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 ^{¾}) 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.