Non-compartmental methods for Below Limit of Quantification (BLOQ) responses.

Helen Barnett

Work with:

Thomas Jaki[1], Helena Geys[2], Tom Jacobs[2]

October 2018

[1]Lancaster University

[2]Janssen Pharmaceutica

Below Limit of Quantification Responses

- The Limit of Quantitation (LOQ) is the lowest concentration at which the analyte can not only be reliably detected, but at which some predefined goals for bias and imprecision are met.
- Measurements below this concentration are reported as BLOQ.
- These measurements are often present at the end of the sampling window.
- How can these left censored values be included in the PK analysis?

Example

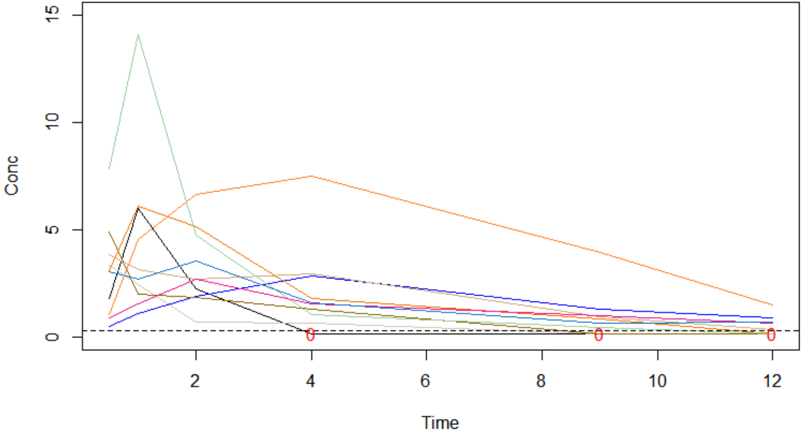


Figure: Dataset with some BLOQ measurements

Below Limit of Quantification Responses

- 7 methods for dealing with these censored observations from a modelling perspective are described by Beal (2001).
- These all involve fitting a PK model, either by discarding the BLOQ measurements, replacing the BLOQ measurements with some other value, or treat them as censored, then using maximum likelihood or least squares estimation to fit the model.
- Apart from simply replacing the censored values with 0 or $\frac{LOQ}{2}$ (which lead to biased estimates), NCA methods for such situations seems to be lacking.

Proposed Methods

We introduce 7 methods for comparison:

- **Method 1**: Replace BLOQ values with 0.
- **Method 2**: Replace BLOQ values with $\frac{LOQ}{2}$.
- **Method 3**: ROS Imputation.
- **Method 4**: Maximum Likelihood per timepoint Means.
- **Method 5**: Maximum Likelihood per timepoint Imputation.
- **Method 6**: Full Likelihood.
- **Method 7**: Kernel Density Imputation.

For all methods, we consider two different error structures: additive and multiplicative. For $i = 1, \ldots, n$ subjects observed at $j = 1, \ldots, J$ timpeoints.

Error Structures

1. **Additive**
   - Here we have:

     $$y_{ij} = \mu_j + \epsilon_{ij}$$

     where y_{ij} is the observed response for subject i at the jth timepoint.
   - The μ_j represent the population mean response at the jth timepoint.
   - The ϵ_{ij} are the differences between the y_{ij} and μ_j and the ϵ_{ij} are normally distributed.
   - In this case we use the arithmetic mean $\bar{y_j} = \frac{1}{n}\sum_{i=1}^{n} y_{ij}$ when we are estimating the AUC.

Error Structures

2. **Multiplicative**
   - Here, we instead have:

     $$y_{ij} = \mu_j e^{\epsilon_{ij}}$$

     which we can rewrite as:

     $$\log y_{ij} = \log \mu_j + \epsilon_{ij}$$

     where y_{ij} is the observed response for subject i at the jth timepoint.
   - Letting $c_{ij} = \log x_{ij}$, we have the geometric mean of the observations per time point $e^{\bar{c_j}}$ where $\bar{c_j} = \frac{1}{n}\sum_{i=1}^{n} c_{ij}$ and use this as our mean estimate of the response per timepoint when estimating AUC.

Simple Imputation

Method 1: Replace BLOQ values with 0
Simply replace any value that is below the limit of quantification by 0. When considering geometric means this method is mostly infeasible for calculating any estimate of variance.

Method 2: Replace BLOQ values with $\frac{LOQ}{2}$
Simply replace any value that is below the limit of quantification by $\frac{LOQ}{2}$.

