Review of ORAUT-RPRT-0071 on External Dose Coworker Methodology

Richard Griffiths, SC&A

Advisory Board on Radiation and Worker Health, Subcommittee for Procedure Reviews

March 14, 2024

Overview of ORAUT-RPRT-0071

- Describes a multiple imputation (MI) method for filling in censored – less than the limit of detection (LOD) – readings
- Current method: one-half of LOD
- MI fills in censored measurements with multiple replicates
- Uses average as imputed value
  - Combines with uncensored measurements in further analyses
- Procedure described has two components
  - Imputation method: MI
  - Probability model underlying the imputations

Summary of SC&A review of RPRT-0071

- MI justifiable and likely improves on LOD/2
- MI generally regarded as state-of-the-art for imputation
  - Can reduce bias
  - Allows for measurement of estimator uncertainty
- Application of lognormal probability model can be problematic in some situations
  - Lognormal assumption should be validated case-by-case
- SC&A views MI positively but believes there are several topics to be explored further
- Leads to four high-level observations

Observation 1: RPRT-0071 does not include estimates of uncertainty

- Significant benefit of MI is to accurately account for error in estimation
- RPRT-0071 does not capitalize on this benefit
- Could help understand downstream uncertainty
  - in co-exposure model
  - in probability of causation model

Observation 2: Explore mixture models

- Nonpositive measurements come from statistical measurement error
- Applicable to all measurements, not just nonpositive ones
- Mixture models explored in ORAUT-RPRT-0096
- Mixture models could be combined with MI to develop better inferences

SC&A

Observation 3: Determine probability model for each case individually

- RPRT-0071 notes lognormal is not optimal in all situations
- Report focuses only on lognormal
- Misspecification of underlying model will undermine imputations
- Analysts need to be aware of other possibilities
- Guidelines for evaluating each situation individually could be helpful

Observation 4: Account for relationship of doses to covariates

- May be cases where covariate information is more important than underlying statistical distribution
- For example, dosages may relate to occupation
- Could stratify by occupation
- Could include occupation in underlying probability model
  - Lognormal assumption can still be supported in a generalized linear model

SC&A's comments by section of RPRT-0071: section 1.0, "Introduction"

- Dose reconstruction
  - Doses in table 1-1 "were reconstructed to eliminate the censoring"
  - How doses were reconstructed is not explained
- **Observation 5:** NIOSH does not provide adequate information on how doses were reconstructed
- Negative dose measurements
  - Important to think about this type of measurement error
  - We discuss statistical measurement error more fully later

SC&A's comments on RPRT-0071 introduction: Linear imputation model

- NIOSH: "These linearly imputed doses are given in the Impute C column in Table 1-1"
  - Take the x-axis of a graph to be the dates of the measurements
  - Take the y-axis of same graph as imputed measurement for each dose
  - Draw line starting at y = 0 for first date to y = 0.05 (LOD) for last date
  - Impute the value of y for the measurement for each date on the x-axis
  - Amounts to $y = 0.05 \times t$, where *t* indexes date
- We think model is meant to illustrate *one* of the imputation methods
- SC&A worries someone might think this is a legitimate model
- **Observation 6:** Report would benefit from a disclaimer about the linear imputation model

