Review of Multiple Imputation Methods Applied to Censored Bioassay Datasets

Bob Barton, CHP
Dr. Carl Gogolak

To the Joint Savannah River Site and SEC Issues Work Groups

November 20, 2020

Background

◆ Co-exposure model development is dependent on the interpretation of "censored results" (e.g., less than the detection limit, decision level, or other threshold level)

◆ Co-exposure modeling requires statistical interpretation to fit data to distributions

◆ Some datasets have large portions that are censored (greater than 90%)

SC&A

Multiple imputation

- NIOSH developed methods to impute (or infer) censored data based on the positive results in the dataset

- Documented in ORAUT-RPRT-0096, "Multiple Imputation Applied to Bioassay Coworker Models" (NIOSH, 2019a)

- First application was in the development of co-exposure models for SRS (NIOSH, 2019b)

SC&A

Initial SC&A concerns

◆ SC&A reviewed the SRS co-exposure models and had 1 finding and 1 observation related directly to general multiple imputation methods (SC&A, 2020a)

◆ Finding 2:

– Imputation analysis resulted in estimates of co-exposure bioassay results that were often much less than ½ the minimum detectable activity (MDA)

– "Missed dose approaches" assume ½ the MDA (finding 2)

◆ Observation 1:

– Method is mathematically correct but has the potential to bias results low

– Maximum possible mean method may be preferable

SC&A

Initial discussions and followup

◆ December 5, 2019: Finding and observation discussed during joint SRS-SEC Issues work group meeting

◆ SC&A tasked with performing a broader technical evaluation of the multiple imputation methodology

◆ June 2020: SC&A delivered technical memorandum, "Review of Multiple Imputation Methods Applied to Censored Bioassay Datasets" (SC&A, 2020b)

SC&A

Summary of relevant literature (Helsel, 2009)

◆ Helsel (2009), "Much Ado about Next to Nothing: Incorporating Nondetects in Science"

  – "In general, do not use substitution. . . . Substitution is NOT imputation, which implies using a model such as the relationship with a correlated variable to impute (estimate) values" (p. 261)

  – "Method evaluations for estimating a mean do not necessarily carry over to the more difficult issues of how to compute interval estimates, upper percentiles, a correlation coefficient, a regression slope and intercept" (p. 261)

SC&A

Summary of relevant literature (Helsel, 2020)

- ◆ Helsel (2020), "Why Not Substitute ½ DL for Nondetects?"
  - Creates problems of "invasive data" (e.g., artificial lowering of the standard deviation)
  - May create artificial trends in the data that do not actually exist
  - Especially problematic for datasets with multiple censoring levels (e.g., bioassay methods where the limit of detection improved over time)

SC&A

SC&A general technical comments

◆ Multiple imputation uses information in the "detected data" to generate values below the detection limit

◆ Co-exposure modeling generally assumes a common lognormal distribution (detected and nondetected data come from the same distribution)

◆ Therefore, imputation uses more of the available information from the dataset than substitution (imputation is statistically preferable)

SC&A

SC&A general technical question

◆ When is the proportion of censored data too much?

◆ Krishnamoorthy et al. (2009) suggest that the performance of imputation is dependent more on the total number of censored results than on the proportion

  – SC&A does not recommend a universal upper limit on the percentage of censored results

  – Each dataset should be evaluated individually with emphasis on total number of available uncensored results

  – Example: 1 out of 10 uncensored results may be problematic, where 10 out of 100 uncensored results may be OK

SC&A

Related topic for further discussion

◆ Several example cases in ORAUT-RPRT-0096 indicate a mix of positive and negative values (i.e., some values do not have a log)

◆ Aitchison and Brown (1957) discuss an alternate distribution, called the delta distribution

 – Mixture of lognormal and a discrete probability at zero

 – Gogolak (1986) describes methods for estimating the three key statistical parameters (delta, mu, and sigma)

◆ NIOSH may want to consider the delta distribution for imputation methods when a large proportion of unexposed workers are mixed with a much smaller proportion of exposed workers

