REPLY TO LETTER TO THE EDITOR

Hormesis and High Throughput Studies: Crump’s Analysis Lacks Credibility

Edward J. Calabrese,* 1 John W. Staudenmayer,† Edward J. Stanek, III,‡ and George R. Hoffmann§

*Department of Public Health, Environmental Health Sciences Program; †Department of Mathematics and Statistics; ‡Department of Public Health, Biostatistics and Epidemiology Program, School of Public Health and Health Sciences, University of Massachusetts, Amherst, MA 01003; and §Department of Biology, College of the Holy Cross, Worcester, MA 01610-2395

Received March 15, 2007; accepted April 2, 2007

We recently reported results of an analysis of a large National Cancer Institute (NCI) database on effects of chemicals in yeast, concluding that dose–response relationships in the low-dose range are more compatible with a hormetic dose–response model than a threshold model (Calabrese et al., 2006). Crump suggests that our interpretation may be incorrect based on differences in understandings of how response ratios were calculated. Since the original data were not available due to a computer malfunction at the Fred Hutchinson Cancer Research Center (FHCRC), we described the methods used to construct the NCI database as reported by Dr Simon, the Principal Investigator. A response was formed by averaging results from two optical density ratios, each of which was formed by dividing the response in a well by the mean of eight control values on the plate.

Crump stated that the unavailability of the original data makes it impossible to know with certainty how the response ratios were calculated. He then proposed an alternative calculation method and suggested that the response ratios may have been calculated in this way. Crump’s method divides each of eight control-well values separately into each single treatment-well value, and the response is the mean of those eight ratios. Using his hypothesized averaging method and simulated data, “hormetic” dose responses occurred at a higher than random frequency, even if no hormesis was present. These apparent hormetic effects were considered false positive responses. He suggested that his methodology was “not implausible” and that our findings should not be considered reliable since one could not be certain about which calculation method was used.

The idea underlying Crump’s results relates to variability in control response. With higher control variability, there will be a larger expected number of false positive “hormetic” responses. Forming ratios by individual control values instead of an average control value increases control variability, and hence increases the bias. Crump’s hypothetical method created variability in controls that was eight times larger than the variability using the method reported by the study investigators. Understandably, the simulation exercise produced a large bias.

The critical question is whether the Crump method or that reported by Simon and the NCI was used to produce the NCI data. We have ample evidence demonstrating our due diligence and accuracy with respect to the calculation method. There is strong evidence that the hypothetical method proposed by Crump was not used. Finally, Crump’s claim that his method of constructing responses is “not implausible” will be shown to lack objective credibility.

In February 2005, Dr Julian Simon of the FHCRC was contacted by Edward J. Calabrese about our intention to analyze the NCI yeast database. We learned that the original data had been lost and were irretrievable. Since the calculated response ratios were available, we thought it possible that we could still proceed. However, we took special care to iteratively interview Dr Simon on all aspects of the study methodology, including how the ratios were calculated. We reinterviewed him on the same points to ensure a clear understanding. These interviews were filed as a detailed written record that we shared with Dr Simon to ensure his concurrence. We also contacted the NCI, the collaborative funding source for the yeast screening data of Dr Simon, to obtain an independent confirmation of the methodology by which response ratios in the NCI testing program were calculated. This interview with the NCI (June 2005) revealed a response ratio calculation method identical to that reported by Dr Simon. It was kept as a written record. After we were convinced that we understood the details of the yeast assay and calculation methods, and the calculation methodology was confirmed by Dr Simon’s sponsors, we proceeded with our analysis. Based on this record, we consider Crump’s assumptions to be not plausible.

1 To whom correspondence should be addressed at Department of Public Health, Environmental Health Sciences Program, Morrill I, N344, University of Massachusetts, Amherst, MA 01003. Fax: (413) 545-4692. E-mail: edwardc@schoolph.umass.edu.

