequency throughout each 24 hour period for a 21-30 day interval and then determine the UCL as above.

4. Take a one-time 24 hour sample and apply some pre-determined variability factor such as a factor of three or 10 (note: the concept of a variability factor has a parallel with seasonal adjustment factors used in indoor radon sampling).

There are advantages and disadvantages to each of the four methods listed above. The first three approaches are presented with respect to one sampling events; each of these methods would need to be repeated at least during the year to account for expected seasonal variation. The cost of multiple 24-hour composite samples is likely to be prohibitive in many applications. The use of passive or active samplers does not distinguish between indoor air vapor intrusion and background contributions. As noted by U.S. EPA (2011b) indoor air background can often be similar in concentration to the regulatory action level. Cost of real-time on-site GC or GC/MS sampling may also be prohibitive, however, it should be noted that this method has the real advantage of identifying peaks and also providing valuable information about potential indoor air background. It appears the most cost effective way to address temporal variability may be to use

[2] For example, U.S. EPA Pro-UCL (U.S. EPA, 2013) can be used to determine the 95% upper confidence limit on the mean concentration (*i.e.*, the 95% UCL, representative of the reasonable maximum exposure (RME) concentration, as defined by U.S. EPA, 1989). Other software packages also may be used to develop a representative concentration.

passive indoor air samplers over a few 21-30 day periods throughout the year. The objective of this approach is to approximate the worst case 21-day indoor air exposure concentration and account for temporal variability.

There are continuing efforts addressing the questions about how to best address variability when devising strategies for assessing vapor intrusion. A recent EPA workshop, "Looking Beyond Natural Variation in Vapor Intrusion: Understanding, Controlling and Addressing Site Variables for Improved Practices and Effective Protection Strategies" (https://iavi.rti.org/WorkshopsAndConferences.cfm), was convened in March 2013 to examine different aspects of spatial and temporal variability based on research in single buildings and studies of data from multiple buildings and sites. Efforts such as this workshop will contribute to evolving best practices for sampling and investigating indoor air at vapor intrusion sites.

Section 2:

Issues Surrounding The Potential Developmental Cardiac Malformations

Risk Management Considerations for the Evaluation of a Developmental Toxic Effect for TCE at Exposure Levels above the RfC

The intent of this section is to gain a broader understanding of the level of confidence in the weight of evidence linking TCE exposure with fetal heart malformations and the overall margin of safety built into the process that defined the RfC. This information will provide a more qualitative understanding of the risk from various TCE exposure levels and provide a higher level of confidence when making site-specific decisions on acceptable exposure levels.

U.S. EPA (2011a) evaluated a large number of studies for derivation of the RfC. For each study they developed a "candidate RfC", as if each study were being used as the principal RfC study. Three candidate RfCs were close in magnitude and as a group, well below other candidate RfCs. The low grouped studies were the Johnson *et al*., (2003); Keil *et al*., (2009); and, NTP , 1988 studies. EPA (2011a) then decided two studies were the main critical effects studies (Johnson *et al.*, 2003 and Keil *et al.*, 2009) and used the average of the two main critical effects studies to derive the RfC, citing the third study (NTP, 1988) as support for the two main critical effects studies.

One critical effects study was linked with developmental effects (*i.e.*, fetal heart malformations as described by Johnson, *et al.*, 2003). The other two studies were longer-term (Keil *et al*., 2009) and chronic (NTP, 1988) exposure studies and the critical effects, decreased thymus weight and kidney toxicity, were associated with long-term exposure. Although no short term or developmental effects were reported in either the Keil *et al*. or the NTP studies, short-term exposures are of concern with respect to the developmental effect reported by Johnson *et al*. (2003). The nature of this toxic effect is serious and many risk managers nationally have become concerned about the risk for fetal heart malformations at exposures above the RfC.

There are two major methods used to evaluate the risk of a developmental effect. The first method is quantitative and addresses the magnitude of acceptable exposure levels for both acute and chronic exposures. These responses are contained in the RfC range discussion included in the "Management Considerations of Noncancer Risk Value Ranges Using Trichloroethylene as an Example" section of this document as response to Charge Question 1.

The second evaluation method is more qualitative or semi-quantitative in nature, addressing four main considerations:

1. The weight of evidence that links exposure with the toxic effect examining both the strength of the epidemiology and animal evidence linking fetal heart malformations to TCE exposure

2. Information on how other countries regulate TCE and how they address the risk of a developmental effect

3. The margin of safety incorporated into entire RfC derivation process.

4. The use of the critical effects study: Johnson *et al*. (2003).

Weight of Evidence

U.S. EPA in their 2011 Toxicological Profile provides considerable information on the evidence for TCE exposure and cardiac malformations. U.S. EPA concludes, based on weakly suggestive evidence, but overall consistent epidemiological data in combination with evidence from experimental and mechanistic studies that TCE exposure poses a potential hazard for congenital malformations (U.S. EPA, 2011a: pgs. 6-10, 11). EPA's conclusions were based on the following:.

1. EPA cites three principal human studies indicating an increased risk of cardiac malformations: ATSDR (2006a; 2008 two studies same population) and Yauck *et al*. (2004).

    - ATSDR concluded that there was an increased risk of heart malformations in the exposed population among Endicott, New York residents living in the area where volatile organic compounds (VOCs) were found in soil vapor.

    - The Yauck *et al*. (2004) study was conducted on children born between 1997-99 with congenital heart defects in Milwaukee, Wisconsin, using TCE emissions data from industrial sources and distance between the emission source and maternal residence within 1.3 miles of the source. Yauck *et al*. concluded that offspring of women >38 years of age showed an increased risk of congenital heart defects while offspring of women <38 years of showed no increased risk.

2. A series of animal studies was used to support EPA's opinion. EPA (2011a) cited a number of animal studies where fetal cardiac malformations occurred after oral dosing of TCE during gestation (Dawson *et al*., 1993, 1990; Johnson *et al*., 2005, 2003) and fetal heart malformations from oral dosing during gestation with the metabolites of TCE (Johnson *et al*., 1998b; Johnson *et al*., 1998a; Epstein *et al*., 1992; Smith *et al*., 1992; Smith *et al*., 1989). Additionally, EPA (2011a) cites support for fetal heart malformations from avian studies (Rufer *et al*., 2008; Drake *et al*., 2006a; Drake *et al*., 2006b; Mishima *et al*., 2006; Boyer *et al*., 2000; Loeber *et al*., 1988; Bross *et al*., 1983). The Rufer *et al*. (2008), Drake *et al*. (2006a, 2006b), Loeber *et al*. (1988), and Bross *et al*. (1983) are all studies where TCE was injected into the developing egg and cardiac malformations observed. Mishima *et al*. (2006)

and Boyer *et al*. (2000) were studies presenting a plausible mechanism for fetal cardiac malformations in chick embryos

3. EPA concludes that although the study used to set the developmental endpoint has significant limitations, there is insufficient evidence to dismiss their findings. The EPA Science Advisory Board (2011a) recommended the two endpoints for immune effects from Keil *et al*. (2009) and the cardiac malformations from Johnson *et al*. (2003) to be considered the principal studies supporting the RfC.

It is important to evaluate the supporting evidence in more detail and also to consider many additional lines of evidence.

1. ATSDR (2006a, 2008) used strongly cautionary language in the interpretation of the results of their two studies; stating that while risk was significantly elevated, an extremely small number of infants born with birth defects was observed. ATSDR notes (2008) that smoking and women working in the electronics industry confounds the results. The elevated smoking rates likely contributed to some of the outcomes and that workplace chemical exposure was also likely.

2. U.S. EPA also concluded that the evidence for an association between TCE exposures in the human population and the occurrence of congenital cardiac defects is not particularly strong. Many of the epidemiological study designs were not sufficiently robust to detect exposure-related birth defects with a high degree of confidence.

3. The only animal studies conducted on TCE exposure and cardiac malformations came from one study facility. Other studies were done with the metabolites of TCE or were direct injections into developing avian embryos and did not include maternal exposure. It is difficult to determine the relevance of direct embryo injection to the weight of evidence.

It is apparent that the significant weight of evidence comes from animal studies done by one facility. There is a considerable weight of evidence that the fetal cardiac malformations are not a relevant inhalation critical effect. In particular:

1. Studies evidencing fetal heart malformations (Johnson *et al*., 2005, 2003; Dawson *et al*., 1993, 1990) were limited to one testing facility. Following these results, two studies were designed to duplicate the fetal cardiac malformations results of Johnson *et al*. (2005, 2003) and Dawson *et al*.(1993, 1990) but were unable to detect fetal cardiac anomalies. In one study by Carney *et al*. (2006), TCE was administered via inhalation and in the second study (Fisher *et al*., 2001), TCE was administered via oral dosing. Johnson collaborated on the Fisher *et al*. (2001) study. The inability of the guideline, GLP-quality study (Carney *et al*., 2006) and the study by Fisher *et al*. (2001) both designed to reproduce the Johnson *et al*. (2003) results is significant. Despite considerable effort to reproduce the results of the Johnson work, the fetal heart malformation critical effect does not appear to be reproducible outside one testing facility.

2. EPA (2011a) concludes that a number of well conducted epidemiology studies did not report a significant increase in fetal cardiac malformations (Bove, 1996; Bove *et al*., 1995; Goldberg *et al*., 1990; Lagakos *et al*., 1986).

3. A significant number of animal studies did not report induction of cardiac effects in rats by inhalation (Carney *et al*., 2006; Healy *et al*., 1982; Dorfmueller *et al*., 1979; Schwetz *et al*., 1975), in rabbits by inhalation (Hardin *et al*., 1981), in rats by gavage (Fisher *et al*., 2001; Narotsky and Kavlock, 1995; Narotsky *et al*., 1995) or mice by gavage (Cosby and Dukelow, 1992).

4. EPA (2011a: pg 4-560) indicated that the only identified cardiac "anomalies" following gestational exposure to TCE or its metabolites were conducted in rats and dosed by the oral route of exposure (gavage or drinking water).