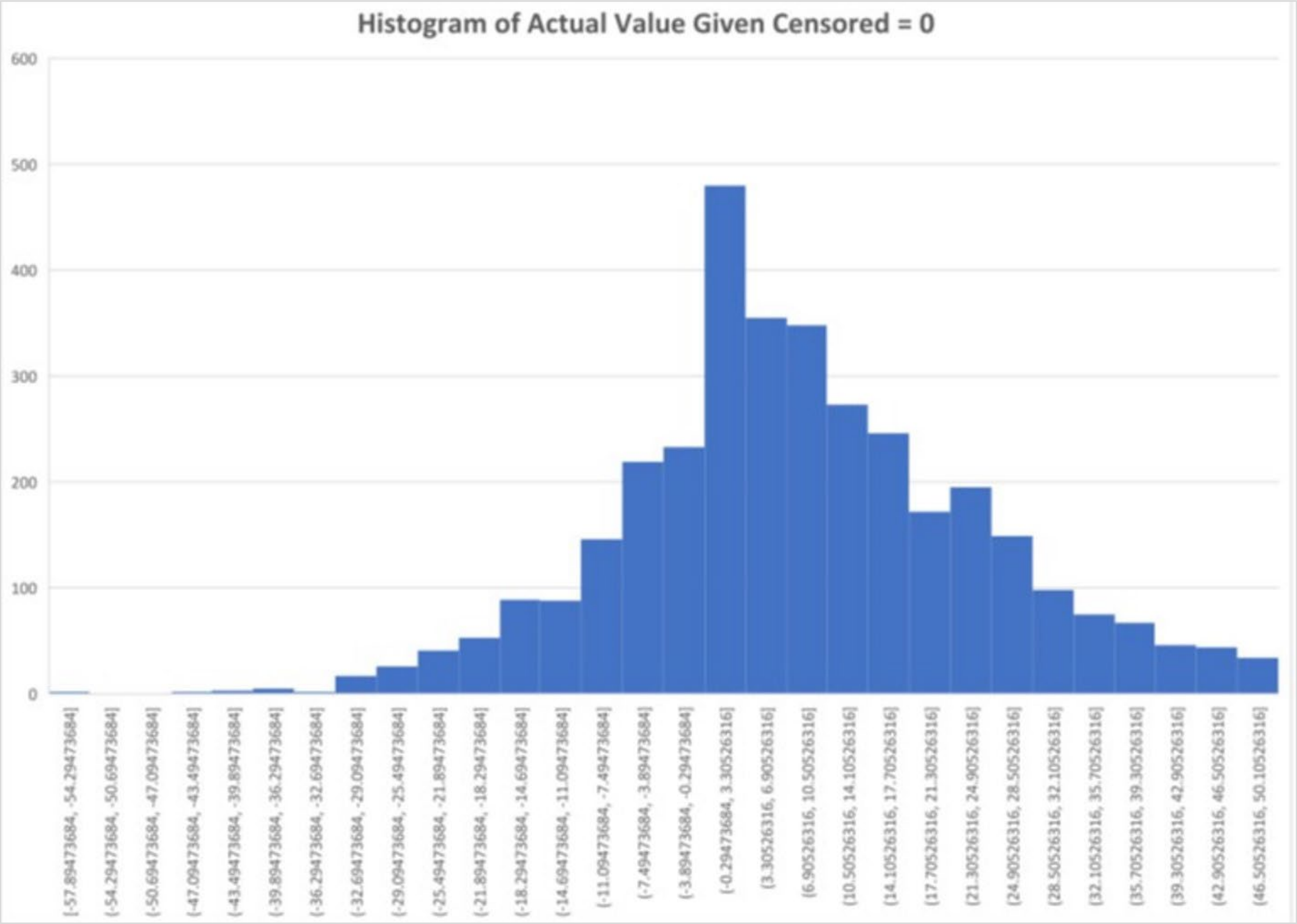
SC&A's comments on RPRT-0071 section 3.0, "Imputation Models and Multiple Imputation"

- Authors fit a lognormal distribution to data with 3,736 observations from 732 workers
- Average about 5 observations per worker
- So, data are clustered by worker
- If intracluster correlation is not small, need to adjust distribution fitting
- **Observation 7:** Acknowledge the impact of clustering

SC&A's comments on RPRT-0071, section 3.0, figure 3-1

- Figure 3-1 data:
  - Report indicates preponderance of data below LOD
  - Hard to see since the figure shows the entire range
  - Can't tell how well lognormal distribution describes data
  - SC&A graphed data below the LOD (graph on next slide)
  - More normal than lognormal
  - Highlights the need for individual analysis of each case
- **Observation 8:** Provide advice for data that are not lognormal

