ARCADIS
Infrastructure · Water · Environment · Buildings

Imagine the result

Flint Hills Resources Alaska, LLC

Supplement to the Revised Draft Final Human Health Risk Assessment

Flint Hills North Pole Refinery
North Pole, Alaska

May 30, 2014

ARCADIS

Brian Magee, Ph.D.
Vice President, Principal Toxicologist
Risk Assessment Technical Leader

Supplement to the Revised Draft Final Human Health Risk Assessment

Flint Hills North Pole Refinery
North Pole, Alaska

Prepared for:
Flint Hills Resources Alaska, LLC

Prepared by:
ARCADIS U.S., Inc.
1 Executive Drive
Suite 303
Chelmsford
Massachusetts 01824
Tel 978 937 9999
Fax 978 937 7555

Our Ref.:
B0081981.0085

Date:
May 30, 2014

This document is intended only for the use of the individual or entity for which it was prepared and may contain information that is privileged, confidential and exempt from disclosure under applicable law. Any dissemination, distribution or copying of this document is strictly prohibited.

ARCADIS

Table of Contents

ARCADIS

Supplement to the Revised Draft Final Human Health Risk Assessment

Flint Hills North Pole Refinery
North Pole, Alaska

1. Introduction

ARCADIS, U.S., Inc. (ARCADIS) prepared a *Revised Draft Final Human Health Risk Assessment for the Flint Hills North Pole Refinery* (ARCADIS, 2012). During the preparation of this risk assessment report, ARCADIS scientifically evaluated the existing Reference Doses (RfDs) and equivalent toxicological reference values for sulfolane, including those derived by the following sources: CCME (2006), ATSDR (2010, 2011), ToxStrategies (2010), TCEQ (2011) and EPA (2012a). More specifically, ARCADIS evaluated the United States Environmental Protection Agency (EPA's) Provisional Peer-Reviewed Toxicity Value (PPRTV) for sulfolane (EPA, 2012a) and concluded that EPA did not follow the best available science and EPA guidance in reaching the conclusions described in its PPRTV. In addition, ARCADIS evaluated the approach followed by ToxStrategies (2010) as described in a White Paper, entitled *Assessment of Toxicological Data for Sulfolane - Update II*. ARCADIS concluded that the RfD developed for sulfolane by ToxStrategies (2010) was based on the best available science. Nonetheless, ARCADIS independently derived a RfD for use in the risk assessment using an approach that differed modestly from that followed by ToxStrategies (2010). ToxStrategies (2010) and ARCADIS (2012) derived equivalent chronic and subchronic RfDs in accordance with the best available science and EPA guidance for evaluation of primary toxicology studies and the derivation of RfDs:

Chronic RfD	0.01 mg/kg-day
Subchronic RfD	0.1 mg/kg-day

As discussed in Section 2 below, since ARCADIS submitted ARCADIS (2012), the ToxStrategies (2010) study was independently and professionally peer-reviewed and published in the *Journal of Applied Toxicology* (Thompson, et al., 2012) ("*JAT* article"). The Thompson et al. (2012) study is a "professionally peer-reviewed document" and should be used to establish a site-specific clean-up level consistent with the toxicity hierarchy established by Alaska Department of Environmental Conservation (ADEC) in its *Risk Assessment Procedures Manual* (ADEC, 2000), which is incorporated by reference into 18 Alaska Administrative Code (AAC) 75.345. By this *Supplement to the Revised Draft Final Human Health Risk Assessment,* now that the ToxStrategies (2010) RfD derivation has been published, ARCADIS adopts and endorses the peer-reviewed analysis provided in that article and the corresponding chronic sulfolane RfD of 0.01 mg/kg-day. Based on this RfD, ARCADIS proposes and the JAT article supports a cleanup level of 362 µg/L as a health-protective standard for groundwater containing sulfolane in a residential setting.

2. The *JAT* Article Is a Newly Available, Independently Peer-Reviewed Publication Regarding the Sulfolane Reference Dose

The *Journal of Applied Toxicology* published an article based on the 2010 ToxStrategies analysis that derived a chronic RfD for sulfolane of 0.01 mg/kg-day (Thompson, et al., 2012) after the ARCADIS (2012) risk assessment was submitted. According to the publisher, the "*Journal of Applied Toxicology* publishes

peer-reviewed original reviews and hypothesis-driven research articles on mechanistic, fundamental and applied research relating to the toxicity of drugs and chemicals at the molecular, cellular, tissue, target organ and whole body level *in vivo* (by all routes of exposure) and *in vitro / ex vivo*" (Wiley, 2014). Accordingly, the work of Thompson and co-authors was carefully peer-reviewed by independent reviewers before it was published on August 31, 2012.

This peer-reviewed publication used benchmark dose modeling to derive a point of departure (POD) dose based on a benchmark response of one standard deviation difference from the mean white blood cell level in the control animals. Utilizing this approach, the authors derived a chronic RfD of 0.01 mg/kg-day. Notably, one of the co-authors of the article is Dr. David Gaylor, who has been performing original research on the benchmark dose method used in the report for decades. Dr. Gaylor is widely cited by EPA and was, in fact, a co-author of EPA's Benchmark Dose Technical Guidance (EPA, 2012b). He is a preeminent and recognized expert in the derivation of the benchmark dose and in the use of EPA's benchmark dose modeling software. As a peer-reviewed article consistent with EPA guidance, the *JAT* article supports the adoption of a chronic sulfolane RfD of 0.01 mg/kg-day under ADEC's toxicity hierarchy (ADEC, 2000).

3. The *JAT* Analysis Is Consistent with EPA Guidance and Precedent Regarding Benchmark Dose Modeling and Dose Transformation

The application of benchmark dose modeling is the principal difference between the *JAT* article and EPA's PPRTV for sulfolane (EPA, 2012a). Because the methodology of the *JAT* article is consistent with the preferred approach identified in EPA guidance (1995, 2000, 2012b, c, d; NRC, 2014; Appendix A), the *JAT* article is a more scientifically valid chronic RfD for sulfolane. The *JAT* article is consistent with EPA guidance in the following respects:

- The *JAT* article applies benchmark dose modeling, which is strongly preferred over the No Observed Adverse Effect Level (NOAEL) approach.
- The *JAT* article performed dose transformation, which is recommended by EPA when necessary to obtain adequate data fits to one or more models.
- The *JAT* article accounted for uncertainty in the RfD using uncertainty factors (UFs) selected in a manner consistent with EPA guidance and precedent.

3.1 Benchmark Dose Modeling – Guidance

The *JAT* article followed EPA guidance and used benchmark dose modeling instead of the NOAEL or Lowest Observed Adverse Effect Level (LOAEL) approach to derive the POD dose. Since the 1980's, EPA has recognized that the traditional approach for deriving RfDs by defining NOAELs or LOAELs from

toxicological studies is deficient. For instance, EPA (1995) discussed the advantages of the benchmark dose modeling approach as compared to the NOAEL approach as follows:

> Using the NOAEL in determining RfDs and RfCs [Reference Concentrations] has many limits (reviewed by Kimmel and Gaylor [1988] and others and noted by EPA's Science Advisory Board [U.S. EPA, 1986, 1988a, b, 1989]). These limitations include the following:
>
> - The experimental dose called the NOAEL is based on scientific judgment and is often a source of controversy....
> - The slope of the dose response plays little role in determining the NOAEL....
> - The NOAEL is limited to the doses tested experimentally....
>
> The EPA believes that the Benchmark Dose (BMD) approach presents a significant opportunity to improve the scientific basis of noncancer risk assessment. This document aims to encourage further application and development of the method by outlining the benchmark approach. It is hoped the BMD will add a new perspective to risk assessment and overcome some limitations of the NOAEL. To do this, the risk assessment community must first become familiar with the benchmark approach and its opportunities and limitations.

