Regulatory Toxicology and Pharmacology 81 (2016) 512–521

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

The need for transparency and reproducibility in documenting values for regulatory decision making and evaluating causality: The example of formaldehyde

Cynthia Van Landingham [a], Kenneth A. Mundt [b], Bruce C. Allen [c], P. Robinan Gentry [a, *]

[a] Ramboll Environ US Corporation, 3001 Armand St., Suite I, Monroe, LA 71201, United States
[b] Ramboll Environ US Corporation, 28 Amity St., Suite 2A, Amherst, MA 01002, United States
[c] Independent Consultant, 101 Corbin Hill Circle, Chapel Hill, NC 27514, United States

ARTICLE INFO

Article history:
Received 15 July 2016
Received in revised form
17 October 2016
Accepted 18 October 2016
Available online 19 October 2016

Keywords:
Reproducibility
Dose-response analysis
Formaldehyde
Inhalation unit risk estimates
Quantitative methods
Epidemiology

ABSTRACT

Reproducibility and transparency in scientific reporting is paramount to advancing science and providing the foundation required for sound regulation. Recent examples demonstrate that pivotal scientific findings cannot be replicated, due to poor documentation or methodological bias, sparking debate across scientific and regulatory communities. However, there is general agreement that improvements in communicating and documenting research and risk assessment methods are needed. In the case of formaldehyde, the peer-review conducted by a National Academy of Sciences (NAS) Committee questioned the approaches used by the Integrated Risk Information System (IRIS) in developing draft unit risk values. Using the original data from the key study (Beane Freeman et al., 2009) and documentation provided in the draft IRIS profile, we attempted to duplicate the reported inhalation unit risk values and address the NAS Committee's questions regarding application of the appropriate dose-response model. Overall, documentation of the methods lacked sufficient detail to allow for replication of the unit risk estimates, specifically for Hodgkin lymphoma and leukemias, the key systemic endpoints selected by IRIS. The lack of apparent exposure-response relationships for selected endpoints raises the question whether quantitative analyses are appropriate for these endpoints, and if so, how results are to be interpreted.

1. Introduction

Reproducibility and transparency in scientific research and reporting, both in the published literature and in documentation of decisions related to public health reached by authoritative bodies, have received significant discussion and debate (Bustin and Nolan, 2015; Campbell, 2014; Igbal et al., 2016; Jilka, 2016). The National Institutes of Health (NIH) are exploring ways to provide greater transparency of the data that are the basis for published manuscripts (Collins and Tabak, 2014) and have noted that the greater scientific community must take steps to correct this issue. In addition, recent commentaries and surveys highlight the growing lack of reproducibility in scientific research (Anonymous, 2016). One of the most immediate and impactful consequences for a lack of transparency or reproducibility is in the direct reliance on published but un-replicated scientific findings for human health risk assessment, including the derivation of cancer unit risk estimates.

In 2011, the National Research Council (NRC) of the National Academy of Sciences (NAS) convened a Committee to Review USEPA's Draft of the *Toxicological Review of Formaldehyde — Inhalation Assessment* in support of the Integrated Risk Information System (IRIS) (NRC, 2011). The Committee noted:

> *"Problems with clarity and transparency of the methods appear to be a repeating theme over the years, even though the documents appear to have grown considerably in length"*

A further review of the IRIS process in 2014 (NRC, 2014) noted progress in meeting the NRC (2011) recommendations, but further noted:

* Corresponding author.
E-mail address: rgentry@ramboll.com (P.R. Gentry).

http://dx.doi.org/10.1016/j.yrtph.2016.10.011
0273-2300/

"However, NRC committees have conducted several reviews of some of the more complex and challenging IRIS assessments in the last decade and have identified methodologic problems and pointed out deficiencies in EPA's approaches."

Formaldehyde provides one such complex database that introduces significant challenges for consideration in a standard IRIS assessment. It is an endogenously generated compound and, for selected endpoints, multiple studies provide inconsistent results, a few of which have suggested associations with formaldehyde exposure. Some have interpreted these findings (generally at face value and apart from the larger body of results) as reflecting causal associations. As an example, there has been much scientific debate regarding whether there is a causal association between formaldehyde exposure and selected lymphohematopoietic (LHP) endpoints, especially acute myeloid leukemia. Multiple authoritative bodies (IARC, 2012; NTP, 2014) have made hazard classification decisions (sufficient evidence in humans, known to be a human carcinogen) based on conclusions that the available evidence is sufficient to conclude that there is a causal association. For the LHP cancers, these conclusions have been based on the grouping of different types of cancers from a limited number of epidemiological studies (Zhang et al., 2009; Beane Freeman et al., 2009), with little or no consideration of findings reported in many other studies or the animal or mechanistic information, much of which lends no support for or even contradicts these conclusions. It is important to note that in reviewing the same critical studies for formaldehyde as IARC (2012) and NTP (2014), the European Chemicals Agency (ECHA, 2011) concluded that

"Altogether, in absence of convincing evidence for a biologically plausible mechanism and considering the discrepancy of results in epidemiological studies, a causal relationship between formaldehyde exposure and induction of myeloid leukaemia cannot be concluded."