Worldwide TCE Regulation

It is also important to consider how other countries and regulatory entities regulate TCE as concerns cardiac malformations. The following table considers some national and global perspectives on TCE exposure and Fetal Cardiac Malformations as a viable critical effect.

Table 2: Regulation of TCE for a Developmental Critical Effect

Regulatory Entity	**Year**	**Fetal Cardiac Malformations (FCM) as critical effect**
European Union	2009	Did not define FCM as critical effect
National Research Council	2006	Notes FCM came from single laboratory, flat dose-response, study needs to be repeatable in another laboratory
National Institute of Public Health and Environment (RIVM)	1999, 2001	Did not consider cardiac effect or Dawson *et al*. (1993) study
Cal-EPA	2009	Johnson FCM was not a meaningful or interpretable dose-response relationship, not consistent with other work. Re-evaluating in consideration of 2011a U.S. EPA RfC
WHO	2005, 2008, 2009, 2010	Notes Johnson *et al*. did not show a clear dose-response as might first appear on closer examination of the data, but corroborated the preferred Dawson *et al*. that showed a clearer dose-response relationship.
Health Canada	2005, 2007	Notes cardiac anomalies reported in Dawson *et al*. (1993) were corroborated by Johnson *et al*. (2003), but dose–response from Johnson *et al*. is not as clear as might first appear on closer examination of the data; endpoint deserves close scrutiny, but deemed Dawson *et al*. more appropriate as the key study, because it showed a clearer dose–response relationship

NICNAS (Australian *National Industrial Chemicals Notification and Assessment Scheme)*	2010	Adopted WHO position on the developmental effects of TCE

Margin of Safety Analysis

An understanding of the risk from a developmental effect from exposure above the RfC can be supported with an analysis of the margin of safety built into the entire process of the derivation of the fetal health malformation critical effect candidate RfC. At each decision point in the process for the candidate RfC there was uncertainty. EPA responded to this uncertainty by incorporating a safety factor. For instance, when extrapolating from the animal study to human exposure, EPA used a 1% response rate as opposed to the more common 5% or 10% response rates. Each decision point incorporates a similar safety factor whose sum, considered as a whole, provides an overall awareness of the entire margin of safety.

EPA used a considerable margin of safety in deriving the candidate RfC from the Johnson *et al*. (2003) evidencing the cardiac malformations.

1. Physiologically based pharmacokinetic (PBPK) modeling was used to determine the internal or absorbed dose in the animal resulting from the administered dose. It is reasonable to assume the model outputs were conservative and provided an underestimation of the true value. The output of this model is used as the input to a second model.

2. The second model, a dose-response model, was used to determine the response rate protective of 99% of the dosed population; termed the Benchmark Dose for a 1% Response rate (BMR 1% or BMD_{01}). The BMR used was 1%; the most protective response rate. It is common to use a BMR of 10% (U.S. EPA, 2000). Response rates are tied to study sensitivity and a 1% BMR is reserved for the most sensitive. By comparison, nested reproductive or developmental studies are considered to be sensitive and they generally use a BMR of 5% (U.S. EPA, 2000). Casarret and Doull (2001) point out that the most extensive studies with developmental toxicity indicate that BMD_{05} (a BMR of 5%) is similar to a statistically derived NOAEL for a wide range of developmental toxicity endpoints. Thus, the 1% response rate would add a considerable margin of safety above what is normally used.

3. Confidence bounds were placed on the $BMR_{1\%}$ value for a BMDL or Benchmark Dose Lower Confidence Limit. That is, the 95% lower confidence bound was placed on the BMD of 1% to obtain a BMDL. The ratio of the BMD to the BMDL was about three (EPA, 2011a: pg 5-45). Thus, use of the lower confidence limit on the BMD introduces a factor of three margin of safety

4. Next, the animal BMDL is converted to a Human Equivalent Concentration (HEC_{99}) at the 99^{th} percentile using a third model. That is, the BMDL modeling projects the lower

confidence bound on a level at which 99% of exposed study population would not have a sufficient internal dose to elicit a toxic response; then, a third model was used to determine the HEC of a human dose that is protective of 99% of the human population. The difference between the HEC_{50}, what the average population would experience, and the HEC_{99} which is protective of 99% of the population, was also about a factor of three.

5. It is important to recognize that the output of each of the models is the input for the next. It is reasonable to assume each model is developed with a margin of safety that under-predicts the true value. If the first model under-predicts the true value and its output is used as an input into the next successive model, which also under predicts the true value, then the margin of safety is compounded. This occurs for three successive models.

6. An additional factor of 10 margin of safety is added.

Section 3:
Margin Of Safety Measures Used To Set The TCE RfC: Hazard Ranges Above the Reference Concentration (RfC)

Introduction

Waste sites around the world often contain chemicals for which target safe levels are exceeded. In cases where the target safe levels are established for chemicals identified as known or potential carcinogens, the acceptable cancer risks are generally presumed to fall within an upper bound, excess cancer hazard range of 10^{-6} to 10^{-4}. Screening levels are typically set at a point value at the lower end of this range (*e.g.,* 10^{-6}) and risk management decisions regarding the implementation of remedial actions (if any) are made within this range based on various considerations (*e.g*., long-term effectiveness, mass reduction of contaminant, cost, community acceptance of the proposed remedial action). In the case of noncancer toxicity, few if any hazard ranges are determined, primarily because the evaluation of noncancer hazards is predicated upon the determination of a "safe dose" below which no toxic effect is expected, rather than upon an acceptable (if non-zero) probability that a cancer-causing event may occur.

The National Academy of Sciences (2009) suggested the development of methods for noncancer toxicity that have the capability of determining hazard ranges. The Alliance for Risk Assessment (*ARA*) project entitled "Beyond Science and Decisions: From Problem Formulation to Dose Response" is a response to this NAS suggestion, and six of its case studies attempt to do this. These case studies are:

- Comparison of Hattis strawman approach and BMDs/UFs for noncancer endpoints (carbonyl sulfide and tetrachlorobenzene) by Greco *et al*.
- Implications of Linear Low-Dose Extrapolation from Benchmark Dose for Noncancer Risk Assessment by Kroner and Haber.
- Review and application of data fusion methodologies for toxicological dataset analysis to resolve data quality issues in predictive toxicology and contaminated sites risk assessment by Mohapatra *et al*.

- Estimate Risk Above the RfD Using Uncertainty Factor Distributions by Spalt and Kroner.
- Use of biomarkers in the benchmark dose method Gentry *et al*.
- Use of Categorical Regression – Risk Above the RfD by Danzeisen *et al*.

The Science Panel (http://www.allianceforrisk.org/Workshop/Panel.htm) of this workshop series vetted each of these case studies and found five of them to have merit. [3] Of these five, we focus on 2 for the purposes of this work, specifically:

- Use of biomarkers in the benchmark dose method; and
- Estimate Risk Above the RfD Using Uncertainty Factor Distributions

The remaining three case studies might be used at a future time.

Use Of Biomarkers In The Benchmark Dose Method

Method

The case study has been adapted to use with experimental animal data, and takes EPA (2011a) determinations of benchmark dose (BMD) and the lower 95% confidence interval on the benchmark dose (BMDL) for critical endpoints of TCE at face value. The original case study is an extension of the BMD method that allows the development of a hazard range at doses above the Reference Dose (RfD) when the existing data are based on human responses.

The appropriate BMD is chosen in the usual fashion using existing EPA software and criteria, including *p*-values, visual fit, residuals, BMD to BMDL ratios and Akaike's Information Criterion (AIC). The data are modeled to an appropriate POD using the usual judgment, and then four different procedures were investigated to determine the risk range. These procedures are:

1. A straight line is drawn from both the chosen BMDL and BMD to the RfC, where the RfC is considered to be without risk;
2. The chosen model, rather than a straight line from the BMD/L, is extrapolated to the RfC and then the risk at the RfC is considered to be zero;
3. The appropriate BMD model is extrapolated to the RfC and the risk at the RfC is considered to be no greater than its upper bound; and
4. The appropriate BMD model is extrapolated using a threshold term, where the threshold value is judged to be the RfC, or some higher value.

[3] The Science Panel discourage the use of Implications of Linear Low-Dose Extrapolation from Benchmark Dose for Noncancer Risk Assessment for purposes other than for comparison amongst chemicals, because the modeled results did not appear to be consistent with a general understanding of noncancer toxicity.

Results

Results of modeling procedure 1 for fetal heart malformations are shown in Table 3 and Figures 1 and 2. Here the hazard range is determined from a straight line drawn from each of two candidate PODs, *i.e.*, the $BMDL_{01}$ (the BMDL at a predicted response 1% above background) and BMD_{01} (the BMD at a predicted response 1% above background) to the RfC, where the RfC is considered to be without risk (*i.e.*, 100% probability that a safe human dose has not been overestimated). The $BMDL_{01}$ and the BMD_{01} are the median human equivalent concentrations (HECs) taken from EPA (IRIS *Toxicological Review of Trichloroethylene,* Appendix F, p. F-35, 2011a) and adjusted by a three-fold uncertainty factor to account for the expected differences in toxicodynamics when extrapolating from the experimental animal (rats) to the human, thereby accounting for the potential overestimation of the safe human dose based on observed effects in test animals. Using the BMD methodology, a dose-response curve is associated with the derivation of each POD (*i.e.*, the BMD_{01} and $BMDL_{01}$); each dose-response curve represents the range of the expected human response, with the low dose region of each curve reflecting sensitive individuals. EPA's HEC_{99} value (*i.e.*, the human equivalent concentration based upon an individual at the 99th percentile of the range of predicted human exposures, based on toxicokinetic variability in the human population) is specifically not used in this procedure, since its use conflates the results of procedure 1, specifically the use of the HEC_{99} would lower the BMD/L_{01} by an undetermined amount.

Modeling procedures 2 and 3 were also attempted for fetal heart malformations. Unfortunately, only one of EPA's nested models fit the available data, so that a comparison of results from different models was not possible. Furthermore, EPA's model gave a linear response in the low dose region, similar to those found with procedure 1, and thus the results would be similar. Future work with either procedure 2 or 3 would be to try different models, since EPA's sole fitting model does not reflect the low dose data, that is, the response looks to be hormetic at the lowest, non-zero dose (see Table 4).