EPA continued to state its preference for the benchmark dose modeling versus the NOAEL approach over the years (EPA, 1995, 2000, 2012b, 2012c, 2012d), and in 2012, EPA issued guidance further explaining the reasons for considering benchmark dose modeling to be a more scientifically valid approach to deriving a POD dose than the use of a NOAEL (EPA, 2012b):

> The NOAEL is sometimes taken as an important point for describing a dose-response relationship in a study because of a presumed correspondence between such NOAELs and true thresholds (i.e., true no-effect levels). However, the NOAEL, which has generally been defined by a lack of statistical significance of the effect, is really a consequence of the fact that any finite study has an inherent limit of detection.
>
> ...Specific limitations of the NOAEL/LOAEL approach are well known and have been discussed extensively (Crump 1984; Gaylor 1983; Kimmel and Gaylor 1988; Leisenring and Ryan 1992; U.S. EPA 1995a):
>
> - The NOAEL/LOAEL is highly dependent on dose selection since the NOAEL/LOAEL is limited to one of the doses included in a study. . . .

- More generally, the NOAEL/LOAEL approach does not account for the variability and uncertainty in the experimental results that are due to characteristics of the study design such as dose selection, dose spacing, and sample size. . . .

- Other dose-response information from the experiment, such as the shape of the dose-response curve (e.g., how steep or shallow the slope is at the BMD, providing some indication of how near the POD might be to an inferred threshold), is not taken into account.

EPA's stated preference for the benchmark dose modeling approach extends to development of the PPRTVs. In the *Standard Operating Procedures (SOPs) for Developing Provisional Peer Reviewed Toxicity Values* (EPA, 2004), EPA states:

> If the available data are sufficient, dose-response modeling is the preferred method for determining the POD. . . . [P]roper use of this approach allows for a greater consideration of the dose-response function than the traditional NOAEL/LOAEL approach.
>
> Some advantages of using a dose-response modeling approach include:
>
> - It does not rely on the doses used in the study to determine a point of departure for calculation of risk values
>
> - It allows for the consideration of the entire dose-response curve, rather than relying on the results of a single dose level to describe the data [and]
>
> - It allows for an estimation of a NOAEL value even if the available studies report effects at every dose level examined . . . .

EPA scientists have also stated in scientific publications that the benchmark dose modeling approach is the preferred approach for the derivation of RfDs (Davis, et al., 2011; Zhao, et al., 2010). More importantly, the National Research Council recently reviewed EPA's standard practices for deriving RfDs and concluded: "Although the NOAEL-LOAEL approach remains in practice, the BMD approach is preferred because it provides and uses dose-response information to a greater extent and reduces uncertainty (EPA 2012)" (NRC, 2014). The use of benchmark dose modeling rather than the NOAEL approach in the *JAT* article is consistent with this preference.

3.2 Benchmark Dose Modeling – Precedent

EPA's preference for benchmark dose modeling is reflected in the numbers of RfDs and RfCs (both IRIS values and PPRTV values) that it has derived based on benchmark dose modeling: 68 chronic RfDs, 47

subchronic RfDs, 48 chronic RfCs, and 27 subchronic RfCs. The files for all 68 chronic RfDs were reviewed to determine the date of the RfD derivation. All 68 were dated 1997 or later, 57 (84%) were derived in 2005 or later, and 48 (71%) were derived in 2009 or later. More than half were derived in 2009 or later, showing that the benchmark dose modeling approach is being used more and more frequently.

3.3 Dose Transformation – Guidance

The *JAT* article followed EPA guidance in its approach to benchmark dose modeling. The *JAT* authors executed EPA's benchmark dose modeling software with the Huntington Life Sciences (HLS) (2001) data on white blood cells and lymphocytes and found that the data did not adequately fit any of the models. Following EPA guidance and the precedent set by EPA in the benzene reference dose derivation (discussed below), the authors log transformed the dose data and re-ran the models, obtaining adequate model fits for several of the models in the EPA software. Significantly, EPA did not take this step in preparing the sulfolane PPRTV (EPA, 2012a) and instead resorted to a NOAEL approach.

Log transformation of the data is explicitly recommended by EPA in guidance (EPA, 1995, 2000, 2012 b, c, d; Appendix A). For instance, EPA (1995) states: "...it may be necessary to transform continuous data in some cases so that they better satisfy the assumptions of a normal distribution. A log-transform is often used for this purpose." EPA (2012b) states: "Whenever none of the available models provides an adequate fit to the data, the modeler should first (re)consider data quality or experimental problems that may have been missed in the initial study evaluation (e.g., opportunistic infections, dosing errors; see Section 2.1.). Sometimes, adjustments to the data (e.g., a log-transformation of dose or adjustments for unrelated deaths) may be necessary." Similarly, when discussing acceptable adjustments to the data in the BMD Methodology Software Tutorial, EPA (2012d) states: "In certain cases, the typical models for a standard study design cannot be used with the observed data as, for example, when the data are not monotonic, or when the response rises abruptly after some lower doses that give only the background response. In these cases, adjustments to the data (e.g., a log-transformation of dose) or the model (e.g., adjustments for unrelated deaths) may be necessary." The authors of the *JAT* article followed the benchmark dose modeling approach as recommended in these EPA guidance documents.

3.4 Dose Transformation - Precedent

The approach in the *JAT* article is consistent with EPA precedents regarding the benefits of log transforming data. Importantly, dose transformation was used by EPA in its published IRIS document, *Toxicological Review of Benzene, (Noncancer Effects)* (EPA, 2002a). In the IRIS profile for benzene, EPA (2014a) states: "Most of the data were supralinear (i.e., the magnitude of the reductions in lymphocyte count decreased with increasing unit dose), and it was necessary to transform the dose data according to the formula d' = ln(d+1) in order to fit the available models." This regulatory precedent for log dose transformation concerns a data set that matches the data set for sulfolane. In both cases, the critical effect was defined as decreased white

blood cell counts; in both cases the data were supralinear; and in both cases simple log transformation of the raw data provided acceptable model fits.

Moreover, EPA routinely log transforms data when it executes the benchmark dose modeling software through the application of log-based models for dichotomous data. For instance, EPA (2012b) states with respect to dichotomous data sets: "In the absence of a biologically based model, dose-response modeling is largely a curve-fitting exercise among the variety of available empirical models." EPA further states that because there is no reason to apply one particular model "we fit a number of models to the data as show in Table A.1.2." In this table, EPA lists logistic, log-logistic, probit, and log-probit models, among others. EPA's software lists these four models in a "pull down menu" in a manner that allows the user to easily execute them both with and without log transforming the dose data. In practice, EPA routinely runs the models with log transformed doses. In addition, when running any model, whether for continuous or dichotomous data, the software contains a "pull down menu" that allows the user to transform the dose data in many different ways, including log dose transforming the data.

In addition to the many instances where EPA has run benchmark dose models after log transforming the dose data, many RfDs are specifically based on benchmark dose model runs in which the dose data were log transformed. In IRIS, there are 7 chronic RfDs based on log transformed doses out of 40 based on benchmark doses (18%) and there are 5 chronic RfCs based on log transformed doses out of 30 based on benchmark doses (17%). With regard to PPRTVs, there are 9 chronic RfDs based on benchmark doses out of 28 based on benchmark doses (32%), and there are 6 chronic RfCs based on log transformed doses out of 18 based on based on benchmark doses (33%).