The 2010 draft IRIS *Toxicological Review of Formaldehyde — Inhalation Assessment* provided the first quantitative estimates of a dose-response relationship between two lymphohematopoietic endpoints, Hodgkin lymphoma (HL) and all leukemias (combined category), and exposure to formaldehyde based on the results from a single epidemiological study (Beane Freeman et al., 2009). The use of these two endpoints by USEPA (2010) for the estimation of unit risk factors was based on the conclusion that the weight of the epidemiologic evidence supported a link between formaldehyde exposure and LHP cancers, particularly myeloid leukemias. In addition to HL largely being considered unrelated to environmental exposures, no other key epidemiological study demonstrates such an association, raising questions as to the validity of the finding in Beane Freeman et al. (2009). As for the combination of all leukemias, little scientific basis is provided for aggregating what increasingly are understood to be diverse diseases with different etiologies, prognoses and treatments.

In 2011, the NRC Committee review noted many uncertainties in the approach used by USEPA (2010) to estimate risk values. The Committee recognized that USEPA (2010) had relied upon selected associations reported between formaldehyde and various LHP cancers from a single study (Beane Freeman et al., 2009). The NRC (2011) Committee further recommended that USEPA conduct an independent analysis of the dose-response models to confirm the degree to which the models fit the data appropriately, as well as consider the use of alternative extrapolation models for the analysis of the cancer data. The NRC (2011) Committee concluded that this is especially important, given the use of a single study, the inconsistencies in the exposure measures, and the uncertainties associated with the selected cancers. In addition to the impact of these assumptions, the NRC (2011) Committee noted that while the National Cancer Institute (NCI) cohort studies, including Beane Freeman et al. (2009), may be the only studies with sufficient exposure and dose-response data needed for risk estimation, they are not without weaknesses and these need to be considered. This recommendation from the NRC (2011) Committee raised several challenges. While there is some guidance provided for the use of animal data for dose-response modelling (USEPA, 2012), the use of epidemiological data in the estimation of inhalation unit risk (IUR) estimates does not have guidance that provides a "road map" for conducting these types of assessments. When using epidemiological data for the estimation of unit risk values, more extensive documentation in the IRIS profile is needed to be able to clearly understand the data relied upon and the methods applied.

In a separate study (Checkoway et al., 2015), the raw data from the NCI cohort study (Beane Freeman et al., 2009) were obtained through a Technology Transfer Agreement (TTA) with the objective of replicating the findings reported by Beane Freeman et al. (2009), as well as conducting additional analyses not reported by Beane Freeman, specifically, acute myeloid leukemia (AML). The availability of these data provided an opportunity to attempt to replicate the unit risk estimates derived by USEPA (2010), as well as address some of the questions raised by NRC (2011). In addition, it offered the opportunity to conduct alternate independent analyses to evaluate specific leukemias, rather than all leukemias combined, and the impact of alternate dose-response models on the estimates of inhalation unit risk. The methods and results of the attempt to duplicate the USEPA (2010) unit risk values, as well as conduct alternate and independent analyses to address the questions raised by NRC (2011) are reported here.

2. Methods

2.1. Duplication of USEPA (2010) reported unit risks

Our goal was to follow the same process and methods used by USEPA (2010) in the estimation of unit risk factors for the two LHP cancers (Hodgkin Lymphoma and all leukemias (combined category)). However, as noted by NRC (2011), the documentation provided in USEPA (2010) related to the assumptions and processes used in the estimation of the unit risk values was limited. NRC (2011) has outlined five steps that it appears USEPA (2010) used in the estimation of formaldehyde unit risks:

1. Evaluate the association between formaldehyde exposure and LHP endpoints;
2. Convert the relative risk estimates into lifetime risk for the exposed population;
3. Compute lifetime risks for Hodgkin Lymphoma and/or all leukemia for the unexposed population;
4. Determine the maximum likelihood and lower bound estimates of the point of departure; and
5. Estimate inhalation unit risks.

Using these five steps, we attempted to duplicate the USEPA (2010) reported unit risks for Hodgkin lymphoma and "all leukemias" using the raw data from the Beane Freeman et al. (2009) study. In order to conduct this estimate, the following were needed:

- *An estimate of cumulative dose for each individual in the cohort.* This information was not provided in either USEPA (2010) or Beane Freeman et al. (2009) and must be determined from the raw data.