Table 3. Modeling results for fetal heart malformations using procedure 1. All units are in ppm, but convertible to µg/m^3 by multiplying by 5.8.

Johnson *et al*. (2003): Fetal heart malformations		
	Concentration	Hazard up to
RfC	0.00037	0
HEC, $BMDL/3UF_{ad}$	0.0040	0.01
EPA, 2011a, Figure F-15 (page F-35)		
RfC	0.00037	0
HEC, $BMDL/3UF_{ad}$	0.011	0.01
EPA, 2011a, Table F-7 (page F-12) and EPA, 2011a, Figure F-15 (page F-35)		

Notes with the cardiac analysis:

- erudite developmental toxicologists think that the effect is spurious
- assessment is based on pup statistics, against EPA guidelines
- $BMDL_{01}$ is used; not typical choice and several models failed
- High dose group dropped

$3UF_{ad}$ = three-fold uncertainty factor for toxicodynamic uncertainty when extrapolating from the experimental animal (rats) to the human

3UFad = three-fold uncertainty factor for toxicodynamic uncertainty when extrapolating from the experimental animal (rats) to the human

Figure 1. Hazard Range of Heart Malformations POD = BMDL

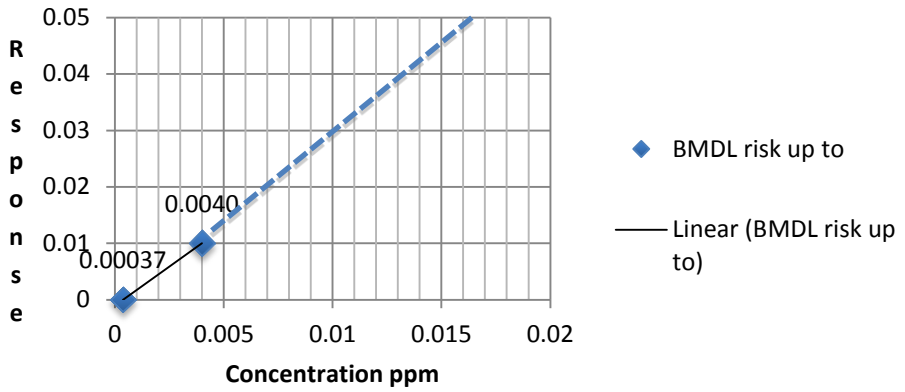


Figure 2. Hazard Range of Heart Malformations POD = BMD

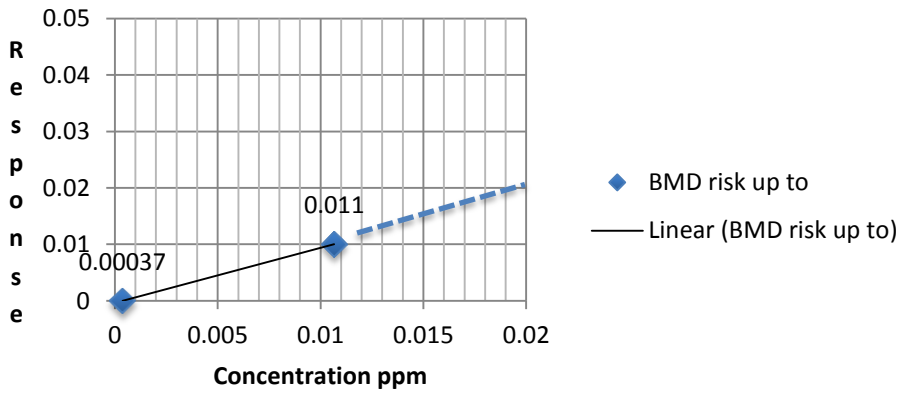


Table 4. Data on fetuses and litters with abnormal hearts from Johnson *et al*. (2003)

Dose (mg/kg/d):	**0**	**0.00045**	**0.048**	**0.218**	**129**
Number of pups:	606	144	110	181	105
Abnormal heart	13	0	5	9	11
Litters	55	12	9	13	9
Abnormal heart	9	0	4	5	6

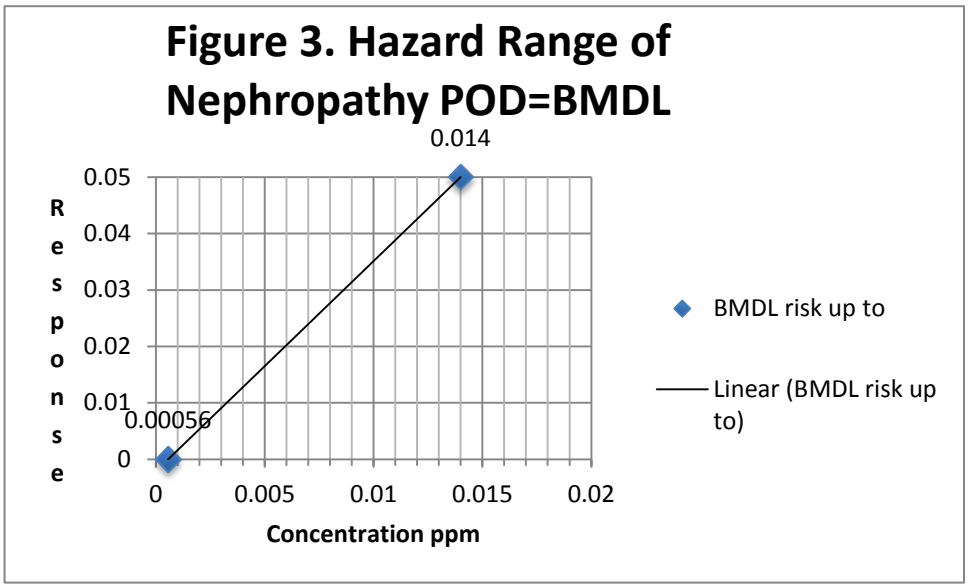
From EPA, Table F-4, Page F-9

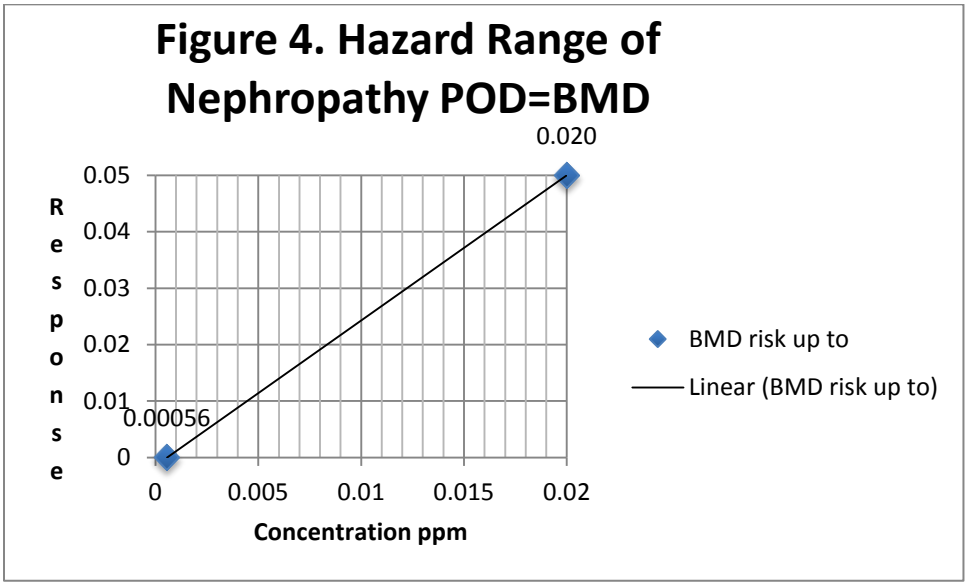
Results of modeling procedure 1 for nephropathy are shown in Table 5 and Figures 3 and 4. As before, hazard ranges above the RfC are determined from a straight line drawn from both the BMDL05 (the BMDL at a predicted response 5% above background) and BMD05 (the BMD at a predicted response 5% above background) to the RfC, where the RfC is considered to a concentration associated with be without risk, and the $BMDL_{05}$ and the BMD_{05} are the HECs taken from EPA (IRIS *Toxicological Review of Trichloroethylene*, Appendix F, p. F-30, 2011a) and adjusted by a three-fold uncertainty factor for toxicodynamic uncertainty when extrapolating from the experimental animal (rats) to the human. Again, the use of the HEC_{99} is contra-indicated because its use conflates with the results of procedure 1.

Modeling procedures 2 and 3 were also attempted for nephropathy. Unfortunately, the lowest data point in these data is greater than 60% response, and, thus, these data are insufficient in the low dose region to confidently predict a response with any model. In such cases, responses are solely dependent on the choice of model, and would vary widely. Future work with either procedure 2 or 3 is not possible on this data set.

Table 5. Modeling results for nephropathy using procedure 1. All units are in ppm, but convertible to ug/m^3 by multiply by 5.8.

NTP (1988): toxic nephropathy		
	Concentration	Hazard Range
RfC	0.00056	0
HEC, BMDL/3UFad	0.014	0.05
(EPA, 2011a, Figure F-11 (page F-30)		
RfC	0.00056	0
HEC, BMD/3UFad	0.020	0.05
EPA, 2011a, Figure F-10 (page F-29) and		
EPA, 2011a, Figure F-11 (page F-30)		





An interesting finding when comparing the results of both endpoints is that when hazard ranges above the different RfCs are compared, the hazard ranges above the RfC for nephropathy are more severe than for fetal heart malformations. For example, compare the BMD results in Figures 4 and 2. Here, the risks associated with nephropathy at a concentration of 0.01 ppm approximate may range up to approximately 2.5%, whereas the risks associated with fetal heart malformations at the same concentration may range up to approximate 1%. This difference in hazard ranges is also true when risks at the BMDL are compared in Figures 3 and 1.