Furthermore, log dose transformation has been used in a number of peer-reviewed scientific studies in which reference doses and reference concentrations were derived by benchmark dose modeling of critical effects data. Examples include:

- Budtz-Jorgensen, E., P. Grandjean, N. Keiding, R.F. White, and P. Weihe. 2000. Benchmark Dose Calculations of Methylmercury-Associated Neurobehavioural Deficits. *Toxicology Letters.* 112-113:193-9.
    - Benchmark doses that related both cord-blood and maternal hair mercury concentrations to neurobehavioral deficits in 7-year old Faroese children were calculated using a power function. The authors log (dose + 1) transformed the mercury dose parameter for benchmark dose modeling in exactly the same manner in which dose data were transformed in the *JAT* article and the EPA's benzene assessment. It was found that log transforming mercury cord-blood concentrations resulted in better model fits.
- TERA. 2005. Use of Benchmark Concentration Modeling and Categorical Regression to Evaluate the Effects of Acute Exposure To Chloropicrin Vapor Part I. Technical Report.

- Toxicology Excellence for Risk Assessment (TERA) performed benchmark concentration modeling using categorical regression to calculate a POD dose for ocular irritation associated with acute exposure to chloropicrin vapor. The final model used a log-transformed concentration (dose) parameter.

- Deutsch, R. C., & Piegorsch, W. W. 2012. Benchmark Dose Profiles for Joint-Action Quantal Data in Quantitative Risk Assessment. *Biometrics*, *68*(4), 1313-1322.

  - Benchmark dose modeling was performed using log-transformed dose data to estimate a POD dose for rates of cellular damage after human hepatic cells were exposed to various combinations of DDT and nano-TiO. The dose transformations applied in this study include $\log_{10}$(DDT dose)+4 and $\log_{10}$(TiO_2 dose)+3.

- Jiao, J., Feng, N. N., Li, Y., Sun, Y., Yao, W., Wang, W., ... & Xia, Z. L. 2012. Estimation of a safe level for occupational exposure to vinyl chloride using a benchmark dose method in central China. *J Occup Health*, *54*(4), 263-270.

  - Benchmark modeling was performed using a logistic model on log-transformed exposure concentrations to estimate safe levels of vinyl chloride exposure in workers from central China. The log-logistic model was used as the final model because it provided the best fit of three models to dose-response data.

- Wang, Q., Tan, H. S., Ma, X. M., Sun, Y., Feng, N. N., Zhou, L. F., ... & Xia, Z. L. 2013. Estimation of benchmark dose for micronucleus occurrence in Chinese vinyl chloride-exposed workers. *International journal of hygiene and environmental health*, *216*(1), 76-81.

  - Benchmark dose modeling was used to assess the dose-response relationship between occupational vinyl chloride exposure and chromosome damage in Chinese workers. Exposure concentrations were log-transformed and related to micronucleus frequency using a logistic model. The log-logistic model was selected as the final model because fitting statistics indicated that it outperformed the other five models considered.

- Wignall, J. A., Shapiro, A. J., Wright, F. A., Woodruff, T. J., Chiu, W. A., Guyton, K. Z., & Rusyn, I. 2014. Standardizing benchmark dose calculations to improve science-based decisions in human health assessments. *Environ Health Perspect*, *122*(5), 506-512.

  - Approaches were investigated for standardizing benchmark dose calculations to improve science-based decision making in human health assessments. The analysis, which included 255 chemicals with dose-response data, revealed that the log-logistic model was frequently the best performing model for describing dichotomous data sets.

4. The *JAT* Article Reports an Appropriately Protective Reference Dose That Accounts for Uncertainty in a Conservative Manner Consistent with EPA Guidance

The RfD derived in the peer-reviewed *JAT* article is conservative and appropriately health-protective for a variety of reasons. A principal reason that the RfD is health-protective is that it is based on the most sensitive endpoint reported in the HLS (2001) study, and this effect is likely not an *adverse* effect. The *JAT*

article also appropriately accounted for uncertainty by applying uncertainty factors in a manner that is consistent with EPA guidance as described below.

4.1 Endpoint Selection

The *JAT* article selected diminished white blood cell counts in female rats in the HLS (2001) study as the relevant endpoint (Thompson, et al., 2012). The *JAT* article approach is health-protective, because it is not known if the degree of white blood cell reduction observed in the study indeed resulted in an *adverse* effect or if the reduction in cells was an adaptive and reversible response. For instance, lymphocytes have a life span of 60-100 days in the rat (Suckow, 2006). Notably, EPA's IRIS dossier (EPA, 2014a) for the benzene assessment, which also used changes in blood cell counts as the relevant endpoint for setting an RfD, discussed the uncertainty associated with considering blood cell counts as adverse effects by stating: "With continuous endpoints such as hematological parameters, there is uncertainty about when a change in a parameter that has inherent variability becomes an adverse effect." Accordingly, this approach was a health protective choice that accounted for uncertainty regarding potential adverse effects of sulfolane.

As explained by Dr. Farland (2012, Appendix A), toxicologically relevant effects are those that are adverse, rather than adaptive. EPA guidance plainly provides that RfDs are based on adverse effects. EPA (1995) differentiates between adverse effects and non-adverse effects in defining a NOAEL as: "An exposure level at which there are no statistically or biologically significant increases in the frequency or severity of adverse effects between the exposed population and its appropriate control. Some effects may be produced at this level, but they are not considered as *adverse*, nor precursors to *adverse* effects." [Emphasis added] EPA recently highlighted the importance of this issue in clarifying its definition of NOAELs to require that an adverse effect be "biologically significant," not merely "statistically significant" (EPA, 2012b). A NOAEL is now defined as: "The highest exposure level at which there are no *biologically significant* increases in the frequency or severity of adverse effects between the exposed population and its appropriate control; some effects may be produced at this dose level, but they are not considered adverse or precursors of adverse effects." [Emphasis added] As described in the previous paragraph, the biological significance of the POD dose used by EPA based on the NOAEL approach is unclear.

In the absence of specific knowledge about the biological significance of a drop in circulating white blood cells from 8 billion per liter to 5 billion per liter in female rats (HLS, 2001), the *JAT* authors assumed that this was an adverse effect as a health-protective measure. Because there was no decrement in circulating white blood cells in the male animals and because that effect is not known to be an adverse effect, this approach is conservative, and the conservativeness of this approach should be considered when selecting uncertainty factors. Furthermore, the EPA default approach of setting the benchmark response at 1 standard deviation difference from the controls, which the *JAT* authors used, is also conservative given that normal ranges for circulating blood cell counts in rats are often set at +/- 2 standard deviations from the mean of control animals. Overall, assuming that this effect was adverse and using a conservative benchmark response to

define when a white blood cell decrement would be large enough to be considered *adverse* are health-protective choices that support relying on the *JAT* article for deriving a sulfolane reference dose.

4.2 Uncertainty Factors

The authors of the *JAT* article selected uncertainty factors in a manner that is consistent with EPA guidance and precedent and that resulted in derivation of an appropriately protective RfD (Appendix A). In accordance with EPA guidance (2002b), uncertainty factors were considered for the following:

- UF_A = animal-to-human (interspecies)
- UF_S = subchronic-to-chronic duration
- UF_D = database uncertainty
- UF_H = inter-individual (intraspecies)

The *JAT* article and the PPRTV do not differ with respect to first three of these UFs. The authors of the *JAT* article applied a UF_A of 3 based on the use of $BW^{3/4}$ scaling, consistent with EPA guidance on body weight scaling (EPA 2011). As discussed below, this UF_A is effectively greater than the UF_A of 10 used in the PPRTV (EPA, 2012a). The authors of the JAT article also considered a value of 1 to be potentially appropriate as endorsed by TCEQ (2011), but decided to apply the more conservative UF_A of 3.