Method 3: ROS Imputation

- Regression on Order Statistics is a semi-parametric method of dealing with censored values, with application in environmental statistics.
- It involves replacing the BLOQ values with different values, as opposed to methods 1 and 2 which replace all BLOQ values with the same value.
- In order to apply this method, we consider each time point in turn, starting with the earliest time point where a BLOQ value is observed.

Method 3: ROS Imputation

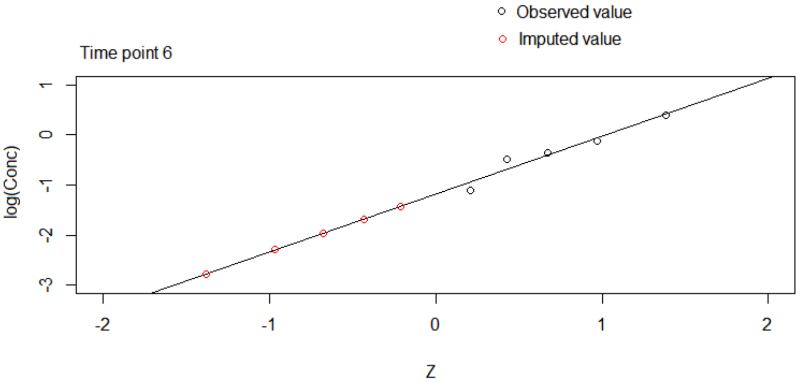
Assumptions per timepoint:

- n total measurements, m of which are left-censored by some value.
- All n measurements are from the same underlying distribution.
- This distribution can be transformed to Normal (e.g. Log-Normal).

Method 3: ROS Imputation

- Use empirical estimate of probability of exceeding the LOQ to calculate cumulative probabilities for each of the ordered $n - m$ observations.
- Calculate normal quantiles for these probabilities.
- Plot these quantiles against the transformed data, fitting a linear regression $f(X) = \widehat{a} + \widehat{b} \cdot Z \ (+\epsilon)$.
- Similarly calculate a second set of quantiles associated with the BLOQ measurements, and use the linear relationship to calculate the imputed value, assigning to a particular subject based on previous measurements for those subjects with BLOQ measurements.

Method 3: ROS Imputation



Method 4: Maximum Likelihood per timepoint (Means)

- For the jth timepoint, we consider a censored likelihood:

$$L(\mu_j, \sigma_j^2) = \left(\Phi \left(\frac{\log LOQ - \mu_j}{\sigma_j} \right) \right)^m \prod_{i=1}^{n-m} \frac{1}{y_{ij}\sigma_j\sqrt{2\pi}} \exp\left(\frac{-(\log y_{ij} - \mu_j)^2}{2\sigma_j^2} \right)$$

- We maximize this over μ_j and σ_j^2 to gain estimates $\hat{\mu_j}$ and $\hat{\sigma_j^2}$ for each timepoint.
- These are used to estimate the mean and variance of AUC.

Method 5: Maximum Likelihood per timepoint (Imputation)

- This method is a hybrid of methods 3 and 4,
- Using censored maximum likelihood to obtain values of $\hat{\mu}_j$ and $\hat{\sigma}_j^2$ for each timepoint,
- Then using these estimates to impute values onto the BLOQ responses in a similar fashion to ROS.
- Order in the same way as ROS.

Method 6: Full Likelihood

- This method takes into account correlation between responses at different timepoints.
- Whereas previously we consider them independently, and then we may induce some correlation by imputing values in a certain order.
- We here estimate the covariance matrix of observations assuming a multivariate normal or lognormal distribution.
- We consider that the (log-transformed) data are n iid observations from a $MVN_J(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution.
- The MLE of $\hat{\boldsymbol{\mu}}$ is a vector of the sample arithmetic means of the log transformed data.
- The exponential of this is the same as the geometric means.

Method 6: Full Likelihood

- Our objective is to estimate $\hat{\boldsymbol{\mu}}$ & $\hat{\boldsymbol{\Sigma}}$, the MLEs for the mean and variance of the observations assuming a multivariate normal distribution .
- For each subject $i = 1, \ldots, n$, we partition into censored and non-censored observations:

$$\boldsymbol{Y_i^{(c)}} = \boldsymbol{Y_{i,\{j:c_{ij}=1\}}}$$

$$\boldsymbol{Y_i^{(-c)}} = \boldsymbol{Y_{i,\{j:c_{ij}=0\}}}$$

where c_{ij} is an indicator, taking the value 1 if the observation is censored and 0 otherwise.