Discussion

This development of hazard ranges for two these noncancer endpoints, considered as critical effects in the EPA (2011a) TCE IRIS text, is in keeping with the NAS (2009) suggestion to

explore quantitative aspects of noncancer dose response assessment. The stated risks can be confidently expected to not exceed the given values, but the corollary is that the risks are then likely to be less and could even be zero.

In fact, since uncertainty factors are generally considered to be protective from the standpoint of the expected average value (Dourson and Stara, 1983; Swartout *et al*., 1998), and since noncancer effects are thought to be associated with thresholds (EPA, 2013), the actual risks at or around the RfC are likely to be much less than that determined from a linear response. Thus, the proper interpretation of these stated values is as a risk range, and not as a single value. This same reason also leads to considering the hazard range associated with the BMD extrapolation to be more appropriate than the hazard range associated with the BMDL extrapolation. This is because the BMDL is used to estimate the RfC in a protective sense; *i.e*., it works with the protective uncertainty factor to drive the RfC to an expected NOAEL in a sensitive human population, or below a population threshold dose. The use of the BMDL extrapolation to project an accurate hazard range is correspondingly diminished.

The choice of hazard range associated with the two different critical effects depends in part on the level of confidence assigned to each RfC, and in part on the severity of the associated hazard ranges developed in this analysis. EPA's confidence in the RfC is "high" with nephropathy and "medium" with heart malformations. The hazard range is more severe with nephropathy when compared with heart malformations. In either case, the more appropriate choice of hazard range is that associated with nephropathy.

Thus, the most reasonable hazard range to use is the one based on the BMD of nephropathy. This choice can then be compared with other methods developed in this Alliance for Risk Assessment (*ARA*) project.

Estimate RfD Probabilities Using Uncertainty Factor Distributions

Methods

This method uses theoretical distributions of EPA's five uncertainty factors (i.e., intrahuman, experimental animal to human, subchronic to chronic, LOAEL to NOAEL and lack of database), based on the work of Swartout *et al*. (1998). Three of these factors are based on estimating a NOAEL of one dose group to a comparable NOAEL of another (i.e., the expected same response; e.g., experimental animal NOAEL to human NOAEL, subchronic NOAEL to chronic NOAEL, and lack of database), and not estimating a different population response (*i.e*., interhuman and LOAEL to NOAEL). Thus, this method cannot develop a dose response relationship in humans for the range of interest. Rather the probabilities are interpreted as the likelihood that the stated RfD is a sensitive human NOAEL, which is the intent of the RfD's definition.[4] Furthermore, if uncertainty factors are seen, individually, as "upper bounds" on the dose-scaling factor for sources of uncertainty, then determining comparable upper bounds for combinations of uncertainty factors can be accomplished by treating uncertainty factors as

[4] The RfD is an estimate (with uncertainty spanning perhaps an order of magnitude) of a daily exposure to the human population (including sensitive subgroups) that is likely to be without an appreciable risk of deleterious effects during a lifetime (Barnes and Dourson, 1988).

distributions, which can be combined by probabilistic techniques. The Swartout *et al.* (1998) approach does not attempt to distinguish one uncertainty factor from another based on empirical data or biological mechanisms, but rather uses a simple displaced lognormal distribution as a generic representation of all uncertainty factors.

Determining the probabilistic limits for the uncertainty factors used in the derivation of the RfD is an important step toward the goal of characterizing the risk of noncarcinogenic effects from exposure to environmental pollutants. The probabilities developed have applicability in comparisons among RfDs and/or for determining different RfDs based on different choices of probability.

Results

Table 5 shows the modeling results for EPA's three different TCE IRIS RfCs at probabilities of 50%, 95% and 99%. The candidate RfC value associated with each of these probabilities reflects the level of confidence that the underlying RfC truly reflects a sensitive human NOAEL (*i.e*, the probability that the candidate RfC value does not exceed the NOAEL for any sensitive human individual or subpopulation). Modeling results for two of these RfC values, Johnson *et al.* (2003) and NTP (1998) are based on identical probabilities, since the uncertainty factor of 10-fold is the same in both cases. The modeling results for Keil *et al*. (2009) are based on a different probability distribution, since the uncertainty factor for this RfC is larger.

BOLD printed entries, shown in either black or red text, indicate matches to the IRIS RfC. Note that whereas the IRIS RfCs for Johnson *et al*. (2003) and NTP (1998) are both associated with a 95% probability, the IRIS RfC for Keil *et al*. (2009) is associated with a 99% probability. RED printed entries show consistency in 95% probability. Of note here is that the RfC based on Keil *et al*. (2009) would be 2-fold higher than its IRIS equivalent if based on a 95% probability. Also notable is that the RfCs based on Keil *et al*. (2009) and NTP (1998) are identical when based on the same probability, whereas they are different when seen on EPA's IRIS.

Table 6. 50th, 95th, & 99th % for TCE IRIS RfCs with Uncertainty Factors of 10 & 100; units in ug/m3

Study	IRIS UF	Confidence	IRIS RfC	50th %	95th %	99th %
Johnson et al. (2003)	10	Medium	**4E-04**	1E-03	**4E-04**	2E-04
NTP (1988)	10	High	**6E-04**	2E-03	**6E-04**	4E-04
Keil et al. 2009	100	Medium	**3E-04**	3E-03	6E-04	**3E-04**

Discussion

This development of probabilities with this method is not a dose response relationship in humans for the effect of interest. Rather the probabilities are interpreted as the likelihood that the stated RfC is a sensitive human NOAEL, which as stated before, is the intent of the RfC's definition. This method is also in keeping with the NAS (2009) suggestion to explore quantitative aspects of noncancer dose response assessment.

The stated probabilities can be confidently expected to approximate the given values, but with the realization that the Swartout *et al*. (1998) approach does not attempt to distinguish one

uncertainty factor from another based on empirical data or biological mechanisms, but rather uses a simple displaced lognormal distribution as a generic representation of all uncertainty factors. Specific, refined distributions could be used to replace these generic ones (see for example, Baird *et al*., 1996).

If managers wish to focus on RfCs with a consistent probability, then the comparison among RfCs would be on the basis of the values associated with the chosen probability, and not the value as stated on IRIS. For example, in the choice of a 95% probability, the associated RfCs would be 4E-04 for heart malformations, 6E-04 for nephropathy and 6E-04 for immune effects. In this case, the RfCs for immune and nephropathy are the same and the value for heart malformations is different, whereas the values on IRIS suggest that the RfCs for immune and heart malformations are similar and that the RfC for nephropathy is different. Different choices of probability lead to correspondingly different RfCs.

Section 4.

Risk Management Synthesis

Introduction

Waste sites throughout the United States are often managed using risk assessment information and methods from EPA. Risk values such as Regional Screening Levels (RSLs) that are derived with these EPA methods (U.S. EPA, 2012b) are used as part of the process of managing site risks. For chemicals that cause cancer, acceptable cancer risks are typically defined as a risk range that spans two orders of magnitude, or a factor of 100-fold, that is, a range from 10^{-6} to 10^{-4} (U.S. EPA, 1991). For noncancer toxicity, the RfD or its inhalation equivalent, the RfC is used, without specifying any risk or hazard range. However, the RfD and RfC as defined by EPA (Barnes and Dourson, 1988; U.S. EPA, 1994; and U.S. EPA, 2002) are associated with "uncertainty spanning perhaps an order of magnitude." So can a range be determined from this definition, and if so, might a risk manager be able to use it?

The reasoning behind this definition of RfD/RfC, and, specifically, with the use of the word "perhaps" by Barnes and Dourson (1988) had everything to do with the variation in the underlying toxicity databases of different chemicals. In some cases, the database has sufficient information in humans and experimental animals so that the resulting RfD/RfC can be developed with a one-fold or 10-fold uncertainty factor. Uncertainty or imprecision in such an RfD/RfC is smaller, that is, perhaps less than an order of magnitude. In other cases, fewer data are available so that the resulting RfD/RfC is developed with a 1000-fold or larger uncertainty factor. Uncertainty or imprecision in these RfD/RfCs is greater, that is, perhaps more than an order of magnitude.

Each RfD/RfC value has both imprecision and uncertainty. The imprecision part comes from the repeatability of the overall process; i.e., how close to the first RfD/RfC would a second expert body come, if it were to repeat the process. In such cases, the imprecision might be best characterized as perhaps three-fold on either side of the RfD/RfC (Felter and Dourson, 1998; U.S. EPA, 2002).

The uncertainty part comes from the fact that each RfD/RfC is protective, because it is developed using uncertainty factors that are each protective based on the observed behavior of the "average" chemical (Dourson and Stara, 1983; Dourson *et al.*, 1996). This protectiveness assures that any RfD/RfC is an underestimate of the expected value. Furthermore, the use of multiple uncertainty factors results in even more protective RfD/RfC values (Swartout *et al.*, 1998). Because of this, the uncertainty in an RfD/RfC might be best characterized as perhaps 10-fold above the RfD/RfC, where the latter is considered as the floor to this range, *i.e.*, the RfD/RfC is the lowest value within the range (Dourson and Stara, 1983; Felter and Dourson, 1998).

Risk managers who need to manage waste sites throughout the U.S. often address risk for carcinogenicity using a risk range of 10^{-6} to 10^{-4} and for noncancer toxicity using a hazard index of 1 for mixtures, or a Hazard Quotient of 1 for single chemicals. The purpose of this paper is to determine a range in the hazard index or quotient that is consistent with the range of uncertainty in the RfD/RfC, and that is similar in concept to the risk range of 10^{-6} to 10^{-4} for carcinogenicity. This range in the hazard index or quotient might allow managers to have comparable flexibility in closing or otherwise managing waste sites where the evaluation of noncancer effects drives the risk assessment. The evaluation of the noncancer effects associated with the chlorinated solvent trichloroethylene (TCE) is used as an example to illustrate the determination of, and application of, this hazard index range.

Defining the Range in the RfD/RfC and Hazard Index or Quotient

In the IRIS Summary for TCE, U.S. EPA (2011a) has identified three candidate RfC values in its evaluation of the noncancer inhalation toxicity of TCE. These three candidates are described below:

- a candidate RfC of 2 $\mu g/m^3$ based on decreased thymus weight in female mice (Keil *et al.*, 2009);
- a candidate RfC of 2 $\mu g/m^3$ based on fetal heart malformations in rats (Johnson *et al.*, 2003); and
- a candidate RfC of 3 $\mu g/m^3$, based on toxic nephropathy in female rats (NTP, 1988).