The *JAT* article used the default UF_S of 10 as did the PPRTV (EPA, 2012a).

Finally, like the PPRTV (EPA, 2012a), the *JAT* article applied a UF_D of 3. The *JAT* article could arguably have been less conservative with respect to this uncertainty factor. EPA (2002b) guidance states:

> The database UF is intended to account for the potential for deriving an under protective RfD/RfC as a result of an incomplete characterization of the chemical's toxicity. In addition to identifying toxicity information that is lacking, review of existing data may also suggest that a lower reference value might result if additional data were available. Consequently, in deciding to apply this factor to account for deficiencies in the available data set and in identifying its magnitude, the assessor should consider both the data lacking and the data available for particular organ systems as well as life stages.

Given that the *JAT* article accounted for uncertainty by choosing an endpoint effect that may be adaptive and reversible, and applying a UF_A of 3 rather than 1, there was arguably no need to further account for uncertainty through a database uncertainty factor higher than 1.

With respect to the UF_H, the *JAT* authors applied a UF_H of 3 rather than the UF_H of 10 selected by EPA in the PPRTV because decreases in white blood cells were observed only in female rats, indicating that female rats represented a sensitive subpopulation. This approach reflects EPA guidance that supports reducing the interspecies uncertainty factor from the default value of 10 in situations where "data are sufficiently representative of the exposure/dose-response data for the most susceptible subpopulation(s)" (EPA, 2002b). As a result, using the more conservative default UF_H of 10 was not necessary to achieve an appropriately health-protective RfD because the uncertainty associated with inter-individual differences in susceptibility was accounted for in the selection of the most sensitive toxicological endpoint. Both EPA in its PPRTV Standard Operating Procedures document (EPA, 2004) and Dr. Farland (2012, Appendix A) have recognized the potential for "overlap between the identified areas of uncertainty." Accordingly, in deriving a RfD, making the most conservative choices at every turn will result in a RfD that is lower than necessary to protect human health.

The composite uncertainty factor in the *JAT* article was 300, although this was effectively the same as a composite uncertainty factor of 1,200 because the authors also addressed uncertainty related to interspecies extrapolation by using body weight scaling to a Human Equivalent Dose (EPA, 2011). EPA specifies that the UF_A of 10 is comprised of a two elements, one for pharmacokinetic differences between species and one for pharmacodynamic (sensitivity) differences between species. Each of these takes a value of approximately 3. When body weight scaling is used to compute a Human Equivalent Dose, the pharmacokinetic portion of the UF_A is not needed. One only uses the remaining factor of 3 as the UF_A. The Human Equivalent Dose factor in this case is 4, not 3, so the combined UF_A is essentially equal to 12, versus the default factor of 10. The composite uncertainty factor of 300 applied after converting the rodent dose to a Human Equivalent Dose was sufficient to provide an appropriately health-protective RfD. Overall, the *JAT* author's decisions are consistent with EPA guidance, which recognizes that: "Sound scientific judgment should be used in the application of UFs to derive reference values that are applied to the value chosen for the POD derived from the available database" (EPA, 2002b).

An overview of the health-protectiveness of the *JAT* RfD can be presented by calculating the Margin of Exposure as discussed by Dr. Farland (2012). The Margin of Exposure is defined as the ratio of the dose that a human receptor receives compared to the dose that was associated with an adverse effect in the study from which the RfD was derived, in this case, an animal study. With the HLS (2001) study, the dose associated with adverse decrements in white blood cells in the entire population is unknown, because no decrements were seen in males and the decrements seen in females are not known to be adverse or irreversible. What is known is that the dose at which truly adverse effects would be expected is greater than 16 mg/kg-day, which is the POD dose based on a one standard deviation decrement from the female control animals. The ratio of >16 mg/kg-day to 0.01 mg/kg-day is >1,600. So a person receiving the RfD of 0.01 mg/kg-day every day for their entire life would be receiving a dose of sulfolane that is >1,600 times *lower* than the lowest level that caused an effect in the HLS (2001) animal study, which *may* have been an adverse effect, but also may have just been an adaptive, reversible response. A Margin of Exposure of >1,600 is

certainly adequate to protect human health, so the RfD of 0.01 mg/kg-day is conservative and health-protective.

5. Existing ADEC and EPA Guidance Supports Using an Adult Receptor as a Health-Protective Basis for an Alternative Cleanup Level

The *JAT* article presented a sulfolane "screening value" of 365 µg/L based on calculations using an adult receptor that is sufficient to protect individuals over a full lifetime of exposure. Although this level was characterized as a screening level, it was based on the same exposure assumptions regarding the appropriate receptor and other appropriate variables identified in ADEC (2008) for use in calculating a groundwater cleanup level. The *JAT* article's choice of an adult receptor is also consistent with EPA (2012e) guidance and practice in calculating health-protective drinking water standards. Overall, the result provides a health-protective level that ADEC should now adopt as a cleanup level (subject to minor adjustments for site-specific factors related to intake from fruit and vegetable consumption). As explained in ARCADIS (2012) applying site-specific assumptions about consumption of fruits and vegetables assumed to be watered with sulfolane-containing groundwater results in a slightly lower cleanup level of 362 µg/L.

5.1 ADEC Guidance

ADEC's 2008 Cleanup Level Guidance specifically provides an equation (Equation1, Table C) to use when deriving Groundwater Cleanup Levels for noncarcinogenic constituents as incorporated by reference in 18 AAC 75.340(e)(1). This equation is shown below:

$$\text{Cleanup Level (mg/L)} = \frac{\text{THQ} \times \text{RfD}_o \times \text{BW} \times \text{AT} \times 365\ \text{d/yr}}{\text{IR} \times \text{EF} \times \text{ED} \times \text{A}}$$

The parameters and their required values are listed below:

Parameter/Definition (units)	Default
THQ/target hazard quotient (unitless)	1
BW/body weight (kg)	70
AT/averaging time (yr)	30
RfD_o/oral reference dose (mg/kg-d)	Chemical-specific (Table 2)
EF/exposure frequency (d/yr)	350
ED/exposure duration (yr)	30
IR/ ingestion rate (L/d)	2
A/absorption factor	1

This equation for a 70 kg receptor who consumes 2 liters of water a day for 30 years defines an adult receptor. There is no equation for groundwater consumption by a child. The ADEC guidance specifies that the ACL for groundwater should be based on an adult receptor. These are the same exposure assumptions used in the *JAT* article and ARCADIS (2012)

5.2 EPA Guidance

ADEC's guidance is consistent with calculations used by EPA and states in accordance with the Safe Drinking Water Act (SDWA) to compute the drinking water equivalent level-the concentration of a contaminant in drinking water that will have no adverse health effect over a lifetime of potential exposure, including potential exposures to sensitive subpopulations (EPA, 2012e). The drinking water equivalent level serves as the basis for the federal drinking water standards. In addition, EPA calculates drinking water unit risk factors, which are measures of the risk associated with a chemical in drinking water, in the Integrated Risk Information System (IRIS) using adult (70 kg) exposures of 2 L/day over a lifetime. As an example, the IRIS profile for benzene (EPA, 2014a) states the following: "The drinking water unit risk was then calculated from the oral slope factor assuming a drinking water intake of 2 L/day."