Method 6: Full Likelihood

- We then partition our parameters:
  - $\boldsymbol{\mu}^{(c)}$ = censored mean
  - $\boldsymbol{\mu}^{(-c)}$ = uncensored mean
  - $\boldsymbol{\Sigma}^{(c)(c)}$ = censored variance matrix
  - $\boldsymbol{\Sigma}^{(-c)(-c)}$ = uncensored variance matrix
  - $\boldsymbol{\Sigma}^{(c)(-c)}$ = censored/uncensored covariance

Method 6: Full Likelihood

- Then the conditional (on the uncensored values) distribution of the censored observations is multivariate normal with
  - Mean $\boldsymbol{\mu}^{(c)\prime} = \boldsymbol{\mu}^{(c)} + \boldsymbol{\Sigma}^{(c)(c)} \left(\boldsymbol{\Sigma}^{(-c)(-c)}\right)^{-1} \left(\boldsymbol{y}^{(-c)} - \boldsymbol{\mu}^{(-c)}\right)$
  - Variance $\boldsymbol{\Sigma}^{(c)\prime} = \boldsymbol{\Sigma}^{(c)(c)} - \boldsymbol{\Sigma}^{(c)(-c)} \boldsymbol{\Sigma}^{(-c)(-c)^{-1}} \boldsymbol{\Sigma}^{(c)(-c)^T}$
- The log likelihood is then:

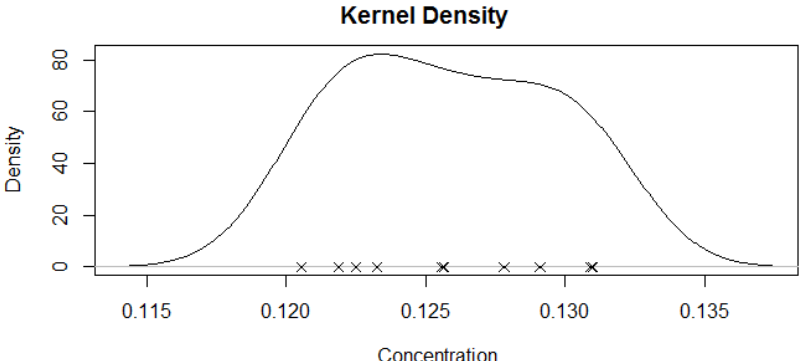
$$\sum_{i=1}^{n} \left(\log \left(F(\boldsymbol{\mu}^{(c)\prime}, \boldsymbol{\Sigma}^{(c)\prime}, \log(LOQ)) \right) + \log \left(f(\boldsymbol{\mu}^{(-c)}, \boldsymbol{\Sigma}^{(-c)}, \boldsymbol{y}_i^{(-c)}) \right) \right),$$

where F and f are the cdf and pdf of the multivariate normal distribution.

- This is maximized over $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ to give MLE $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$.
- The point estimate and variance of the AUC can be estimated from this.

Method 7: Kernel Density Imputation

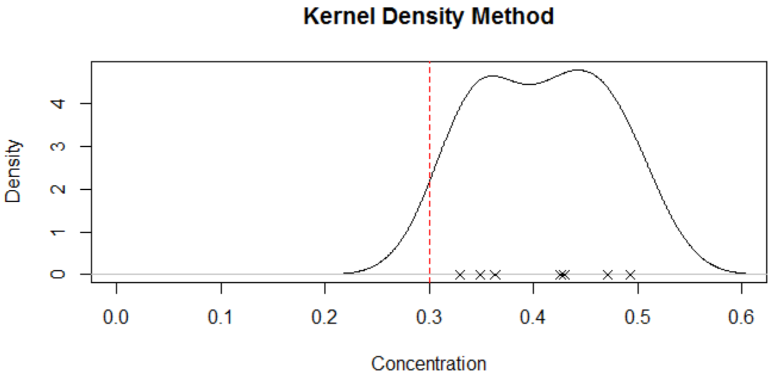
- This method is very different to the others, we do not assume any error distribution.
- We still consider the two cases, using arithmetic and geometric means of the responses.
- We consider each time point in turn.