SC&A

Practical implications for dose reconstruction

◆ SC&A (2020a) provided a scoping evaluation comparing imputation co-exposure values that were much less than ½ the MDA to standard missed dose approaches (evaluated at ½ the MDA)

  – Evaluated Sr-90, Co-60, Np-237, Pu-239, and U-234

  – Intakes and doses are significantly higher for missed dose methods, but only minor differences found in resulting probability of causation (POC) (observation 2)

  – Specific to uranium, imputed co-exposure values were a factor of 4 or more higher than the missed dose approach (observation 3)

◆ Two methods showed remarkable agreement and had limited effect on POC (exception was uranium)

◆ Significant effect of uncertainty on POC when applying co-exposure assignment

◆ Unmonitored radiation workers who should have been monitored can (and in SC&A's opinion should) be assigned the 95th percentile co-exposure values, which are likely reflective of detected values rather than imputed values

SC&A

Summary of SC&A technical evaluation of multiple imputation

◆ Multiple imputation is a mathematically accurate method for assessing censored bioassay data in the absence of other information (e.g., raw data measurements)

◆ Total number (rather than the percentage) of uncensored results should be used in evaluation of appropriate statistical methods

◆ NIOSH may consider use of the delta distribution for cases of large proportions of unexposed workers mixed with a much smaller proportion of exposed workers (i.e., when lognormal fits are less ideal)

◆ Substitution, in comparison to imputation, has many analytical drawbacks as described in Helsel (2020), such as artificial lowering of the uncertainty

◆ Scoping calculations indicate very little practical difference in POC values when evaluated at the 50th percentile (effect of applied uncertainty to co-exposure assignment)

◆ Unmonitored workers can be assigned the 95th percentile, which is likely reflective of actual positive monitoring results

SC&A

SC&A Conclusion

SC&A finds that the use of multiple imputation in evaluation of bioassay datasets with censored results is technically appropriate, scientifically defensible, and likely of small practical significance when considering its effect on resulting POC calculations.

SC&A

References

◆ Aitchison, J., & Brown, J. A. C. (1957). *The lognormal distribution*. Cambridge: Cambridge University Press.

◆ Gogolak, C. V. (1986). *Parameter estimation for censored samples from delta-ized distributions with application to air quality monitoring data* [Doctoral dissertation, Polytechnic University of Brooklyn]. Proquest.com Pub. ID 27948318

◆ Helsel, D. R. (2009). Much ado about next to nothing: Incorporating nondetects in science. *Annals of Occupational Hygiene, 54*(3), 257–262.

◆ Helsel, D. R. (2020). Why not substitute ½ DL for nondetects? *PracticalStats.com*. https://www.practicalstats.com/resources/Webinar-pdfs/Why-Not-Sub.pdf

◆ Krishnamoorthy, K., Mallick, A., & Mathew, T. (2009). Model-based imputation approach for data analysis in the presence of non-detects. *Annals of Occupational Hygiene, 53*(3), 249–263.

◆ National Institute for Occupational Safety and Health (NIOSH). (2019a). *Multiple imputation applied to bioassay coworker models* (ORAUT-RPRT-0096, rev. 00). SRDB Ref. ID 175396

◆ National Institute for Occupational Safety and Health (NIOSH). (2019b). *Internal coworker dosimetry data for the Savannah River Site* (ORAUT-OTIB-0081, rev. 04). https://www.cdc.gov/niosh/ocas/pdfs/tibs/or-t81-r4-508.pdf

◆ SC&A, Inc. (2020a). *Review of ORAUT-OTIB-0081, revision 04, "Internal Coworker Dosimetry Data for the Savannah River Site"* (SCA-TR-2019-SEC004, rev. 1). SRDB Ref. ID 181746

◆ SC&A, Inc. (2020b). *Review of multiple imputation methods applied to censored bioassay datasets* [Memorandum to Savannah River Site and SEC Issues Work Groups]. https://ftp.cdc.gov/pub/FOIAREQ/182225-508.pdf

SC&A

Questions?

SC&A