Children are considered by the federal Office of Water in the calculation of One-day and Ten-day Health Advisories (HAs). The *lifetime* HA, however, is based on the adult. EPA publishes "concentrations of drinking water contaminants at which noncancer adverse health effects are not anticipated to occur over specific exposure durations - One-day, Ten-day, and Lifetime - in the *Drinking Water Standards and Health Advisories* (DWSHA) tables. The One-day and Ten-day HAs are for a 10 kg child and the Lifetime HA is for a 70 kg adult." The lifetime HA is always more protective than a One-day or Ten-day Health Advisory.

In each of the above cases, an adult receptor is used to calculate a level sufficient to provide protection against adverse effects over a lifetime of exposure (Appendix A). ARCADIS acknowledges that EPA derives Regional Screening Levels (RSLs) for all media, including groundwater, based on a child receptor. However, these are not cleanup levels or drinking water standards. They are, as their name implies, *screening levels.* These screening levels are used in the Constituent of Potential Concern selection step of a human health risk assessment to identify the constituents that will be quantitatively included in the risk assessment. In fact, the EPA webpage (EPA, 2014b) that discusses the RSLs specifically states: "The SLs presented in the Generic Tables are chemical-specific concentrations for individual contaminants in air, drinking water and soil that may warrant further investigation or site cleanup. **It should be emphasized that SLs are not cleanup standards.**" [Note: bold typeface is in the original source.]

6. Conclusions

In conclusion, the *JAT* article has reviewed the scientific data on sulfolane and has derived an appropriately health-protective RfD that is consistent with EPA guidance and precedent. The *JAT* authors' determined the POD dose using best available science by performing benchmark dose modeling and also conservatively

assumed that an observed toxicological endpoint was an *adverse* effect despite any information that such an effect was truly *adverse*. In addition, uncertainty factors were chosen using standard EPA guidance, and these factors were health-protective. Lastly, the Alternative Cleanup Level of 362 µg/Lin groundwater derived by ARCADIS using the RfD derived in the *JAT* article was derived using an adult receptor to calculate a groundwater concentration level that provided sufficient protection against adverse effects over a lifetime of exposure. The use of the adult receptor is consistent with ADEC (2008) and EPA (2012e) guidance. ARCADIS proposes and the *JAT* article supports a health-protective cleanup level of 362 µg/L for groundwater containing sulfolane in a residential setting.

7. References

ADEC. 2000. Alaska Department of Environmental Conservation. Contaminated Sites Remediation Program. Risk Assessment Procedures Manual. June 8, 2000.

ADEC. 2008. Alaska Department of Environmental Conservation. Cleanup Level Guidance. June 9, 2008. http://dec.alaska.gov/spar/csp/guidance/cleanuplevels.pdf.

ARCADIS. 2012. Revised Draft Human Health Risk Assessment. Flint Hills North Pole Refinery, North Pole, Alaska. May 2012.

ATSDR (Agency for Toxic Substances and Disease Registry). 2010. Health Consultation. Sulfolane. Department of Health and Human Services, Division of Toxicology and Environmental Medicine Prevention, Response and Medical Support Branch Emergency Response Team. Atlanta, GA. (February 3, 2010).

ATSDR (Agency for Toxic Substances and Disease Registry). 2011. Health Consultation. Sulfolane. Department of Health and Human Services, Division of Toxicology and Environmental Medicine Prevention, Response and Medical Support Branch Emergency Response Team. Atlanta, GA. (May 2, 2011)

Budtz-Jorgensen, E., P. Grandjean, N. Keiding, R.F. White, and P. Weihe. 2000. Benchmark Dose Calculations of Methylmercury-Associated Neurobehavioural Deficits. *Toxicology Letters* 112-113:193-9.

CCME (Canadian Council of Ministers of the Environment). 2006. Canadian soil quality guidelines for the protection of the environmental and human health; Sulfolane. Winnipeg, Canada.

Clewell, H. J., G.A. Lawrence, D.B. Calne, and K.S. Crump. 2003. Determination of an Occupational Exposure Guideline for Manganese Using the Benchmark Method. *Risk Analysis* 23: 1031–1046.

Davis, J.A., J.S. Gift, and Q.J. Zhao. 2011. Introduction to benchmark dose methods and US EPA's benchmark dose software (BMDS) version 2.1.1. *Toxicol. Appl. Pharmacol.* 254:181-191.

Deutsch, R. C., & Piegorsch, W. W. 2012. Benchmark Dose Profiles for Joint-Action Quantal Data in Quantitative Risk Assessment. *Biometrics*, *68*(4), 1313-1322.

EPA (United States Environmental Protection Agency). 1995. The Use of the Benchmark Dose Approach. In Health Risk Assessment. Risk Assessment Forum. EPA/630/R-94/007. February.

EPA. 2000. Benchmark Dose Technical Guidance Document. External Review Draft. EPA/630/R-00/001. Washington, DC: Risk Assessment Forum

EPA. 2002a. Toxicological Review of Benzene (Noncancer Effects) (CAS No. 71-43-2) In Support of Summary Information on the Integrated Risk Information System (IRIS). EPA/635/R-02/001F.

EPA. 2002b. A Review of the Reference Dose and Reference Concentration Processes. EPA/630/P-02/002F. Risk Assessment Forum. December 2002.

EPA. 2004. Standard Operating Procedures (SOPs) for Developing Provisional Peer Reviewed Toxicity Values. March 2004.

EPA. 2011. Recommended Use of Body Weight$^{3/4}$ as the Default Method in Derivation of the Oral Reference Dose. Final. EPA/100/R11/0001. Office of the Science Advisor, Risk Assessment Forum, Washington, DC.

EPA. 2012a. Provisional Peer-Reviewed Toxicity Values for Sulfolane (CASRN 126-33-0). Final. 1-30-2012. Washington, DC: Office of Research and Development, National Center for Environmental Assessment, Superfund Health Risk Technical Support Center.

EPA. 2012b. Benchmark Dose Technical Guidance. EPA/100/R-12/001. Risk Assessment Forum. Washington, DC. June 12.

EPA. 2012c. BDMS 2.2. Benchmark Dose Modeling Software Manual.

EPA. 2012d. Benchmark Dose (BMD) Methodology. Software Tutorial.

EPA. 2012e. Drinking Water Standards and Health Advisories. April. http://water.epa.gov/action/advisories/drinking/upload/dwstandards2012.pdf.

EPA. 2014a. IRIS Profile: Benzene (CASRN 71-43-2). Accessed online at: http://www.epa.gov/iris/subst/0276.htm

EPA. 2014b. Regional Screening Level User's Guide. http://www.epa.gov/reg3hwmd/risk/human/rb-concentration_table/usersguide.htm

Farland, William. 2012. Sulfolane Hazard Characterization – Considerations. Appendix K. Revised Draft Human Health Risk Assessment. Flint Hills North Pole Refinery, North Pole, Alaska. May 2012 (ARCADIS, 2012)

Gaylor, D., L. Ryan, D. Krewski, Y. Zhu. 1998. Procedures for Calculating Benchmark Doses for Health Risk Assessment. *Regulatory Toxicology and Pharmacology* 28: 150-164.

HLS (Huntingdon Life Sciences). 2001. *Sulfolane toxicity study by oral administration via the drinking water to Cd rats for 13 weeks: Volume one.* Huntingdon, England: Huntingdon Life Sciences Ltd.

Jiao, J., Feng, N. N., Li, Y., Sun, Y., Yao, W., Wang, W., ... & Xia, Z. L. 2012. Estimation of a safe level for occupational exposure to vinyl chloride using a benchmark dose method in central China. *J Occup Health*, *54*(4), 263-270.

NRC. 2014. Review of EPA's Integrated Risk Information System (IRIS) Process.

Suckow M.A., S.H. Weisbroth, and C.L. Franklin. 2006. *The Laboratory Rat.* Elsevier Academic: Burlington, MA.