Each of the candidate RfC values may be evaluated with respect to the imprecision and the uncertainty inherent in its derivation.

As further described in Section 1, the imprecision of each of the candidate RfCs may be considered as a uniform or equal distribution around the RfC as its central value, that is, as 2 $\mu g/m^3$ for Keil *et al.* (2009) or Johnson *et al*. (2003), or 3 $\mu g/m^3$ for NTP (1988). A floor and ceiling to this uniform distribution would be a value lower and higher than the candidate RfC, perhaps two-fold in each direction for the Johnson *et al*. (2003) or NTP (1988) RfCs, and three-fold in either direction for the Keil *et al.* (2009) RfC.

The uncertainty in each of the candidate RfCs may also be evaluated, but in contrast to imprecision, these uncertainties are associated with ranges of values that are not uniformly

distributed around the RfC as its central value. This is because these three candidate RfCs are developed using different PODs and different uncertainty factors that are known to be protective from the behavior of the "average" chemical. For risk management decisions, it is generally the range of the uncertainty that is more important when compared with the range of imprecision. This is because managers are interested in making decisions that are protective of public health and it is an understanding of the uncertainties in the range of RfCs that is generally more informative.

These non-uniform ranges of uncertainty have as a floor the individual candidate RfCs as on IRIS. This decision to use the individual candidate RfC as the floor of the range for each non-cancer endpoint is reasonable from a practical point of view, because managers are unlikely to take regulatory action below these values, due to the protective nature implicit in the derivation of each candidate RfC, as described above. However, using the RfD/RfC as the floor to a range in its value also has theoretical support (Dourson and Stara, 1983).

Because the uncertainty range varies for each study and each candidate RfC, it may be helpful to develop a midpoint so as to guide managers when exposures fall within the range. The determination of the midpoint of this non-uniform hazard range is a judgment that meshes four considerations for each candidate RfC; specifically:

- the size of the uncertainty factor,
- the steepness of the hazard slope (*i.e.*, the slope of the line describing hypothetical population responses at concentrations above the RfC, see section 3),
- the confidence in the choices of critical effect, and
- the confidence in the PDD.

Midpoints that are closer to their RfCs are associated with a smaller uncertainty factor, a steeper hazard slope, a higher confidence in the critical effect, and a higher confidence in the POD. Midpoints that are further from their RfCs are associated with a larger uncertainty factor, a shallower hazard slope, a lower confidence in the critical effect, and a lower confidence in the POD.

For the fetal malformation endpoint based on Johnson *et al.* (2003), the midpoint of the endpoint-specific uncertainty range (10 $\mu g/m^3$) is judged to be five-fold above the candidate RfC due to its small uncertainty factor of 10, shallower hazard slope (Section 3), low confidence in the critical effect (Section 2), and low confidence in the choice of a benchmark response of 1% ($BMDL_{01}$). For the toxic nephropathy endpoint based on the NTP (1988) study, the midpoint of the endpoint-specific uncertainty range (9 $\mu g/m^3$) is judged to be three-fold above the candidate RfC due to its small uncertainty factor of 10, steeper hazard slope (Section 3), medium confidence in the critical effect, and medium to low confidence in the choice of a benchmark response of 5% ($BMDL_{05}$). For decreased thymus weight endpoint based on the Keil *et al.* (2009) study, the midpoint of the endpoint-specific uncertainty range (20 $\mu g/m^3$) is judged to be 10-fold above the candidate RfC due to its larger uncertainty factor of 100, medium confidence in the critical effect, and medium to low confidence in its choice of a LOAEL as the POD. The effect shown by Keil *et al.* (2009) does not lend itself to dose response modeling, so a judgment of steepness of hazard slope is not possible.

The ceiling value for the endpoint-specific uncertainty range for each RfC is defined as the POD (*i.e.*, the HEC_{99}) for each candidate RfC (*i.e.*, for each non-cancer endpoint), as shown on EPA's IRIS. Therefore, the ceiling value of each endpoint-specific uncertainty range is 20 $\mu g/m^3$ from the Johnson *et al.* (2003) study), 30 $\mu g/m^3$ from the NTP (1988) study and 190 $\mu g/m^3$ from the Keil *et al.* study (2009). This decision is reasonable from a practical point of view, because managers are likely to take regulatory action above these values due to the fact that specific toxic effects can sometimes be associated with their values. The floor, midpoint and ceiling values for each endpoint-specific uncertainty range are shown in Table 7.

Table 7. Different uncertainty ranges for different TCE RfCs. All values are in $\mu g/m^3$. Shaded areas indicate best **overall uncertainty range** for risk management purposes.

Study	**IRIS UF** [a]	**Steep** [b] **Slope**	**Confidence**		**Uncertainty Ranges**		
			Critical [c] **Effect**	**Point of** [d] **Departure**	**Floor**	**Intermediate**	**Ceiling**
Johnson *et al.* (2003)	10	Lower	Low	Low	2	10	20
NTP (1988)	10	Higher	Medium	Medium to Low	3	9	30
Keil *et al*. 2009	100	NA	Medium	Medium to Low	2	20	190

a. Size of the uncertainty factor as on IRIS
b. Steepness of the hazard slope (*i.e.*, the slope of the line describing hypothetical population responses at concentrations above the RfC), as per Section 3.
c. Confidence in the choices of critical effect, as per Section 4.
d. Confidence in the POD, as per Section 4.

Since EPA selected these three studies (two principal and one supporting) as the basis for the derivation of the RfC for non-cancer effects, the endpoint-specific uncertainty range of each of the three studies may be considered in the implementation of risk management decisions. From the collective evaluation of the endpoint-specific uncertainty ranges of all three studies, a "total uncertainty range" of 2 $\mu g/m^3$ to 190 $\mu g/m^3$ may be inferred. However, extraction of a "multi-endpoint uncertainty range" from the broader total uncertainty range is more useful for risk management decision-making. The multi-endpoint uncertainty range may be defined as an estimate of "a daily exposure to the human population (including sensitive subgroups) that is

likely to be without an appreciable risk of deleterious effects during a lifetime", based upon the definition of the RfD (Barnes and Dourson, 1988)[5].

The multi-endpoint uncertainty range of the RfC for TCE (as defined by multi-endpoint floor, midpoint and ceiling values) was determined here by careful discernment of the confidence, imprecision and uncertainty associated with each endpoint-specific floor, midpoint and ceiling value. The floor of the multi-endpoint uncertainty range of the RfC for TCE was determined by comparing the candidate RfC values from each of the three studies (*i.e.*, 2 $\mu g/m^3$ for both the decreased thymus weight and fetal cardiac malformation endpoints, and 3 $\mu g/m^3$ for the toxic nephropathy endpoint). These three floor values are so closely clustered that, based on the imprecision inherent in non-cancer hazard estimates, the values are essentially indistinguishable[6]. The endpoint-specific floor value of 3 $\mu g/m^3$ based on toxic nephropathy not only represents the endpoint-specific floor value of the highest overall confidence from among the three endpoint-specific floor values (see Table 7), but is equivalent to a concentration that is found within the mathematical precision range of each of the three endpoint-specific floor values[7]. Therefore, the

[5] Barnes and Dourson (1988) defined the RfD as a point value "with an uncertainty spanning perhaps an order of magnitude". The point of departure (POD) for the RfD as defined by Barnes and Dourson was typically (although not always) based upon a no observed adverse-effects level (NOAEL) from a single study for a single critical effect. The assessment of confidence levels in the study design, critical effect and POD of a single study and single critical effect enables the assignation of uncertainty factors in a relatively straightforward manner. The multi-endpoint uncertainty range of the RfC for TCE has been defined here (in 2013) in the same terms as the RfD (as stated in 1988). The use of the definition of the RfD to also define the multi-endpoint uncertainty range is a recognition of the complexity associated with the prediction of a "safe dose" or "safe concentration" associated with multiple effects, observed in multiple studies and species, and multiple points of departure (each variously based on a $BMDL_{01}$, $BMDL_{05}$ or lowest observed adverse effect level (LOAEL) value, appropriately averaged over two different exposure periods, as relevant for developmental or chronic effects. Therefore, the concept of a "safe dose" for the non-cancer effects of TCE has been applied here to a range of values (*i.e.*, the multi-endpoint uncertainty range) which represents a "safe concentration" for multiple endpoints and multiple studies, with various degrees of confidence in study design, critical effect and POD, as illustrated in Table 7.

[6] With respect to non-cancer endpoints, the safety of human exposures to environmental concentrations is evaluated on the basis of the hazard quotient (HQ), *i.e.*, the ratio of the environmental concentration of a substance (C_{air}) to its RfC (*i.e.*, C_{air}/RfC). The HQ associated with acceptable exposures is one (1), with a precision of one significant figure (*i.e.*, an implicit HQ range of 0.95 to 1.5). The mathematical precision of the endpoint-specific floor value of 2 $\mu g/m^3$ corresponds to a range of 1.9 $\mu g/m^3$ to 3 $\mu g/m^3$; the mathematical precision of the endpoint-specific floor value of 3 $\mu g/m^3$ corresponds to a range of 2.5 $\mu g/m^3$ to 4.5 $\mu g/m^3$. Thus, the ranges associated with the implicit precision of the three endpoint-specific floor values overlap; at a precision of one significant figure, the value of 3 $\mu g/m^3$ falls within the mathematical precision ranges of all three endpoint-specific floor values. The endpoint-specific floor value of 3 $\mu g/m^3$ based on toxic nephropathy not only represents the endpoint-specific floor value of the highest overall confidence from among the three endpoint-specific floor values (see Table 7), but is equivalent to a concentration that is found within the mathematical precision range of each of the three endpoint-specific floor values; the value of 3 $\mu g/m^3$ was thus selected to represent the floor value of the multi-endpoint uncertainty range.

