

## Beyond Science and Decisions: From Problem Formulation to Dose-Response Assessment: Summary of Case Study: Criteria Requirements for Data-Driven Carcinogenicity Mode of Action (MoA) Determinations as Exemplified by Chloroform Carcinogenicity

This case study provides a process and criteria for determining when a MoA has been sufficiently well established in an animal model that it should be used to determine the human relevance of the tumors, and if relevant, replace policy-based linear, non-threshold, default models for extrapolating high dose animal tumor data to human cancer risks with models that use actual chemical-specific data.

### 1. Summary of the Method Illustrated by the Case Study

All critical reviews and consensus panel reports regarding chloroform's carcinogenic MoA were evaluated according to a subset of components from a recently published, hypothesis-driven weight of evidence framework.<sup>1</sup> Those components provided a consistent metric for evaluation that is independent of the frameworks and criteria used by the reviews and consensus panels to assess chloroform's MoA. Thus, for each key event in chloroform's carcinogenic MoA, the reviews and consensus reports were evaluated for the following components (also listed in Table 2 of the Case Study):

- a) Support for the particular key event (hypothesis) in the MoA;
- b) Evaluation or discussion of data quality supporting (or refuting) the key event;
- c) Evaluation or discussion of counterfactual<sup>2</sup> concepts in experimental design and interpretation for data supporting the key event, and;
- d) Evaluation or discussion of alternative hypotheses or data interpretations regarding the key event.

From the results of this analysis for chloroform, criteria were derived that should allow data-driven carcinogenicity risk determinations in place of default carcinogenicity assumptions

---

1 Borgert, C.J., Mihaich, E.M., Ortego, L.S., Bentley, K.S., Holmes, C.M., Levine, S.L., and Becker, R.A. (2011). Hypothesis-driven weight of evidence framework for evaluating data within the US EPA's Endocrine Disruptor Screening Program. *Regul Toxicol Pharmacol* 61, 185-191.

<sup>2</sup> Counterfactuals test whether the effect of interest still occurs when a putative causal step is prevented under conditions that would otherwise produce the effect of interest.

whenever the data for a specific chemical are similarly compelling. The set of criteria derived were (also listed in Table 4 of the Case Study):

- I. Defined key events should be consistently (not necessarily unanimously) supported among objective analyses;
- II. Issues regarding data quality should not weaken support for the key events; i.e., data supporting the proposed key event should be of equal or higher quality than contradictory data;
- III. At least one of the key events should be counterfactually demonstrated to exhibit a biological threshold;
- IV. The MoA should not differ by route of exposure (if a chemical's MoA does differ by route of exposure, the MoA evaluation should reflect those differences);
- V. Alternative MoA hypotheses that would dictate a linear biological model of tumor development should be consistently ruled out or considered unlikely among objective analyses.

## 2. Problem Addressed by the Method

Updated assessments continue to strengthen non-linear MoA-based approaches for assessing carcinogenesis and biologically based models continue to refine extrapolations to human cancer risks. Nonetheless, clear criteria have yet to be defined for determining when departure from a policy-based default model is justified and should be used so that the policy is consistent with the best available science. As a consequence, policy decisions regarding use of MoA-based approaches tend to rely on subjective, case-by-case judgments about what constitutes sufficiency of the data for any particular chemical. Clear criteria would enhance scientific integrity and transparency in cancer risk assessments and would facilitate consistency in policy-making. This case study is an initial attempt to define a process and criteria for making this determination, using a data-rich chemical with a well-defined MoA: chloroform. Chloroform is a drinking water disinfection byproduct that is widely found in drinking water systems. Alternative disinfection processes that generate less chloroform are available, but they generate other byproducts, some of which may be of greater concern than chloroform.