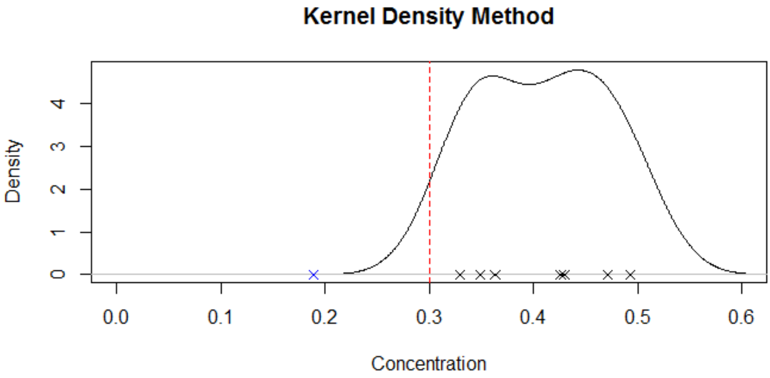
Method 7: Kernel Density Imputation

- For each timepoint:
  1. Calculate $\hat{f}_0$ based on uncensored data $\boldsymbol{y}^{(c)}$.
  2. Compute $E_{\hat{f}_0}(X|X < LOQ) = k_0$.
  3. Initialize $i = 1$.
  4. Calculate $\hat{f}_i$ based on k_{i-1} and uncensored data $\boldsymbol{y}^{(c)}$.
  5. Compute $E_{\hat{f}_i}(X|X < LOQ) = k_i$
  6. Let $i = i + 1$
  7. Repeat steps 4-6 until $|k_i - k_{i-1}| < \epsilon$ for some small value beϵ. This k_i is the value to be imputed.
- This process is repeated as many times as needed to get m imputed values for this timepoint.

