Review of ORAUT-RPRT-0087 on Application of Regression in External Dose Reconstruction

Richard Griffiths, SC&A

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

November 8, 2024

Quantile regression in two reports

- Quantile regression (QR) introduced in two reports:
  - ORAUT (2018): ORAUT-RPRT-0087 describes QR and ordinary least squares (OLS) as methods to impute unknown doses from known
    - Argues that flexibility of QR allows for improvement in imputing unknown quantities (e.g., beta doses) from known quantities (e.g., gamma doses) over regression on order statistics (ROS)
  - NIOSH (2019): White paper is an application of QR
    - Employs QR to predict neutron doses from known photon doses to improve on standard 0.2 neutron-to-photon (N:P) imputation ratio for gaseous diffusion plants
- QR identified as overarching issue

Summary of SC&A's opinion

- SC&A concurs QR can be helpful
- However, other methods are better in some situations
- RPRT-0087 should:
  - Consider that there are datasets for which other techniques will be superior
  - Help the analyst determine which method is best in which situation
  - Provide more guidance about evaluating the chosen method once it is implemented
  - Give guidance about other factors that affect choice of method
  - Consider how to quantify the precision of imputations

Quantile regression vs. traditional regression

Traditional linear regression (OLS) methods

- Used to estimate likelihood of different beta doses values from a gamma dose value
- Estimate mean of beta dose distribution
- Percentile estimates all based on lines with same slope

Quantile regression (QR) method

- Used to estimate likelihood of different beta doses values from a gamma dose value
- Estimates a percentile, e.g., median
- Percentile estimates allow for lines with different slopes

Regression calculations and program goals

- QR and OLS calculations are different
  - **QR:** Minimize absolute value of differences between QR predictions and observed beta values
  - **OLS:** Minimize squared values of differences
- Subtle but important difference that affects
  - Complexity of software calculations
  - Interpretation of results

Is quantile regression the best regression method for dose reconstruction?

- Pre-analysis comparisons can suggest which regression methods might be appropriate
- Post-analysis tests can help determine if a regression method worked well
- RPRT-0087 considers both types of analyses but not in depth
- **Observation 1:** ORAUT-RPRT-0087 would benefit from expanded guidance on how to use preliminary statistical analyses to determine which regression methods are best.

Choice of methods based on dose data

- QR method is distribution agnostic
- What if data distribution can be established, though?
- If normal or lognormal: OLS might be most appropriate method
- If otherwise exponential family: general linear model (GLiM) might be best method
- Q-Q plot but otherwise no in-depth analysis in RPRT-0087
- Exploratory data analysis and statistical tests can help determine appropriate methods
- **Observation 2:** ORAUT-RPRT-0087 would benefit from more extensive discussion of pre-analysis methods such as Q-Q plots, histograms, box plots, and other statistical testing methods.

Post-analysis evaluation

- RPRT-0087 compares the fit of the QR and ROS models
  - Important step to determine which method is best for the given data
- Other evaluation methods beyond goodness-of-fit comparisons
  - Cross-validation and formal statistical testing of the prediction results
- RPRT-0087 would benefit from expanded coverage of post-analysis evaluations
- **Observation 3:** Methods of comparison discussed in ORAUT-RPRT-0087 should be expanded and formalized. Methods such as cross-validation and statistical testing of predicted results are warranted to inform the analyst of the appropriateness of their chosen regression approach.

Sample size consideration

- Because QR estimates percentiles with separate lines, it requires a relatively large sample size
- If the data distribution can be established, QR loses the important benefit of being distribution agnostic
- Might lead to inferences that are less precise than another regression method
- **Observation 4:** The QR method may be less appropriate than other regression methods when the sample size is relatively small and/or the underlying distribution is normal/lognormal.

Other comments on RPRT-0087

Observation 5: SC&A provides the following editorial comments for completeness of review and consideration if/when RPRT-0087 is revised:

1. Section 7.1 (example of a bad fit): specifying a model with a bad fit and comparing it to the QR model in fit doesn't mean QR is best
   - Better to compare the fit of QR model to that of the ROS and OLS models
2. The precision of imputations should be quantified
3. Equation (6-1) needs another subscript to make clear that each quantile has a separate intercept and slope in QR
4. β refers to both a type of dose (equation [4-1]) and a regression parameter (equations [4-1] and [6-1]) – could be confusing
5. Section 2.0: Clarify rationale for using types 1 and 3 data
6. Section 3.0: Clarify how censored gamma doses were treated

Covariate information

- Both RPRT-0087 and NIOSH (2019) use the QR method
  - QR model has same form in both
- But is the model used the best one?
- Are there other features that could improve the model?
- There are factors that can affect relationships between known and unknown dose measurements
  - Type of uranium, facility, location, atmospheric conditions, distance
- **Observation 6:** The usefulness of covariate information should be considered when determining the form of the QR model most appropriate for a given analysis.

Accuracy of QR method

- Measuring the precision of the QR model helps analysts determine if QR is appropriate for a given dataset
  - Precision of model parameter estimates
  - Precision of imputations
  - How well model fits the data
- **Observation 7:** To provide readers with quantified evidence of QR being an accurate method of analysis, NIOSH (2019, SRDB 176609) should measure the precision of the N:P ratio estimates and provide results of statistical testing of the QR fit for the N:P ratio estimates.

Construct practical guidelines

- “This report is intended as reference by statisticians who use the methods described in the report.” (RPRT-0087, Section 1)
- ORAUT-RPRT-0087 should list situations that QR might be best for and those for which other methods might be better
- SC&A’s review provides a start on such a list
- **Observation 8:** NIOSH should consider constructing clear and practical guidelines for the intended analyst that outline situations when QR is and is not appropriate.

Pros and cons of QR

Pro

- Don’t need to know data distribution for QR to work
- Flexibility in modeling percentiles
- No adjustment needed for heteroscedasticity
- Uses relevant information to understand relationships between dose types

Con

- More complex than other methods
- Requires relatively large sample sizes
- Extra work to deal with censored observations
- Can produce nonsensical results
- Often distribution is known, so QR loses its primary advantage

Conclusion

- SC&A believes QR can be a useful tool
- Which regression method to use needs to be determined on a case-by-case basis
- Methods other than QR, OLS, and ROS are available
- RPRT-0087 would benefit from expanded coverage of
  - How to determine which method is best in which situation
  - How to evaluate the chosen method once it is implemented
  - How sample size impacts the choice of method
  - How to quantify the precision of imputations

References

National Institute for Occupational Health (NIOSH). (2019). *Neutron dose assignment for K-25 and Portsmouth Gaseous Diffusion Plants* [White paper]. SRDB Ref. ID 176609

Oak Ridge Associated Universities Team (ORAUT). (2018). *Applications of regression in external dose reconstruction* (ORAUT-RPRT-0087, rev. 00). SRDB Ref. ID 170100