Graph: figure 3-1 data less than the LOD



SC&A's comments on RPRT-0071 section 3.0: Covariate data

- Covariate data
  - Page 8 of RPRT-0071 gives examples of other ways to generate multiple imputations
  - Use of covariate data not mentioned
- Sometimes dosages vary by population of worker
  - Populations may be distinguishable from available information
  - That information could be used to stratify a model
  - Or used as independent variables in a model
- **Observation 9:** Expand discussion of population subsets

SC&A's comments on RPRT-0071 section 3.0: MI variations

- There are many varieties of multiple imputation
- Traditional advice is to apply it within a Bayesian framework (Rubin, 1986)
  - Bayesian framework can be difficult to apply in practice
- RPRT-0071 uses less complex version than the Bayesian one
- Bayesian version might be unnecessarily complicated for our application
  - However, shouldn't assume all benefits of the full MI method apply to RPRT-0071 version

SC&A's comments on RPRT-0071 section 4.0, "Coworker Models"

- NIOSH (p. 9): "The statistician performing the analysis will make the judgment as to whether or not a given dataset is large enough to provide usable parameter estimates"
- Not just how large dataset is or how well model fits
- Statistician should quantify uncertainty in model parameter estimates
- Imputation adds uncertainty, and MI allows statistician to quantify it
- This report on MI is the place to explore how to quantify it

SC&A's thoughts on further research: Measuring uncertainty

- MI method could be implemented with single (k = 1), not multiple, imputation
- Would not alter the bias properties of the model
- Using k > 1 does, though, reduce the uncertainty in the final model estimates and provides a method for assessing that level of uncertainty
  - With k = 1, the level of uncertainty is hard to assess
- RPRT-0071 should highlight and discuss this benefit more
- Using MI data in co-exposure models allows users to
  - Properly account for the extra uncertainty of model parameters from imputation
  - Estimate resultant standard errors of estimates from models

SC&A's thoughts on further research: Measurement error

- Measurement error present in all measurements
  - Not just nonpositive ones
- Measured dose = true value + measurement error
- Simple approach usually models just true value
- RPRT-0071 notes measurement error is at play in nonpositive dose values
  - Attempts to account for that measurement error via imputation
- Since true dose value must be zero or more, nonpositive doses necessarily have negative measurement error
- Accounting for only negative measurement errors potentially biases the model

SC&A's thoughts on further research: Mixture models

- ORAUT-RPRT-0096 examined mixture models
- Mixture models can account for effects of measurement error
- Instead of relying solely on a lognormal probability model, it might make sense to use a mixture model that includes a lognormal component
- RPRT-0071 has a contradiction: It considers negative measurement errors but ignore positive ones
- **Observation 10:** RPRT-0071 does not acknowledge positive measurement error

Conclusion

- MI is state-of-the-art
- It is a credible approach
- The measurements it targets are the smallest ones, so the imputation method may not make much difference to probability of causation estimates in many cases
- Nonetheless, if MI is to be pursued, further exploration of issues related to our observations may benefit the dose reconstruction process

References

Oak Ridge Associated Universities Team. (2015). *External dose coworker methodology* (ORAUT-RPRT-0071, rev. 00).

Oak Ridge Associated Universities Team. (2021). *Multiple imputation applied to bioassay co-exposure models* (ORAUT-RPRT-0096, rev. 01). https://www.cdc.gov/niosh/ocas/pdfs/orau/oraurpts/or-rprt-96-r1-508.pdf

Rubin, D. B. (1986). Basic ideas of multiple imputation for nonresponse. *Survey Methodology, 12*(1), 37−47. https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1986001/article/14439-eng.pdf?st=Rm-mgRKJ

SC&A