


Revisiting Cohen et al. 2015, Cohen et al. 2014 and Waalkes et al. 2014: a bayesian re-analysis of tumor incidences

Ingrid L. Druwe¹ · Lyle Burgoon² 

Received: 17 February 2016 / Accepted: 8 June 2016 / Published online: 20 June 2016
© Springer-Verlag Berlin Heidelberg (outside the USA) 2016

Over the past year, this journal has published a lively discussion in the form of letters to the editor between Waalkes et al. and Cohen et al. regarding an article published in this journal titled “Lung Tumors in Mice induced by ‘whole Life’ inorganic arsenic exposure at human relevant doses” by Waalkes et al. in 2014. Cohen et al. raised a series of thoughtful questions with respect to the reproducibility of the control animal tumor incidences in the arsenic exposure studies published by Tokar et al. (2011) and Waalkes et al. (2014). In addition, Cohen et al. brought into question whether the development of lung tumors in the mice was related to the genetic background of the mice used in the study rather than arsenic exposure. Many of the questions raised by Cohen and colleagues centered around what they deemed to be uncertainty in the tumor incidences in the control animals used in both studies, and by extension, the quality of the studies performed by Tokar and Waalkes.

If the assertions made by Cohen et al. proved true, risk assessors would be unable to use Waalkes et al.’s (2014) study in hazard and dose–response assessments of inorganic arsenic. Thus, we performed the analysis requested

by Cohen et al. Specifically, we tested the null hypothesis: *The control animal tumor incidences reported by Tokar et al. in 2011 are not different from the study reported by the same group in the Waalkes et al. 2014 publication.* Our full analysis can be found in Burgoon and Druwe (2015); however, we briefly discuss the approach and results here.

To test the null hypothesis, we used a Bayesian approach. Initially, we examined whether or not a difference existed without using any prior knowledge of what the tumor incidence in Waalkes’ laboratory was; therefore, we used a flat prior distribution. We modeled the control tumor incidences from the Tokar 2011 and Waalkes studies as Bernoulli distributions. The posterior distributions were calculated and are shown in Burgoon and Druwe (2015). There is a difference of about 13 % between the means of these two distributions.

However, that fact alone does not mean that the incidences are from different distributions. In fact, what we could be observing is a case where the incidences for each study were taken from different sides of the same distribution. Thus, to test the hypothesis, we took samples from the posterior distributions and calculated the difference to obtain a distribution of the differences. If the incidences in both studies were from the same distribution, we would expect a difference of 0, or close to 0, to be a credible value. In order to accomplish this, we used an approach that sets a region of practical equivalence (ROPE) around the zero difference, and the 95 % highest density interval (HDI) of the difference distribution. The ROPE demarcates a region around zero difference that is functionally equivalent to no difference. In general, if any part of the 95 % HDI is within the ROPE, then we accept the null hypothesis that the control tumor incidences from the studies are the same. Else, we reject the null hypothesis. A complete explanation of our decision rules can be found in Burgoon and Druwe (2015).

Disclaimer: The views presented in this article do not necessarily reflect current or future opinion or policy of the US Environmental Protection Agency and the US Army Corps of Engineers.

✉ Lyle Burgoon
lyle.d.burgoon@usace.army.mil

¹ Oak Ridge Institute for Science and Education, US Environmental Protection Agency, National Center for Environmental Assessment, Vicksburg, MS, USA

² US Army Engineer Research and Development Center, Environmental Laboratory, Research Triangle Park, NC, USA

We decided to take the ROPE and 95 % HDI approach to test for a practical equivalence (or more colloquially known as a zero difference test, similar to a *t* test approach) because we wanted to be able to leverage existing prior knowledge, and to calculate posterior probabilities and posterior odds. We did not want to be hamstrung with the notion of *p* values and instead wanted to be able to generate distributions that represented uncertainty and would also represent a notion of actual probability. Our chief concern with *p* values is that they do not represent biological significance. The ROPE and 95 % HDI approach, on the other hand, allows us to use our biological knowledge to define a response range that we define as being equivalent to 0.