[7] *Ibid.*

value of 3 μg/m^3 was selected to represent the floor value of the multi-endpoint uncertainty range.

Similarly, the midpoint of the multi-endpoint uncertainty range of the RfC for TCE is determined by comparing the endpoint-specific midpoint values from each of the three studies (*i.e.*, 20 μg/m^3 for the decreased thymus weight endpoint, 10 μg/m^{3}for the fetal cardiac malformation endpoint, and 9 μg/m^3 for the toxic nephropathy endpoint). The midpoint values for the fetal cardiac malformation and toxic nephropathy endpoints are so closely clustered that the values are essentially indistinguishable. The endpoint-specific midpoint value of 9 μg/m^3 based on toxic nephropathy not only represents the endpoint-specific midpoint value of the highest overall confidence from among the three endpoint-specific midpoint values (see Table 7), but is lower than or equivalent to the other two endpoint-specific midpoint values. Therefore, the value of 9 μg/m^3 was selected to represent the midpoint value of the multi-endpoint uncertainty range.

The ceiling of the multi-endpoint uncertainty range of the RfC for TCE is determined by comparing the endpoint-specific ceiling values from each of the three studies (*i.e.*, 190 μg/m^3 for the decreased thymus weight endpoint, 20 μg/m^3 for the fetal cardiac malformation endpoint, and 30 μg/m^3 for the toxic nephropathy endpoint). The endpoint-specific ceiling values for the fetal cardiac malformation and toxic nephropathy endpoints are closely clustered, and the mathematical precision ranges of the two values overlap[8]; the endpoint-specific ceiling value for decreased thymus weight is substantially higher. Therefore, the endpoint-specific ceiling value for fetal cardiac malformations (20 μg/m^3) has been selected here as the ceiling value for the multi-endpoint uncertainty range. Although the POD is of low confidence, the value is based on the HEC_{99} (calculated from the $BMDL_{01}$), and is a plausible yet conservative estimate of the "safe concentration" for human exposures[9].

In summary, the confidence in each of the endpoint-specific uncertainty ranges was subsequently considered in the determination of the multi-endpoint uncertainty range for risk management purposes (*i.e.*, 3 μg/m^3 to 20 μg/m^3). The higher-confidence results of the NTP study were used

[8] The mathematical precision (at HQ = 1) associated with the endpoint-specific ceiling value of 20 ug/m^3 corresponds to a range of 15 ug/m^3 to 30 ug/m^3; the mathematical precision (at HQ = 1) of the endpoint-specific floor value of 30 ug/m^3 corresponds to a range of 25 ug/m^3 to 45 ug/m^3. Thus, the ranges associated with the implicit precision of each endpoint-specific floor value overlap.

[9] The ceiling value for the toxic nephropathy endpoint (30 ug/m^3) is a plausible alternative representation of the ceiling value for the multi-endpoint uncertainty range. Since the mathematical precision ranges (at HQ=1) of the endpoint-specific ceiling values for the fetal cardiac malformation and toxic nephropathy endpoints overlap, the values may be considered to be approximately equivalent. In this case, the endpoint-specific ceiling value for toxic nephropathy would be favored over the endpoint-specific ceiling value for fetal cardiac malformation on the basis of higher confidence not only in the critical effect (high *versus* low) but also in the point of departure (medium *versus* low); the latter comparison is of particular relevance since the point of departure serves as the endpoint-specific ceiling value for each endpoint. Nevertheless, the endpoint-specific ceiling value for fetal cardiac malformations was selected as the multi-endpoint ceiling value, since the amount of overlap between the two ranges of mathematic precision was not substantial. In addition, the selection of the ceiling value of 20 ug/m^3 based on the HEC_{99} for fetal cardiac malformations, enables the risk manager to conservatively account for developmental endpoints when the multi-endpoint uncertainty range is used as the basis for risk management decisions.

to determine the floor and midpoint of this uncertainty range. The highly controversial results from the Johnson *et al.* (2003) study, while associated with low confidence (see Section 2), were nevertheless used to determine the ceiling level of this uncertainty range. The widest endpoint-specific uncertainty range is embedded within the individual uncertainty range from the Keil *et al.* (2009) study; therefore, this study was considered to be confirmatory of the uncertainty ranges associated with the other two endpoints.

Toxicologists are not able to distinguish the absence of health risk between any two or more values within this overall uncertainty range. That is to say, toxicologists cannot differentiate, for example, the "safety" of a value of 4 $\mu g/m^3$ versus a value of 17 $\mu g/m^3$. Because of this, managers may use different values within the multi-endpoint uncertainty range of 3 $\mu g/m^3$ to 20 $\mu g/m^3$, along with site-specific exposure assessments, and other site considerations to make different decisions on a case-by-case basis.

For example, when a site-specific exposure assessment defines a range of exposures that are primarily below the multi-endpoint uncertainty range of 3 $\mu g/m^3$ to 20 $\mu g/m^3$, then the probability of inducing any noncancer effects in the exposure population is lower and the priority for any management action is reduced (see Figure 3a). In this case, a risk manager may decide to take no action, or to delay action pending further information.

In contrast, when the exposure assessment defines a range in exposures that is primarily above the multi-endpoint uncertainty range of 3 $\mu g/m^3$ to 20 $\mu g/m^3$, then the probability of inducing noncancer effects in the exposure population is higher and the priority for risk management action is increased (see Figure 3c). In this case, a risk manager may decide to take action, or to ask for specific information that would refine the estimates of health risk and/or exposure.

When the exposure assessment defines a range in exposures that are primarily in the multi-endpoint uncertainty range of 3 $\mu g/m^3$ to 20 $\mu g/m^3$, then risk managers can use the intermediate value in this uncertainty range, that is 9 $\mu g/m^3$ and other site considerations, to gauge whether a management action is needed or if further information should be sought (see Figure 3b).

The multi-endpoint uncertainty range can also be used to develop a comparable range of hazard quotient estimates for single-chemical exposures to TCE alone, or hazard index estimates for exposures to mixtures of TCE and other chemicals (*e.g.*, solvent impurities such as 1,1,1-trichloroethane, 1,2-dichlroethane, 1,1-dichloroethene; or degradation products, such as *cis*-1,2-dichloroethene and vinyl chloride), when only a point estimate of exposure is available for comparison. This adaptation is consistent with EPA's definition of the RfD/RfC, and is akin to the range developed for cancer risk (i.e., an excess lifetime cancer risk of 10^{-6} to 10^{-4}). When considering exposures to mixtures, risk managers should be cognizant of the target organ(s), mode(s) of action and mechanism(s) of action of the various chemicals in the mixture, based on critical effects and other effects that may be elicited at environmentally relevant concentrations. Additionally, the role of co-exposures and interactions of chemicals may be considered and/or ruled out in developing the range in the hazard index for actions that are recommended by risk managers.

The multi-endpoint uncertainty range is composed of floor, midpoint and ceiling values for which the appropriate averaging time corresponds to different exposure durations (*i.e.*, developmental or long-term/chronic exposure periods). Therefore, this range can be applied to

both long- and short-term exposures, with the associated differences in exposure averaging times.[10] For shorter-term exposures, the results from the Johnson *et al.* study (2003) might also be used to describe the best averaging time, but if so, this averaging time should be based on the average time of cardiac development in humans during fetal growth, approximately 24 days (based on Marcela *et al*., 2013; Fanaroff and Martin, 2003; Schleich, 2003). The use of a 24-day average time for cardiac development in humans is consistent with the fact that the dosing in the Johnson *et al.* study (2003) occurred during the whole time of cardiac development in the rat, as described in the next section.

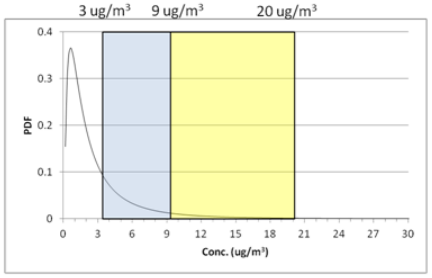


Figure 3a. Exposure distribution of indoor air concentrations primarily below the 3 µg/m^3 to 20 µg/m^3 hazard range. Relatively small proportion of exposures is higher than 3 µg/m^3. Nominal actions or no further action may be warranted for risk management.

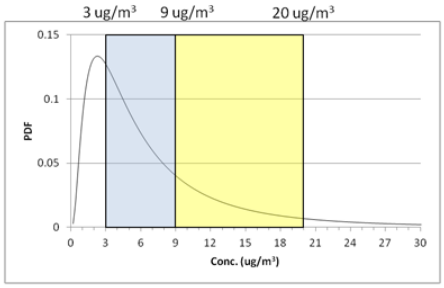


Figure 3b. Exposure distribution of indoor air concentrations falling within the 3 µg/m^3 to 20 µg/m^3 hazard range. Relatively small proportion of exposures is higher than 9 µg/m^3. Limited action may be warranted for risk management

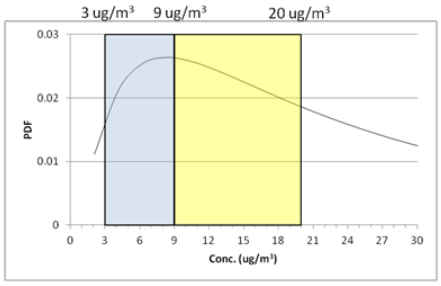


Figure 3c. Exposure distribution of indoor air concentrations primarily above the 3 µg/m^3 to 20 µg/m^3 hazard range. Actions to reduce exposures may be warranted for risk management.