© The Author 2007. Published by Oxford University Press on behalf of the Society of Toxicology. All rights reserved.
For Permissions, please email: journals.permissions@oxfordjournals.org
Even though the original NCI data were lost (a fact that we noted in our paper), there is no reasonable basis for doubting that the methods we described were applied to construct the NCI data. All the evidence contradicts Crump’s assertion that his alternative method may have been used.

Subsequent to Crump’s analysis, we recontacted Dr Simon and revisited the past conversations and our written records, and he confirmed verbally and in writing that the original information provided was accurate. We also contacted the NCI and received written reaffirmation that the NCI has long used the averaging method described by Simon and has never used the one proposed by Crump.

While the absence of original data places limitations on what can be explored and evaluated, the lack of original data may not be uncommon in database analyses. There are 3-year limits on the time for which federally (National Institutes of Health) funded investigators are obligated to retain data. The fact that original data may not be available has not provided a general foundation for investigators to claim that they cannot trust the descriptions of the original investigators about how they obtained and analyzed their data. In the final analysis, believability and operational acceptance are based on the credibility of the investigators, including their experience, record of achievement, and the capacity of the system to provide the data. In the current situation, the NCI yeast data research group at the FHCRC would receive high marks on all such accounts.

The assurances that we received from Dr Simon that we accurately reported the procedures used in the research removed any doubt about how the studies were conducted. Even though the nature of the analysis of the NCI database was not in question, we explored how 96-well plates in high-throughput studies are typically analyzed. If Crump’s assertion of “not implausible” is correct, then one would expect the analysis that he hypothesized to occur with some reasonable frequency in laboratory practice and the published literature. We briefly outline several approaches that we used to evaluate the credibility Crump’s hypothesis:

- We surveyed approximately 200 published studies on cell proliferation in tumor cell lines using 96-well plates. The papers, which had been obtained previously, were chosen based on the strength of the study design and capacity for dose response evaluation (e.g., number and spacing of doses) with no consideration given to how the data were analyzed (e.g., how response ratios were calculated).
- We contacted 55 scientists who published papers using 96-well plates within the past 7 years. They were identified using the Web of Science and Science Direct databases, using the keywords 96-well plate, or word combinations such as 96-well plate and cell proliferation or other terms.
- We contacted 20 professors of biostatistics from Schools of Public Health and other similar institutions.

The results of these complementary efforts were completely consistent, indicating no instance of support for Crump’s hypothesis. A review of the published studies revealed that the method reported by Simon was extremely common, and we identified no occurrence of the hypothetical method proposed by Crump. Our survey of scientists resulted in responses from 40 laboratories (including major pharmaceutical companies, federal agencies, EU research facilities, university laboratories, commercial high-throughput facilities, and private foundation research facilities). It revealed that most labs routinely used the NCI/Simon method. No research team reported using Crump’s analysis method. All groups were asked to cite, if possible, an example where this alternative method had been published; most respondents indicated that they had never seen it performed, and no one could identify a case when it had been used. Written statements from scientists who review testing protocols for the chemical industry and the EU indicate that they have never seen Crump’s method attempted or employed. Thus, all responses were consistent with the statements from the NCI and Dr Simon as noted above.

We consider these results important in providing an assessment of the plausibility of Crump’s approach. The “reach” of this large number of independent researchers is broad, encompassing substantial research experience in diverse settings, varied positions, and service as peer reviewers for journals and governmental research groups. That so many researchers are unaware of an example where Crump’s hypothesized method actually was used in their field argues against its plausibility.

We received information of a generally speculative nature from 15 biostatisticians suggesting settings in which Crump’s hypothetical method might be used. None had ever seen it used in practice, and none could cite any published paper that had employed it, despite extensive association with laboratory scientists performing assays related to growth and cell proliferation.

In conclusion, diverse lines of evidence indicate that the method of analysis hypothesized by Crump clearly does not apply to our study (Calabrese et al., 2006), and it seems implausible more generally. It lacks a reasonable foundation in the literature or in practice. In contrast, all the evidence is consistent with the procedures described by Dr Simon and the NCI and with the interpretation given in our manuscript. We conclude that there is no logical basis for qualifying the conclusions that we reached (Calabrese et al., 2006).

REFERENCE