As depicted in Burgoon and Druwe (2015), we can see that a difference of 0 is a credible value, and that the 95 % highest density interval (the values that are greater than the first 5 % of the distribution) includes a region of practical equivalence (ROPE) of 0 ± 5 %. In other words, we accepted the null hypothesis based on our analysis; thus, the tumor incidences in both studies are likely from the same distribution based on the data we have.

After completing that analysis, we identified another study by Tokar et al. (2012) that was conducted prior to the Waalkes study. The control animals in this study appear to be exchangeable with the Waalkes control animals. Thus, we used the Tokar et al. (2012) study as a prior, used the Tokar et al. (2011) study as the likelihood, and calculated the posterior distribution of the control animal tumor incidences for both Tokar et al. studies. We then compared this posterior to the posterior for the Waalkes data. Again, we found that part of the 95 % HDI was within the ROPE—recapitulating the results we found before.

Next we tested a second null hypothesis: *Inorganic arsenic exposure does not increase the lung tumor incidence in mice when taking the Tokar et al. (2011) control tumor incidence into account*. Again, we used a Bayesian approach. The advantage of a Bayesian approach is that it allows us to combine prior knowledge with information from the current study. In other words, the Bayesian approach, unlike a frequentist approach, allows us to directly combine the control tumor incidences from the three studies to directly assess Cohen et al. assertion. For this analysis, we use the same approach as before, calculating the posterior distributions and then the difference between the posterior of the inorganic arsenic-treated groups and the posterior of the combined controls. We used the control data from both Tokar et al. studies as our prior.

The posterior distribution of the differences for each of the inorganic arsenic treatment groups is shown in our report (Burgoon and Druwe 2015). Our results agree with Waalkes et al. (2014): The 50- and 500-ppb exposure groups show an increase in lung tumor incidence, while the 5000-ppb group shows the same incidence as the vehicle. The posterior odds ratios for the increase in mouse lung tumor incidence for the 50-, 500-, and 5000-ppb exposure groups were 3.1:1, 3.5:1, and 1.2:1.

In conclusion, we found that Cohen et al.'s assertion with respect to the differences in the control mouse lung tumor incidences and the assertion following from there that the differences in the 50- and 500-ppb exposed groups were anomalies were false given the data presented by Tokar et al. (2011, 2012) and Waalkes et al. (2014). Based on our analysis, the data support the assertion that the difference in incidence between the control groups in Tokar et al. (2011) and Waalkes et al. (2014), as well as Tokar et al. (2011, 2012) and Waalkes et al. (2014), is likely due to sampling. Furthermore, upon taking the control tumor incidence in Tokar et al. (2011) into consideration, integrating it with the control tumor incidence in Waalkes et al. (2014) we were able to affirm that the data support the assertion that inorganic arsenic increases the tumor incidence in the 50- and 500-ppb exposed groups, but not the 5000-ppb group, with odds ratios of 3.1:1, 3.5:1, and 1.2:1, respectively (Burgoon and Druwe 2015).

Thus, based on the published experimental data, we confirm that Waalkes et al. (2014) study demonstrates low-dose inorganic arsenic exposures increase the lung tumor incidences in CD1 male mice.

References

- Burgoon and Druwe (2015) https://github.com/DataSciBurgoon/arsenic_mouse_lung_tumor_reanalysis
- Tokar EJ, Bhalchnadra AD, Ward JM et al (2011) Carcinogenic Effects of “Whole-Life” exposure to inorganic arsenic in CD1 mice. *Toxicol Sci* 119(1):73–83
- Tokar EJ, Bhalchnadra AD, Waalkes MP (2012) Renal, hepatic, pulmonary and adrenal tumors induced by prenatal inorganic arsenic followed by dimethylarsinic acid in adulthood in CD1 mice. *Toxicol Lett* 209(2):178–185
- Waalkes MP, Qu W, Tokar EJ et al (2014) Lung tumors in mice induced by “whole life” inorganic arsenic exposure at human relevant doses. *Arch Toxicol* 88(8):1619–1629