Cardiac development in various species

Cardiac development has been described in several species including humans (Marcela *et al*., 2012; Moorman *et al*., 2003; Zaffran and Frasch, 2002; Srivastava, 2000). The heart is reported to be the first internal organ to form and become functional in the vertebrate embryo (Srivastava, 2000; Richtsmeier, 1999). It develops in several sequential steps, beginning from an epithelium to mesodermal tissue, to a tube, to a looped, two-chambered structure, and finally to the mature four-chambered organ (Fanaroff and Martin, 2002). In the rat, the first primitive cardiac segment appears on embryonic day 9 (ED 9) (Marcela *et al*., 2012). A comparable developmental stage has been described in stage 10 mouse embryo (7 days post coitum) and Carnegie Stage VI-VII human embryos (corresponding to second week of gestation) (Marcela *et al*., 2012) or beginning of third week of gestation) (Fanaroff and Martin, 2003). Schleich (2003)

[10] Use of the RfC values for assessing long-term exposures is well-established in EPA guidance.

lists the following stages as occurring during the third week of gestation in humans: development of the primitive streak—gastrulation (day 15); formation of intra-embryonic mesoderm (day 16); mesoblast differentiation—somatopleura and splanchnopleura (day 17); development of blood islets, the cardiogenic region, and primitive heart tubes (day 19); and formation of the primitive endocardial tube (day 21). A series of internal and external changes occurs, leading to a fully formed heart (*i.e.*, a mature heart, with interventricular foramen completely closed and almost compact interventricular septum) in the rat ED 16 embryo. The heart is also reported to be fully formed in the stage 30 Hamburger–Hamilton (HH) chick embryo (Contreras Ramos *et al*., 2008) and Carnegie Stage XIX human embryo (Goor and Lillehei, 1975) (corresponding to 38 ± 1 days [Marcela *et al*., 2012]). According to Marcela *et al*. (2012), the more relevant developmental events in rat fetuses from 17 to 20 days and 21 days newborn include an increase in weight and crown to tail length. Aside from these changes, the fetus morphology remained almost the same, including the still-closed eyes. Histological maturation of all the organs appears to be almost fully in the ED 16 embryo.

Based on the available information, the human hearts starts developing between days 13 and 15 after gestation or later and may be completely formed by 37-39 days into gestation (Marcela *et al*., 2012) or later, indicating the length of cardiac development in humans during fetal growth to be 24-26 days. Since exposure to TCE in Johnson *et al*. (2003) occurred during the whole time of cardiac development in the rat, it is reasonable to use the time of cardiac development in humans, 24-26 days, as the averaging time for risk management decisions.

References

ATSDR (Agency for Toxic Substances and Disease Registry). (2006a). Health statistics review: Cancer and birth outcome analysis, Endicott area investigation, Town of Union, Broome County, New York. U.S. Department of Health and Humans Services. Atlanta, GA. Available at: http://www.atsdr.cdc.gov/HAC/pha/EndicottAreaInvestigation/EndicottHealthStatsReviewHC052606.pdf

ATSDR (Agency for Toxic Substances and Disease Registry). (2006b). Health consultation: Public health implications of exposures to low-level volatile organic compounds in public drinking water. Endicott area investigation, Broome County, New York. U.S. Department of Health and Human Services. Atlanta, GA. Available at: http://www.atsdr.cdc.gov/HAC/pha/EndicottAreaInvestigation113006/EndicottAreaInvestigationHC113006.pdf .

ATSDR (Agency for Toxic Substances and Disease Registry). (2008). Health consultation: Health statistics review follow-up: Cancer and birth outcome analysis. Endicott area investigation, Town of Union, Broome County, New York. U.S. Department of Health and Human Services. Atlanta, GA. Available at: http://www.atsdr.cdc.gov/hac/pha//EndicottAreaInvestigationFollowUp/EndicottAreaHC051508.pdf .

Baird, S.J.S., Cohen, J.T., Graham, J.D., Shlyakhter, A.I., and Evans, J.S. (1996). Noncancer Risk Assessment: A Probabilistic Alternative to Current Practice. Human Ecol Risk Assess. 2(1): 78-99.

Barnes, D.G. and Dourson, M.L. (1988). Reference dose (RfD): Description and use in health risk assessments. Regul Toxicol Pharmacol. 8: 471-486.

Blossom, S.J., Doss, J.C., Hennings, L.J., Jernigan, S., Melnyk, S., and James, S.J. (2008). Developmental exposure to trichloroethylene promotes CD4+ T cell differentiation and hyperactivity in association with oxidative stress and neurobehavioral deficits in MRL+/+ mice. Toxicol Appl Pharmacol. 231: 344–353.

Bove, F., Fulcomer, M., Klotz, J., Esmart, J., Dufficy, E., and Savrin, J. (1995). Public drinking water contamination and birth outcomes. Am J Epidemiol. 141: 850-862.

Bove, F. (1996). Public drinking water contamination and birthweight, prematurity, fetal deaths, and birth defects. Toxicol Ind Health. 12: 255-266.

Boyer, A., Finch, W., and Runyan, R. (2000). Trichloroethylene inhibits development of embryonic heart valve precursors in vitro. Toxicol Sci. 53: 109-117.

Bross, G., DiFranceisco, D., and Desmond, M. (1983). The effects of low dosages of trichloroethylene on chick development. Toxicology. 28: 283-294.

CalEPA (California Environmental Protection Agency). (2009). Public Health Goal for Trichloroethylene in Drinking Water. Pesticide and Environmental Toxicology Branch, Office of Environmental Health Hazard Assessment. Available online at: http://oehha.ca.gov/water/phg/pdf/TCE_phg070909.pdf.

Carney, E., Thorsrud, B., Dugard, P., and Zablotny, C. (2006). Developmental toxicity studies in Crl:CD (SD) rats following inhalation exposure to trichloroethylene and perchloroethylene. Birth Defects Res B Dev Reprod Toxicol. 77: 405-412. Available at: http://dx.doi.org/10.1002/bdrb.20091.

Casarett, L.J., & Doull, J. (2001). Casarett and Doull's Toxicology: The Basic Science of Poisons, 6th Ed. McGraw-Hill Pub. Division, New York, NY.

Contreras-Ramos, A., Sanchez Gomez, C., Garcia Romero, H., and Cimarosti, L. (2008). Normal development of the muscular region of the inter- ventricular septum. I. The significance of the ventricular trabecu- lations. Anat Histol Embryol. 37: 344–351.

Cosby, N., and Dukelow, W. (1992). Toxicology of maternally ingested trichloroethylene (TCE) on embryonal and fetal development in mice and of TCE metabolites on in vitro fertilization. Toxicol Sci. 19: 268-274.

Dawson, B; Johnson, P; Goldberg, S; Ulreich, J. (1990). Cardiac teratogenesis of trichloroethylene and dichloroethylene in a mammalian model. J Am Coll Cardiol 16: 1304-1309.

Dawson, B., Johnson, P., Goldberg, S., and Ulreich, J. (1993). Cardiac teratogenesis of halogenated hydrocarbon-contaminated drinking water. J Am Coll Cardiol. 21: 1466-1472. Available at: http://dx.doi.org/10.1016/0735-1097(93)90325-U.

Dawson, H.E., and McAlaray, T. (2009). A compilation of statistics for VOCs from post-1990 indoor air concentration studies in North American residences unaffected by subsurface vapor intrusion. Ground Water Monit R. 29: 60-69.

Dorfmueller, M., Henne, S., York, R., Bornschein, R., and Manson, J. (1979). Evaluation of teratogenicity and behavioral toxicity with inhalation exposure of maternal rats to trichloroethylene. Toxicology. 14: 153-166.

Dourson, M.L., and Stara, J.F. (1983). Regulatory history and experimental support of uncertainty (safety) factors. Regul Toxicol Pharmacol. 3: 224-238.

Dourson, M.L., Felter, S.P., and Robinson, D. (1996). Evolution of science-based uncertainty factors in noncancer risk assessment. Regul Toxicol Pharmacol. 24: 108-120.

Drake, V., Koprowski, S., Hu, N., Smith, S., and Lough, J. (2006a). Cardiogenic effects of trichloroethylene and trichloroacetic acid following exposure during heart specification of avian development. Toxicol Sci. 94: 153-162. Available at: http://dx.doi.org/10.1093/toxsci/kfl083.

Drake, V., Koprowski, S., Lough, J., Hu, N., and Smith, S. (2006b). Trichloroethylene exposure during cardiac valvuloseptal morphogenesis alters cushion formation and cardiac hemodynamics in the avian embryo. Environ Health Perspect. 114: 842-847.

Epstein, D., Nolen, G., Randall, J., Christ, S., Read, E., Stober, J., and Smith, M. (1992). Cardiopathic effects of dichloroacetate in the fetal Long-Evans rat. Teratology. 46: 225-235. Available at: http://dx.doi.org/10.1002/tera.1420460306.

EU (European Union). 2009. Recommendation from the Scientific Committee on Occupational Exposure Limits for Trichloroethylene. SCOEL/SUM/142. Available at: http://ec.europa.eu/social/BlobServlet?docId=6405&langId=en

Faranoff, A.A., and Martin, R.J. (2002). Neonatal-Perinatal Medicine, Diseases of the Fetus and Infant, 7th Ed. Vol. 2. Mosby Inc. pp. 1095-1103.

Felter, S.P., and Dourson, M.L. (1998). The Inexact Science of Risk Assessment (and Implications for Risk Management). Human Ecol Risk Assess. 2: 245-251

Fisher, J., Channel, S., Eggers, J., Johnson, P., MacMahon, K., Goodyear, C., Sudberry, G., Warren, D., Latendresse, J., and Graeter, L. (2001). Trichloroethylene, trichloroacetic acid, and dichloroacetic acid: Do they affect fetal rat heart development. Int J Toxicol. 20: 257-267.

Folkes, D., Wertz, W., Kurtz, J., and Kuehster, T. (2009). Observed spatial and temporal distributions of CVOCs at Colorado and New York vapor intrusion sites. Ground Water Monit R. 29: 70-80.

Goldberg, S., Lebowitz, M., Graver, E., and Hicks, S. (1990). An association of human congenital cardiac malformations and drinking water contaminants. J Am Coll Cardiol. 16: 155-164.

Goor, D.A., and Lillehei, C.W. (1975). Congenital malformations of the heart. Embryology, anatomy, and operative considerations. Grune & Strattion, New York, NY.

Hardin, B., Bond, G., Sikov, M., Andrew, F., Beliles, R., and Niemeier, R. (1981). Testing of selected workplace chemicals for teratogenic potential. Scand J Work Environ Health. 7: 66-75.