TERA. 2005. *Use of Benchmark Concentration Modeling and Categorical Regression to Evaluate the Effects of Acute Exposure To Chloropicrin Vapor Part I.* Technical Report.

TCEQ (Texas Commission on Environmental Quality). 2011a. Sulfolane; CASRN 126-33-0. Austin, TX. (September 6, 2011).

Thompson, C., Gaylor D., Tachovsky J., Perry C., Carakostas, M. and Haws L. 2012. Development of a chronic noncancer oral reference dose and drinking water screening level for sulfolane using benchmark dose modeling. *Journal of Applied Toxicology.* August 31.

ToxStrategies (ToxStrategies, Inc.). 2010. Assessment of Toxicological Data for Sulfolane - Update II. August 18, 2010.

Wang, Q., Tan, H. S., Ma, X. M., Sun, Y., Feng, N. N., Zhou, L. F., ... & Xia, Z. L. 2013. Estimation of benchmark dose for micronucleus occurrence in Chinese vinyl chloride-exposed workers. *International journal of hygiene and environmental health*, *216*(1), 76-81.

Wignall, J. A., Shapiro, A. J., Wright, F. A., Woodruff, T. J., Chiu, W. A., Guyton, K. Z., & Rusyn, I. 2014. Standardizing benchmark dose calculations to improve science-based decisions in human health assessments. *Environ Health Perspect*, *122*(5), 506-512.

Wiley. 2014. http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1099-1263/homepage/ProductInformation.html (accessed 5-12-14)

Zhao, Q.J., L. Haber, M. Kohrman-Vincent, P. Nance, M. Daurson. 2010. Quantitative modeling in noncancer risk assessment. In *Quantitative Modeling in Toxicology*, Krishnan K, Andersen ME (eds). John Wiley & Sons: Chichester.

ARCADIS

Appendix A

Perspectives on the Journal of Applied Toxicology Article entitled "Development of a chronic non-cancer oral reference dose and drinking water screening level for sulfolane using benchmark dose modeling"

Perspectives on the Journal of Applied Toxicology Article entitled "Development of a chronic non-cancer oral reference dose and drinking water screening level for sulfolane using benchmark dose modeling"[1]

William H. Farland, PhD, ATS

May 30, 2014

Summary-

The database on sulfolane has been evolving over the last 3 decades. Relatively speaking, compared to many industrial chemicals encountered in the environment, the available data and details of their generation are quite robust. It has been generally recognized that there is sufficient information on sulfolane to derive scientifically-defensible toxicity values based on these data. This Journal of Applied Toxicology article provides a peer-reviewed analysis of the data and demonstrates state-of-the-science approaches to benchmark dose modeling to derive a reference dose and tap-water screening level that adhere carefully to EPA's published methods, guidance and precedents. It provides a clear explanation of the rationale for choices made, while also discussing alternatives. It provides a balanced perspective on uncertainties and opts to use public health protective values in the face of these alternatives. It compares these values with previous attempts to assess the sulfolane database and provides a significant advance over previous NOAEL/LOAEL-based efforts. Inclusion of this study, which was carried out by experienced toxicologist/risk assessors and includes one of the "fathers" of the benchmark dose (BMD) methodology, in a peer-reviewed, well respected journal suggests to me the need to re-evaluate previous efforts carried out by ATSDR and US EPA.

Methods-

The authors have provided an explanation of their approach to collection of the sulfolane toxicity testing database. Their approach is comprehensive and could easily be replicated by others, given the information provided. The only exception to this is the statement that other "proprietary resources were used when available." In reviewing the modeling efforts and results presented, there is no indication that "proprietary resources" had any impact on these efforts. The authors modeled dose-response for the noted effects using the US EPA's BMD Software (BMDS). They followed approaches suggested by EPA for both the continuous and dichotomous data sets. Model fits were evaluated, as suggested by US EPA, using criteria such as p-values, scaled residuals, Akaike information criteria, parsimony and visual inspection.

[1] Thompson, C.M., Gaylor, D.W., Tachovsky, J.A., Perry, C., Carakostas, M.C., Haws, L.C. Development of a chronic noncancer oral reference dose and drinking water screening level for sulfolane using benchmark dose modeling. J Appl Toxicol. 2012 Aug 31. doi: 10.1002/jat.2799.

Database-

The toxicologic testing database on sulfolane is relatively robust although, as mentioned by the authors, is "modest relative to some widely studied compounds." The data available include genotoxicity studies, acute and subchronic toxicity studies in multiple species by various routes of exposure, a chronic oral toxicity study, reproductive and developmental toxicity studies in multiple species by various routes of exposure and carcinogenicity studies involving sulfolene, a structurally-related compound. I have discussed the nature and quality of the database on this relatively well-studied chemical in a previous assessment (Farland, 2012). No additional, new information is included in this article.

Aside from frank effects seen in acute studies within an order of magnitude (factor of ten) of the very high doses causing lethality, other manifestations of toxicity are lacking in longer term, lower dose studies. The partial lifetime (subchronic) studies in particular suggest toxicological investigations without appreciable low dose toxicological effects. Carcinogenicity does not appear to be of concern since genotoxicity studies have been mostly negative and a lifetime cancer study in animals of a similar compound (sulfolene) raised no concerns. The focus of attention at low doses in subchronic studies has been on the observation of subtle changes, which are generally considered to be of unclear toxicological significance. Because these effects are the only ones seen at the lower doses studied, they have been modeled as if they were indications of adverse, rather than adaptive responses, to sulfolane exposure. The use of these data in this way provides a public health conservative approach to generation of a point of departure that suggests little to no toxicologic concern from the animal studies and is an uncertain indicator of potential toxicity to humans.

Given the focus of the assessment in this article on derivation of a chronic oral RfD, acute studies and those related to inhalation, irritation and sensitization were not reviewed. The summary of the studies considered in this assessment are provided in Table 1 in the article. Strengths and weaknesses of the studies are described in the text. For instance, the authors note the limited information available on the reported results of the 90-day toxicity studies by Zhu et al. (1987). They state, "Overall, these data could not be reanalyzed statistically nor were they amenable to quantitative dose-response modeling." This is consistent with a characterization of these study reports by the US EPA in their PPRTV document (US EPA, 2012a). The US EPA states, "This report appears to be an extended abstract of the original study with very little useful information for risk assessment purposes. There is, for example, no clear indication of histopathological examination of any tissues in any test described, save for the spleen and liver in the 6-month study. This lack of results precludes assigning any effect levels at least to the 90-day test reports." I concur with this assessment. The drinking water study in rats from the Huntingdon Laboratories (HLS), on the other hand, although available but unpublished, provided sufficient detail on methods and results to be useful for this assessment. The HLS study was well documented as is required of studies adhering to Good Laboratory Practice (GLP) requirements, and studied lower doses and a wider range of toxicologic endpoints than other available studies.

While likely subject to internal review at the Huntingdon Laboratories, the study was also subject to an independent peer review as part of the EPA PPRTV review process. The balance of the database was similarly assessed by the article's authors, noting similar effects in different studies where evident.

Overall, the concise description of the toxicologic database in this article appears complete and consistent with previous work. It also provides a clear and reasonable basis for the selection of the data to be modeled for dose-response, although as noted above, this represents a public health conservative approach to risk assessment given the uncertain significance of the effects observed and chosen which needs to be fully considered as the outcome of the assessment is considered and applied by decision-makers.