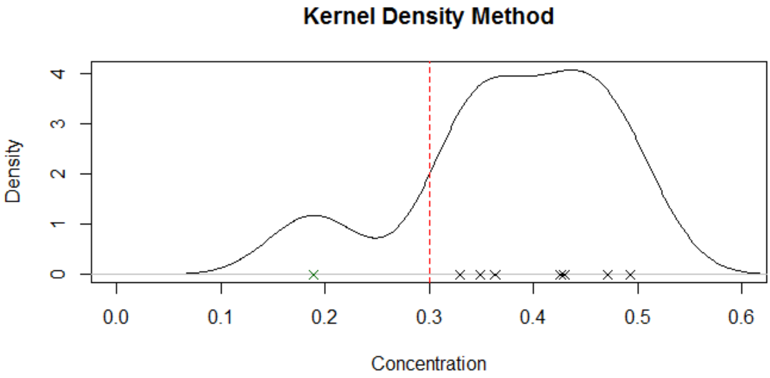
Method 7: Kernel Density Imputation



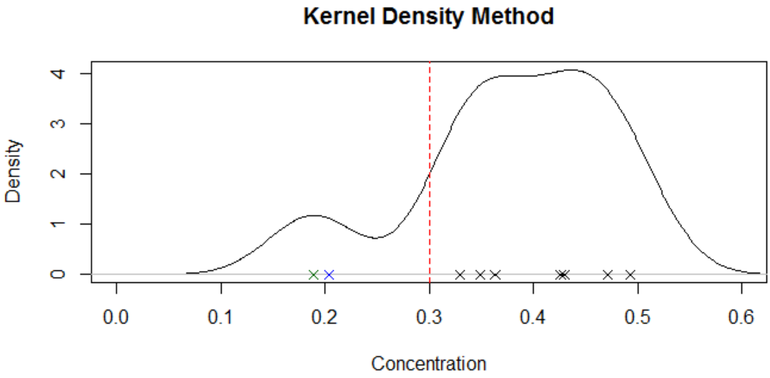
Method 7: Kernel Density Imputation



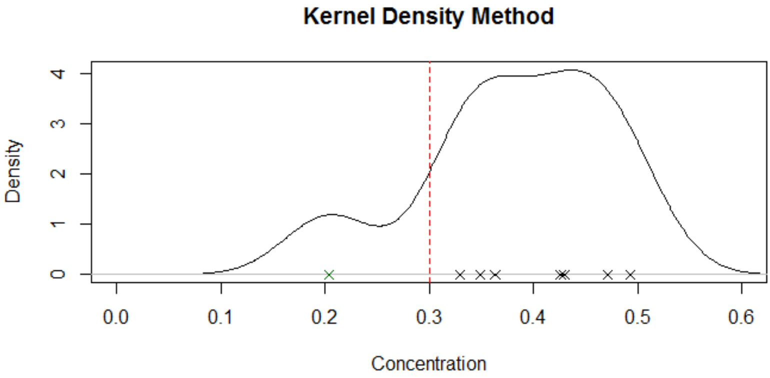
Method 7: Kernel Density Imputation



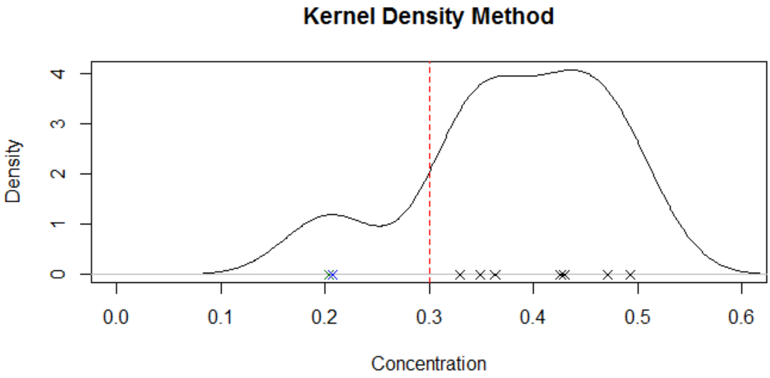
Method 7: Kernel Density Imputation



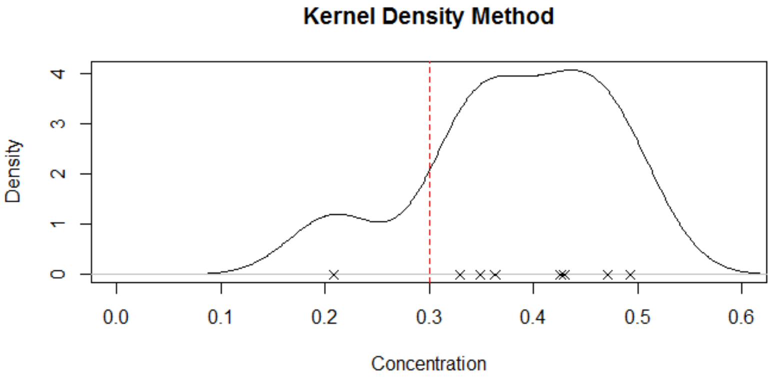
Method 7: Kernel Density Imputation



Method 7: Kernel Density Imputation



Method 7: Kernel Density Imputation



Application

$$y(t) = C(t)\exp(e(t)).$$

The PK model is a one compartmental IV dose model:

$$C(t) = \frac{dose}{V_d}\exp(CL \cdot t),$$

The error model is Normally distributed $e(t) \sim N(0, h(t))$ with

$$h(t) = 0.03 + 0.165\frac{C(t)^{-1}}{C(1.5)^{-1} + C(t)^{-1}}.$$

The time points used are $(0.5, 1, 1.5, 2, 2.5, 3)$.

- Fixed effects: parameter values $CL = 0.693$, $V_d = 1$ and $dose = 1$.
- Mixed effects: $CL = \widetilde{CL}\exp(\eta_1)$ and $V_d = \widetilde{V_d}\exp(\eta_2)$, with $\eta_1 \sim N(0, \omega_1^2)$ and $\eta_2 \sim N(0, \omega_2^2)$, corr$(\eta_1, \eta_2) = 0$ and $\omega_1 = \omega_2 = 0.2$. *(Beal, 2001)*