## 3. General Applicability of the Method

Within its intended context, the method is believed to be broadly applicable to two specific questions: i) whether a MoA for any particular chemical has been sufficiently established that it should be considered the correct model for assessing human cancer risks from exposure to that chemical, as a matter of science and policy, and ii) whether there is sufficient confidence in the

MoA that a non-linear extrapolation approach should be used to assess human cancer risks, as a matter of science and policy. The criteria developed here apply to the first part of the Human Relevance Framework, which is a determination of the MoA in an animal model. They are formulated within the realm of experimental data and may also be appropriate for the second arm of the HRF, which determines whether a MoA in the test species is likely to be relevant for human tumors. Consideration of this second question might also benefit from clear criteria, but was not specifically addressed in formulating the current methodology and case study.

Although the criteria do not assume that the MoA is relevant to humans, their use could be helpful in resolving that question because it would help to ensure firm and consistent decisions about when a MoA has been sufficiently well established in an animal model that it should form the basis for the human relevance determination. Thus, use of clear criteria for deciding the sufficiency of the animal MoA should facilitate the decision to adopt a dose-response model for humans based on that particular MoA.

#### 4. Overall Strengths and Weaknesses of the Method

The strengths of the method derive from the simplicity of the criteria and the fact that they relate to fundamental tenets of the scientific method as described in a) through d) above.

By design, the criteria are compatible with the considerations used to identify key events in the HRF and the IPCS framework, yet they are not identical. These are both strengths. That they are compatible lends credibility and ready utility; that they are not identical avoids a potential conundrum similar to that which arises in attempting to validate a model with the same data used to derive the model. Were the HRF or IPCS method used to derive these criteria, they would not represent a separate and independent synthesis of potentially differing interpretations. Furthermore, because the criteria are not identical to those used in the HRF to identify key events, they do not bias an evaluation toward the outcomes of the HRF process. In other words, they provide a fair hearing for evaluations conducted by alternative decision structures, which is considered a strength.

The criteria can leverage other scientific processes, such as the HRF and IPCS framework and published critical reviews, rather than demanding a de novo consideration of all primary data. In this regard, the method lends efficiency and strength to the decision-making process.

The criteria were developed using chloroform as the test case. The carcinogenic MoA for few, if any, chemicals are as well characterized as that of chloroform, and none has been subjected to greater scrutiny or more numerous peer-review evaluations. Therefore, there can be high confidence that the criteria are based on reliable scientific evidence.

The weaknesses of the method arise particularly when the method is applied to leverage other evaluative processes as mentioned above. In this instance, if the existing published work were unanimously incorrect regarding one or more of the criteria considerations, the method would not uncover that error.

Because the method was developed based on evaluations of chloroform, which are particularly numerous, thorough, and rigorous, it may be unnecessarily stringent. This potential weakness could be remedied by testing the method against additional case studies that vary in the depth and breadth of published evaluations, and modifying the criteria as necessary. In this regard, the method should reject datasets for chemicals whose MoAs are clearly uncertain, and should accept others with strong datasets, albeit perhaps less strong than for chloroform.

Although the criteria were intentionally developed with broad application in mind, their development solely from the chloroform literature could render them less practical for carcinogenic MoAs other than the cytotoxic, non-linear type. This potential weakness is considered unlikely, but cannot be dismissed until further evaluation resolves the issue.

#### 5. Minimum Data Requirements and Types of Data Needed to Apply the Method

Although the method is applied here to critical reviews and consensus panel evaluations regarding carcinogenic MoAs, it is applicable to evaluations of primary data as well. In the latter case, additional WoE components regarding primary, secondary and tertiary validity (Borgert et al. 2011) would be applied. In either circumstance, considerable MoA data are required, but the minimum primary data requirements are uncertain as these would vary by individual MoA and perhaps by chemical. However, this ambiguity does not apply to the method directly. For application of the method, a minimum of one comprehensive MoA evaluation or data set is required, providing the data are available for review by independent analysts, e.g., published in a peer-reviewed journal, the official report of an Agency or Institutional panel evaluation, etc. Conceivably, a single comprehensive data set or evaluation could be sufficiently thorough and rigorous to satisfy all five criteria. In practice, however, application of the method will be strengthened with greater numbers of well-constructed studies demonstrating each key event, or of critical reviews and consensus evaluations, as is evident for the case study example, chloroform.