Health Canada (2005). Guidelines for Canadian drinking water quality: supporting documentation. Trichloroethlyene. Ottawa, Ontario.

Healy, T., Poole, T., and Hopper, A. (1982). Rat fetal development and maternal exposure to trichloroethylene 100 ppm. Br J Anaesth. 54: 337-341.

IRIS (Integrated Risk Information System). (2012). Trichloroethylene (CASRN 79-01-6). U.S. Environmental Protection Agency. Available at: http://www.epa.gov/iris/subst/0199.htm

Johnson, P., Dawson, B., and Goldberg, S. (1998a). Cardiac teratogenicity of trichloroethylene metabolites. J Am Coll Cardiol. 32: 540-545. Available at: http://dx.doi.org/10.1016/S0735-1097(98)00232-0.

Johnson, P., Dawson, B., and Goldberg, S. (1998b). A review: trichloroethylene metabolites: potential cardiac teratogens. Environ Health Perspect. 106(Suppl 4): 995-999.

Johnson, P., Goldberg, S., Mays, M., and Dawson, B. (2003). Threshold of trichloroethylene contamination in maternal drinking waters affecting fetal heart development in the rat. Environ Health Perspect. 111: 289-292.

Johnson, P., Goldberg, S., Mays, M., and Dawson, B. (2005). Correction: Threshold of trichloroethylene contamination in maternal drinking waters affecting fetal heart development in the rat. Environ Health Perspect. 113: A18.

Keil, D., Peden-Adams, M., Wallace, S., Ruiz, P., and Gilkeson, G. (2009). Assessment of trichloroethylene (TCE) exposure in murine strains genetically-prone and non-prone to develop autoimmune disease. J Environ Sci Health A Tox Hazard Subst Environ Eng. 44: 443-453. Available at: http://dx.doi.org/10.1080/10934520902719738.

Lagakos, S., Wessen, B., and Zelen, M. (1986). An analysis of contaminated well water and health effects in Woburn, Massachusetts. J Am Stat Assoc. 81: 583-596.

Loeber, C., Hendrix, M., Diez De Pinos, S., and Goldberg, S. (1988). Trichloroethylene: a cardiac teratogen in developing chick embryos. Pediatr Res. 24: 740-744.

Marcela, S.G., Monsalve, M.C.R., Garibay, M.A.P., Martinez, M.A., Diaz-Cintra, S., De La Rosa-Santander, P., Bladimir, R-R., and Gomez, C.S. (2012). Chronological and morphological study of heart development in the rat. The Anatomoical Record. 295: 1267-1290.

Mishima, N., Hoffman, S., Hill, E., and Krug, E. (2006). Chick embryos exposed to trichloroethylene in an ex ovo culture model show selective defects in early endocardial cushion tissue formation. Birth Defects Res A Clin Mol Teratol. 76: 517-527. Availabe at: http://dx.doi.org/10.1002/bdra.20283.

Moorman, A., Webb, S., Nigel, A., Brown, N.A., Lamers, W., and Anderson, R.H. (2003). Development of the heart: (1) Formation of the cardiac chambers and arterial trunks. Heart. 89: 806-814.

Narotsky, M., and Kavlock, R. (1995). A multidisciplinary approach to toxicological screening: II. Developmental toxicity. J Toxicol Environ Health. 45: 145-171. Available at: http://dx.doi.org/10.1080/15287399509531987.

Narotsky, M., Weller, E., Chinchilli, V., and Kavlock, R. (1995). Nonadditive developmental toxicity in mixtures of trichloroethylene, di(2-ethylhexyl) phthalate, and heptachlor in a 5 x 5 x 5 design. Fundam Appl Toxicol. 27: 203-216. Available at: http://dx.doi.org/10.1093/toxsci/27.2.203.

NRC (National Research Council/ Committee on Human Health Risks of Trichloroethylene). (2006). Assessing the Human Health Risks of Trichloroethylene: Key Scientific Issues. Washington, DC, National Academies Press, Available at: http://www.nap.edu/catalog/11707.html

NTP (National Toxicology Program). (1988). Toxicology and carcinogenesis studies of trichloroethylene (CAS No. 79-01-6) in four strains of rats (ACI, August, Marshall, Osborne-Mendel) (gavage studies). Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health. Available at: http://ntp.niehs.nih.gov/ntp/htdocs/LT_rpts/tr273.pdf.

OEHHA (Office of Environmental Health Hazard Assessment). (2007). Trichloroethylene. Chronic Toxicity Summary. Available at: http://oehha.ca.gov/air/chronic_rels/pdf/79016.pdfOEHHA (Office of Environmental Health Hazard Assessment). (2012). Initiation of Process to Update Public Health Goals for Three Chemicals in Drinking Water , September 2012. Available at: http://oehha.ca.gov/water/phg/092512InitRep.html

RSLTs (Regional Screening Level Tables). (2012) Regional Screening Level Tables User's Guide. Available at: http://www.epa.gov/reg3hwmd/risk/human/rb-concentration_table/usersguide.htm

Rufer, E., Hacker, T., Lough, J., and Smith, S. (2008). Low-dose trichloroethylene exposure during valvuloseptal morphogenesis causes ventricular septal defects in hatched chicks. Toxicologist. 102: 314.

Schwetz, B., Leong, B., and Gehring, P. (1975). The effect of maternally inhaled trichloroethylene, perchloroethylene, methyl chloroform, and methylene chloride on embryonal and fetal development in mice and rats. Toxicol Appl Pharmacol. 32: 84-96. Available at: http://dx.doi.org/10.1016/0041-008X(75)90197-0.

Smets, P. (1999). Imperfect Information-Imprecision-Uncertainty. IRIDIA Université Libre de Bruxelles. Available at: http://sites.poli.usp.br/d/pmr5406/Download/papers/Imperfect_Data.pdf

Smith, M., Randall, J., Read, E., and Stober, J. (1989). Teratogenic activity of trichloroacetic acid in the rat. Teratology. 40: 445-451. Available at: http://dx.doi.org/10.1002/tera.1420400506.

Smith, M., Randall, J., Read, E., and Stober, J. (1992). Developmental toxicity of dichloroacetate in the rat. Teratology. 46: 217-223. Available at: http://dx.doi.org/10.1002/tera.1420460305.

Srivastava, D., and Olson, E.N. (2000). A gene tic blueprint for cardiac development. Nature. 407: 221-226.

Swartout, J., Price, P., Dourson, M., Carlson-Lynch, H., and Keenan, R. (1998). A probabilistic framework for the reference dose. Risk Anal. 18(3): 271-282.

TERA (Toxicology Excellence for Risk Assessment). (2012). Risk Methods. Available at: http://www.tera.org/iter/riskmethods.html

U.S. EPA (U.S. Environmental Protection Agency). (1989). Risk Assessment Guidance for Superfund. Volume I: Human Health Evaluation Manual (Part A). Office of Solid Waste and Emergency Response. EPA/540/1-89/002. Available at: http://www.epa.gov/oswer/riskassessment/ragsa/index.htm

U.S. EPA (U.S. Environmental Protection Agency). (1991). Role of the Baseline Risk Assessment in Superfund Remedy Selection Decisions. Office of Solid Waste and Emergency Response, OSWER Directive 9355.0-30, April 22, 1991.

U.S. EPA (U.S. Environmental Protection Agency). (1993). EPA ROD EPA/ROD/R10-93/059 Allied Plating, Inc., EPA ID: ORD009051442 OU 01 Portland OR

U.S. EPA (U.S. Environmental Protection Agency). (1994). Methods for Derivation of Inhalation Reference Concentrations and Application of Inhalation Dosimetry. Office of Health and Environmental Assessment. Washington, DC. EPA/600/8-90-066F. October.

U.S. EPA (U.S. Environmental Protection Agency). (2001), Trichloroethylene Health Risk Assessment: Synthesis and Characterization - External Review Draft, Office of Research and Development, EPA/600/P-01/002A.

U.S. EPA (U.S. Environmental Protection Agency). (2002). A Review of the Reference Dose (RfD) and Reference Concentration (RfC) Processes. Risk Assessment Forum. EPA/630/P-02/002F, December.

U.S. EPA (U.S. Environmental Protection Agency). (2011a). Toxicological Review of Trichloroethylene, in support of summary information on the Integrated Risk Information System (IRIS). September 2011. EPA/635/R-09/011F.

U.S. EPA (U.S. Environmental Protection Agency). (2011b). Background Indoor Air Concentrations of Volatile Organic Compounds in North American Residences (1990 – 2005): A Compilation of Statistics for Assessing Vapor Intrusion. June 2011. EPA 530-R-10-001.

U.S. EPA (U.S. Environmental Protection Agency). (2012b). APTI Atmospheric Sampling Course Appendix G: Significant Figures and Rounding. Available at: http://www.epa.gov/apti/Materials/APTI%20435%20student/Student%20Manual/Appendix_G_no%20TOC-cover_MRpf.pdf

U.S. EPA (U.S. Environmental Protection Agency). (2012a). Human Health Risk Assessment Bulletins, Supplement to RAGS, USEPA Region 4 Front Page. Available at: http://www.epa.gov/region4/superfund/programs/riskassess/healtbul.html

WHO DWG (World Health Organization Drinking Water Guidelines). (2008, 2009). Guidelines for drinking water quality, 3rd Ed., incorporating first and second addenda (2008) and rolling revisions current to 2009. World Health Organization. Available at: <http://www.who.int/water_sanitation_health/dwq/chemicals/en/index.html>.

Williams, A.L., and DeSesso, J.M. (2008). Trichloroethylene and ocular malformations: analysis of extant literature. Int J Toxicol. 27: 81–95.

Yauck, J., Malloy, M., Blair, K., Simpson, P., and McCarver, D. (2004). Proximity of residence to trichloroethylene-emitting sites and increased risk of offspring congenital heart defects among older women. Birth Defects Res A Clin Mol Teratol. 70: 808-814. Available at: http://dx.doi.org/10.1002/bdra.20060.

Zaffran, S., and Frasch, M. (2002). Early signals in cardiac development. Circ Res. 91: 457-469.