Application

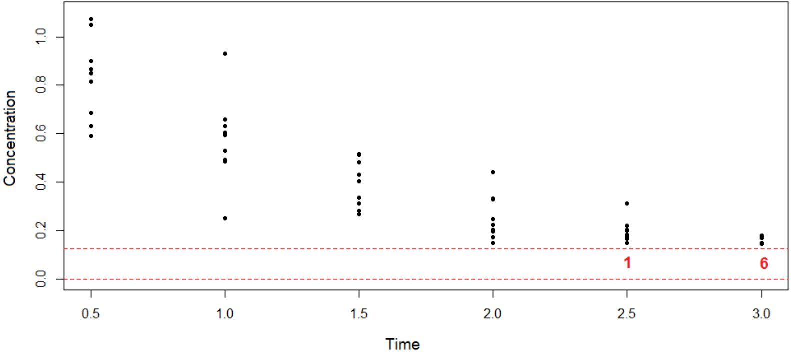


Figure: Example dataset

Application

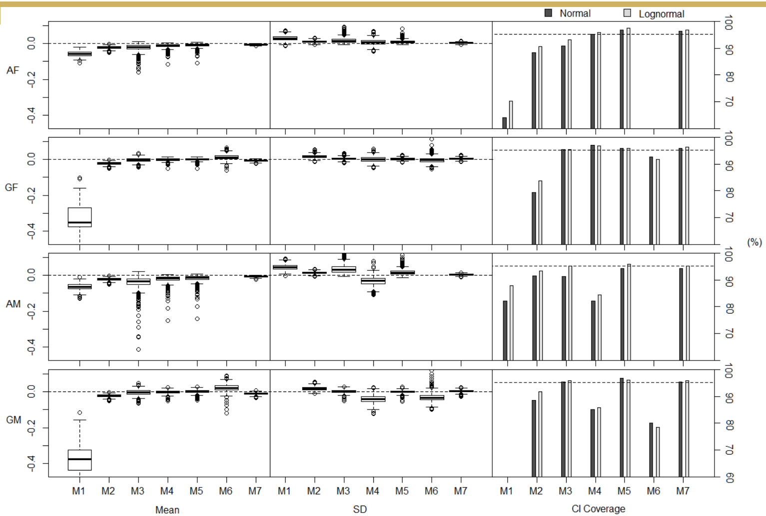


Figure: M1: Replace BLOQ values with 0, M2: Replace BLOQ values with LOQ/2, M3: ROS Imputation, M4: ML per timepoint Means, M5: ML per timepoint Imputation, M6: Full ML, M7: Kernel Density Imputation.

Application

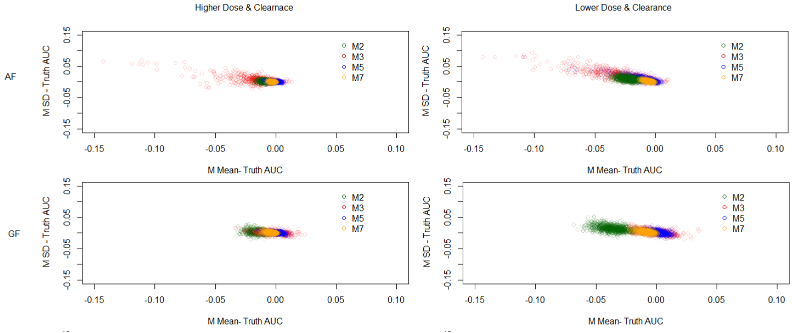


Figure: Comparison of the four best performing methods, fixed effects. M2: Replace BLOQ values with LOQ/2, M3: ROS Imputation, M5: ML per timepoint Imputation, M7: Kernel Density Imputation.

Application

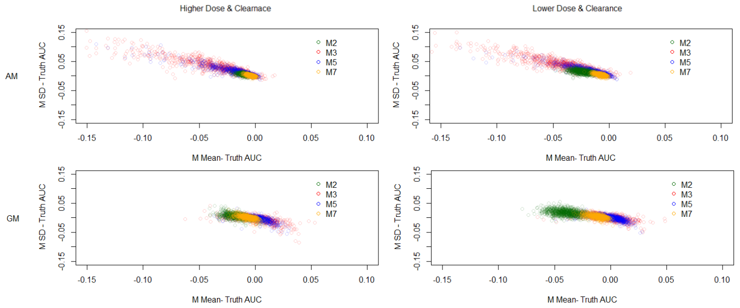


Figure: Comparison of the four best performing methods, mixed effects. M2: Replace BLOQ values with LOQ/2, M3: ROS Imputation, M5: ML per timepoint Imputation, M7: Kernel Density Imputation.

Concluding Remarks

- Simple imputation methods perform poorly, especially in scenarios with a large proportion of BLOQ responses.
- Methods that use maximum likelihood also fail to estimate the $\widehat{AUC}$ and its variance well.
- It is clear that the method of kernel density imputation is the best performing out of all the methods considered and is hence is the preferred method for dealing with BLOQ responses in NCA.
- The main advantage this method offers the advantage of lack of distributional assumptions.
- For application see R package 'BLOQ' available on CRAN (Nassiri et al., 2018).

References

Beal, S. L. (2001). Ways to fit a PK model with some data below the quantification limit. *Journal of Pharmacokinetics and Pharmacodynamics*, 28(5):481–504.

Nassiri, V., Barnett, H., Geys, H., Jacobs, T., and Jaki, T. (2018). BLOQ: Impute and Analyze Data with Observations Below the Limit of Quantification.