Dose-response modeling-

The authors of this article take the approach that, where the data allow, dose-response should be modeled rather than simply using a generally outdated NOAEL/LOAEL approach. They cite several of the nine (9) limitations of the NOAEL/LOAEL approach listed and referenced by the US EPA (2012b) in support of their approach. Use of the BMD approach is wholly consistent with the prevailing thought of the risk assessment community, in my opinion. The authors cite several references supporting this view. These include guidance from the US EPA, and the European Food Safety Agency (EFSA) as well as a recent (2010) text on quantitative methods in no-cancer risk assessment. This topic has also been part of the input by the National Research Council (NRC) on the evolution of risk assessment methods (See, for example, NRC (2009)). In addition, the NRC (2014) in its recent "Review of EPA's Integrated Risk Information System (IRIS) Process" echoes the US EPA as it references the benchmark dose approach as the "preferred" approach to setting a POD, indicating that the NOAEL/LOAEL approach should only be used if the data are inadequate for BMD modeling. The paradigm shift from the NOAEL/LOAEL approach to the preferred BMD modeling, with its use of more of the available data and focus on approximating the lower end of the range of observation, is now clearly established for appropriate data sets.

Modeling of the Zhu et al. (1987) data from the 6-month studies illustrates the authors' approach to the modeling of dichotomous data sets. The best fitting model (log-logistic) was chosen based on best fit for all three data sets. Among the three endpoints, fatty liver (steatosis) provided the lowest $BMDL_{10}$ value. Several other "alternative" dichotomous models (Davis et al., 2011) were tested and rejected based on fit and appropriateness of the model. Only one of these alternatives (dichotomous-Hill model) provided a lower BMDL value. The authors extended their analysis of this model by applying several scenarios where hypothetical higher doses and responses were modeled. Based on the results of these hypothetical scenarios and model behavior the dichotomous-Hill model was not considered appropriate for modeling these data sets. The log-logistic modeling of steatosis in the guinea pig resulted in a $BMDL_{10}$ value of 22.6 $mgkg^{-1}$ per day. I was pleased to see the rigorous attempt by the authors to assess BMD modeling for this

data set and concur with the decision that they made regarding selection of the log-logistic model. The authors reached a reasonable, data-informed conclusion that this was the most scientifically defensible POD value for the Zhu (1987) six month study.

The authors also modeled the continuous data sets from the HLS (2001) study. While this was the best documented (GLP-compliant) study in the database, the toxicologic effects on blood cells were of unclear toxicologic significance to both the study authors and multiple reviewers of the study, as previously discussed (Farland, 2012). As noted by the authors, initially none of the models in the BMDS would reasonably fit the data. One of the approaches recommended by benchmark dose modeling practitioners, including US EPA, in these circumstances is to drop the highest dose to improve the fit and place more reliance of the data closer to POD. However, recognizing that there was no evidence for frank toxicity or a plateauing of the responses, the authors reasonably rejected this approach. They did, however, recognize that the data was characterized by the two lower doses spanning a small percentage (5.5%) of the total dose range. They chose the scientifically supportable approach of log transformation of the doses to more evenly space the doses and reduce the influence of the highest dose without arbitrarily dropping it. Use of log transformation in BMD modeling is discussed by the US EPA (2012b) and is common practice among modelers (see for example, Wignall, et al. (2014)). This decision was further supported by the precedent established by US EPA in their benzene assessment (US EPA, 2002) where US EPA log transformed the doses when they modeled a reduction in lymphocytes in humans exposed to benzene to establish their RfC and RfD values. Applying the same approach, the authors of this article found a reasonable fit for linear as well as other models for total WBC and lymphocyte counts. In addition, the authors considered the use of available historical control data in lieu of the concurrent control data from the HLS study, thereby providing a "much more robust data set for establishing the normal range" which is consistent with US EPA guidance (US EPA, 2000, 2012b). The authors provide a reasonable explanation for their choice of the linear model of the log-transformed data based on well-established model selection criteria and a rationale similar to that used in the US EPA benzene assessment, i.e. parsimony (US EPA, 2002). Additionally, a dichotomous BMD analysis of the blood cell data from the HLS study as well as BMD modeling of the developmental toxicity data described by OECD (2004) was discussed by the authors.

Results of the dose response modeling efforts are presented in Table 7 of the article. PODs based on BMDL values range from 16 to 38.1 $mgkg^{-1}$ per day for the subchronic effects in rats and chronic effects in guinea pigs and at 120 $mgkg^{-1}$ per day for reproductive and developmental toxicity. The PODs represent a relatively narrow (less than one order of magnitude) range based on a variety of effects, several of which are of unclear toxicologic significance.

The approach to BMD modeling presented in this article is consistent with the state-of-the-science, rigorously applied and well explained. It is a good example of how complex data sets should be assessed for use in deriving risk reference values using today's science.

Chronic RfD derivation-

Table 7 also shows the calculated human equivalent doses (HEDs) for the PODs based on allometric scaling ($BW^{3/4}$) in the absence of an available comparative toxicocokinetic model, citing current US EPA practices (US EPA, 2011a). US EPA has stated that in the absence of a toxicokinetic model or other appropriate scaling approaches, "...body weight scaling to the ¾ power (i.e., $BW^{3/4}$) is endorsed as a general default procedure to extrapolate toxicologically equivalent doses of orally administered agents from all laboratory animals to humans for the purposes of deriving an oral Reference Dose (RfD). Use of $BW^{3/4}$ scaling in combination with a reduced default interspecies uncertainty factor, UFA, is recommended as the Agency default approach to replace the previous default approach for this purpose which involved $BW^{1/1}$ scaling with a full uncertainty factor (i.e., a UFA value of 10)." The authors of this article have correctly adopted this approach and have provided a robust discussion of their rationale for choice of values for the four typical uncertainty factors (UFs) employed to derive a reference value. Their clear description of their choices should engender support for this assessment, although, inevitably, as with all assessments requiring scientific judgment, there will be some discussion regarding their choices. Nonetheless, I endorse their choices based on my knowledge of risk assessment and their rationale. Ultimately, selection of the individual uncertainty factors, discussion of alternative approaches and consideration of conservatism in the name of public health resulted in composite UFs of 300 for all the PODs presented and a range of reference values of 4X (0.01-0.04 $mgkg^{-1}$ per day).

Modeled data on leukopenia from the HLS study seem to represent the most sensitive endpoint among the options presented. The authors have chosen to treat this endpoint as "adverse" despite the uncertain toxicologic significance of these effects which were noted by the study authors and reviewers. The issue of adverse versus adaptive responses in this context has been discussed elsewhere (Farland, 2012). As mentioned above, using these effects provides an extra measure of public health conservatism but, in the absence of the linkage of observations like those described above with potential human disease outcomes, the distinction between adverse and adaptive becomes blurred and use of these endpoints for other than screening purposes becomes problematic. This point should be carefully considered when applying such reference values to human health protection.

Risk-based screening level for drinking water-

The presentation of a risk-based screening level for drinking water uses the equation for an adult (70 kg) consumer of 2 liters per day over a 30 year period as is standard practice. I, along with the US EPA Drinking Water Program, have stated that it is appropriately health protective to assess chronic exposure scenarios for a chemical like sulfolane by using an RfD-like value with an adult body weight and ingestion rate. (Farland, 2012). The authors refer to the US EPA Regional screening level equations found in Superfund guidance. However, reference to the

Regional screening guidance is missing. US EPA (2011) in the article reference list is not to that guidance. I refer to it here as US EPA (2009). In addition to the adult tapwater value used in this article, this Regional guidance also shows an equation and parameters for derivation of a screening level for tapwater exposure to a child which results in a 2.3X lower regional screening level (156 versus 365 ppb). While it may be appropriate to use this approach to be fully protective of children for certain irreversibly toxic or accumulative chemicals, this does not appear to be the case with sulfolane exposure. Therefore, I can support the authors' choice of the equation and parameters in setting a risk-based screening level for drinking water. See below for further discussion.

Other points for consideration-

Within the Discussion in the article, the authors compare their findings to previous assessments, recognizing significant differences in selection of the critical study, in differences in methods for the derivation of the POD, and in selection of UFs. They also make a compelling case for the use of BMD modeling as opposed to the NOAEL/LOAEL approach when the data allow. The BMD approach for sulfolane described in this article is an improvement over previous approaches as it uses more of the data and carefully inspects the applicability of various models. In discussing the more recent US EPA provisional peer-reviewed toxicity value (PPRTV), the authors highlight the differences in methods to derive the POD, namely the lack of use of the BMD approach by US EPA. US EPA's failure to explore dose transformation and rigorously test the fit of the models as these authors have done is noteworthy. In addition, the authors point out the differences that led to US EPA using the maximal accepted composite UF (3,000) despite the reasonable scientific case that can be made for the use of lower UFs. It would appear that in the derivation of the PPRTV, US EPA (2012a) missed an opportunity to use the best available approaches and follow their own guidance. These authors have provided a compelling alternative assessment when compared to the PPRTV.

Of particular interest is the discussion of the ATSDR (2011) BMD-derived public health action level. The rationale for the use of the Zhu six-month data remains controversial as described previously. The use of the dichotomous-Hill model is problematic given the reasons articulated in the article, relating to the sensitivity of the model to the assumption of achievement of a maximal response. It is interesting to note that the authors believe that, if ATSDR had used the log logistic model, their action level would have been in the range of the reference values described above.

As mentioned previously, it is my view that it is appropriately health protective to assess chronic exposure scenarios for a chemical like sulfolane by using an RfD-like value with an adult body weight and ingestion rate. As discussed in Farland (2012), the use of an adult body weight and water consumption level has its basis in US EPA Drinking Water Standards and Health Advisories (US EPA, 2011b). In this document a "Lifetime Health Advisory" is defined as "the concentration of a chemical in drinking water that is not expected to cause any adverse non-

carcinogenic effects for a lifetime of exposure. The Lifetime HA is based on exposure of a 70-kg adult consuming 2 liters of water per day." One day or ten day health advisories use different assumptions regarding acute responses and a body weight of 10 kg and 1 liter a day consumption to protect infants for short durations of exposure when their body weight and consumption patterns could result in higher relative exposures. However, the assumption is that these short duration, higher exposure concerns are adequately accounted for by use of chronic RfD-like values for longer term (lifetime) exposures. Studies of "community water" consumption support these default values of 2 liters for lifetime exposure and 1 liter for infants' and children's exposure as representing the 80-90th percentile of the population values with mean consumption values being closer to half these values. It is considered fully protective of health to combine a chronic RfD-like value, which by definition is protective against appreciable risk for a lifetime of exposure for the population, including sensitive subpopulations and life-stages, with exposure values that represent the greatest part of a lifetime exposure.

An alternative approach has been chosen by the EPA Superfund program. The EPA Superfund program has developed a consensus approach to the calculation of screening levels (SLs) which are developed using EPA risk assessment guidance and can be used for Superfund sites. A discussion of SLs can be found at http://www.epa.gov/reg3hwmd/risk/human/rb-concentration_table/index.htm (USEPA, 2009). The SLs are described as "risk-based concentrations derived from standardized equations combining exposure information assumptions with EPA toxicity data. SLs are considered by the Agency to be protective for humans (including sensitive groups) over a lifetime." In the case of drinking water exposure, SLs include an assumption that the use of a chronic RfD-like value, coupled with an assumption of exposure parameters of 1 liter per day consumption for a 15 kg child, will generate a drinking water SL that is protective for the population with a lifetime of exposure. While the SL takes a more conservative approach, the HA value and the SL differ only by a factor of 2.3 times (70kg/2liters/day divided by 15kg/1liter/day). This difference is well within the inherent uncertainty of the RfD-like estimate itself and can be contrasted with the magnitude of the composite uncertainty factor which renders the estimate of the RfD-like value to be 1,000-10,000 times below observed subtle effects in animals. USEPA is clear to point out that SLs are generic screening values, not de facto cleanup standards. The SL approach is used to assess acceptable levels of both carcinogenic and non-carcinogenic effects and accounts for the possibility of shorter-term, age-specific exposures leading to toxicity. The available toxicity database for sulfolane supports neither a concern for irreversible effects of early exposures nor age-specific sensitivity of children. Site-specific decisions, taking this issue and others discussed above into account, should determine how the SLs will impact remediation goals for sulfolane.

References-

Agency for Toxic Substances and Disease Registry (ATSDR). 2011. ATSDR Health Consultation - Sulfolane. Division of Toxicology and Environmental Medicine Prevention, Response and Medical Support Branch.

Davis, JA et al. 2011. Introduction to benchmark dose methods and US EPA's benchmark dose software (BMDS) version 2.1.1. Toxicol. Appl. Pharmacol. 254: 181-191.

Farland, WH . 2012. Sulfolane Hazard Characterization – Considerations. Prepared for Flint Hills Resources, April 5, 2012.

HLS (Huntingdon Life Sciences), 2001. Sulfolane toxicity study by oral administration via the drinking water pathway to CD rats for 13 weeks. Huntingdon Life Sciences Ltd. Huntingdon, England.

National Research Council (NRC). 2009. Science and Decisions, Advancing Risk Assessment. National Academy Press, Washington, DC.

National Research Council (NRC). 2014. Review of EPA's Integrated Risk Information System (IRIS) Process. National Academy Press, Washington, DC.

OECD. 2004. SIDS initial assessment report for: tetrahydrothiophene-1,1-dioxide. CASRN 126-33-0.

United States Environmental Protection Agency (US EPA). 2000. Draft Technical Guidance for Benchmark Dose Modeling (EPA/630/R-00/001). US Environmental Protection Agency, Washington, DC.

United States Environmental Protection Agency (US EPA). 2002. Toxicological review of benzene (non-cancer effects). CASRN 71-43-2.

United States Environmental Protection Agency (US EPA). 2009. Regional Screening Levels for Chemical Contaminants at Superfund Sites. Available at http://www.epa.gov/reg3hwmd/risk/human/rbconcentration_table/index.htm.

United States Environmental Protection Agency (US EPA). 2011a. Recommended use of body weight$^{3/4}$ as the default method in derivation of the oral reference dose. Risk Assessment Forum.

United States Environmental Protection Agency (US EPA). 2011b. 2011 Edition of the Drinking Water Standards and Health Advisories (EPA 820-R-11-002). Office of Water, U.S. Environmental Protection Agency Washington, DC.

United States Environmental Protection Agency (US EPA). 2012a. Provisional Peer-Reviewed Toxicity Values for Sulfolane. CASRN 126-33-0. Superfund Health Risk Technical Support Center, National Center for Environmental Assessment, Office of Research and Development.

U.S. Environmental Protection Agency (US EPA). 2012b. Benchmark Dose Technical Guidance. EPA/100/R-12/001. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC. Available at: http://www.epa.gov/raf/publications/pdfs/benchmark_dose guidance.pdf

Wignall,J, et al. (2014) Standardizing Benchmark Dose Calculations to Improve Science-Based Decisions in Human Health Assessments. Env Health Persp 122(5): 499-505.

Zhu Z, Sun M, Li Z, et al. 1987. An investigation of maximum allowable concentration of sulfolane in surface water. J West China Univ Med Sci 18(4):376-80.