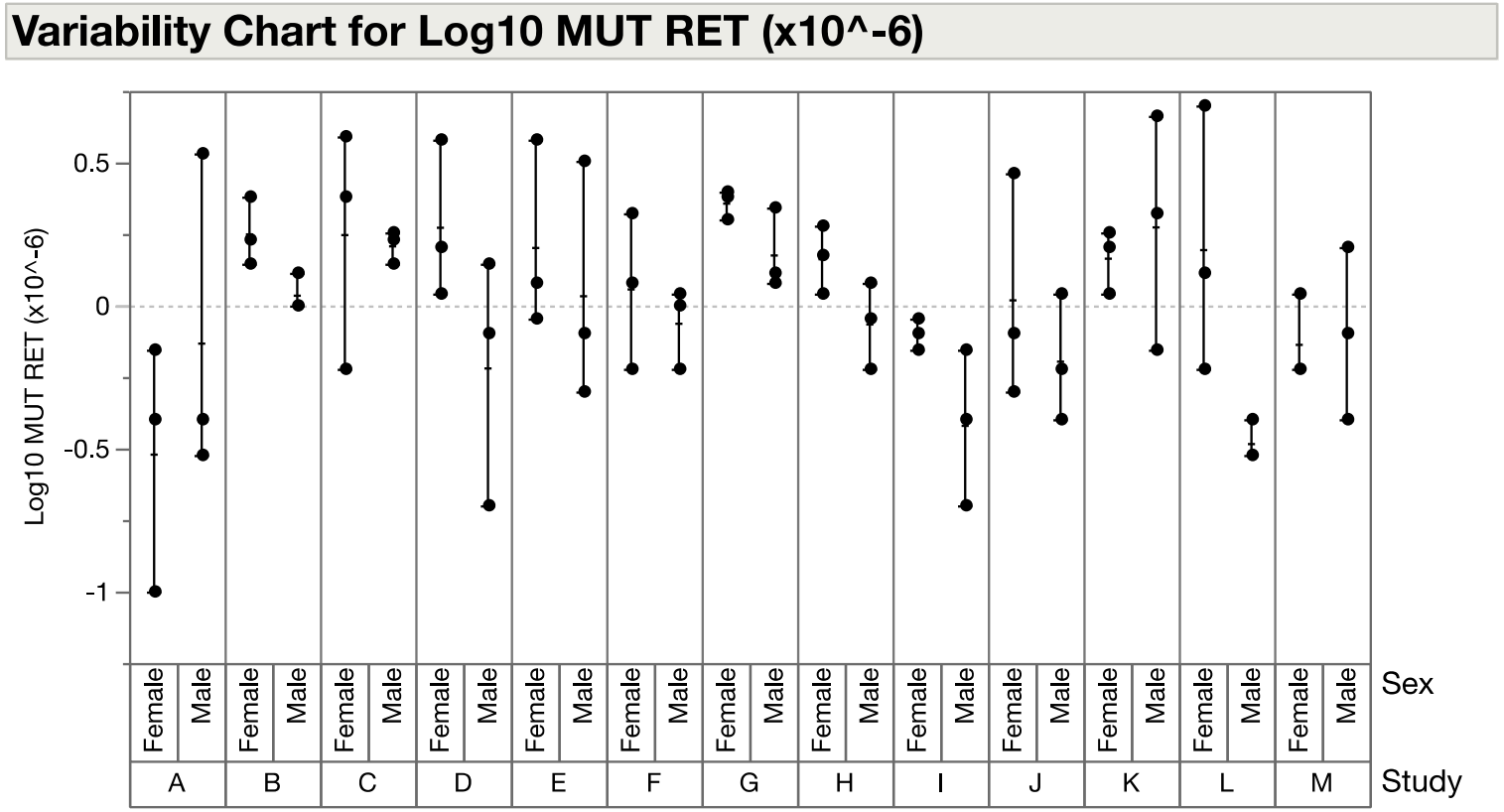
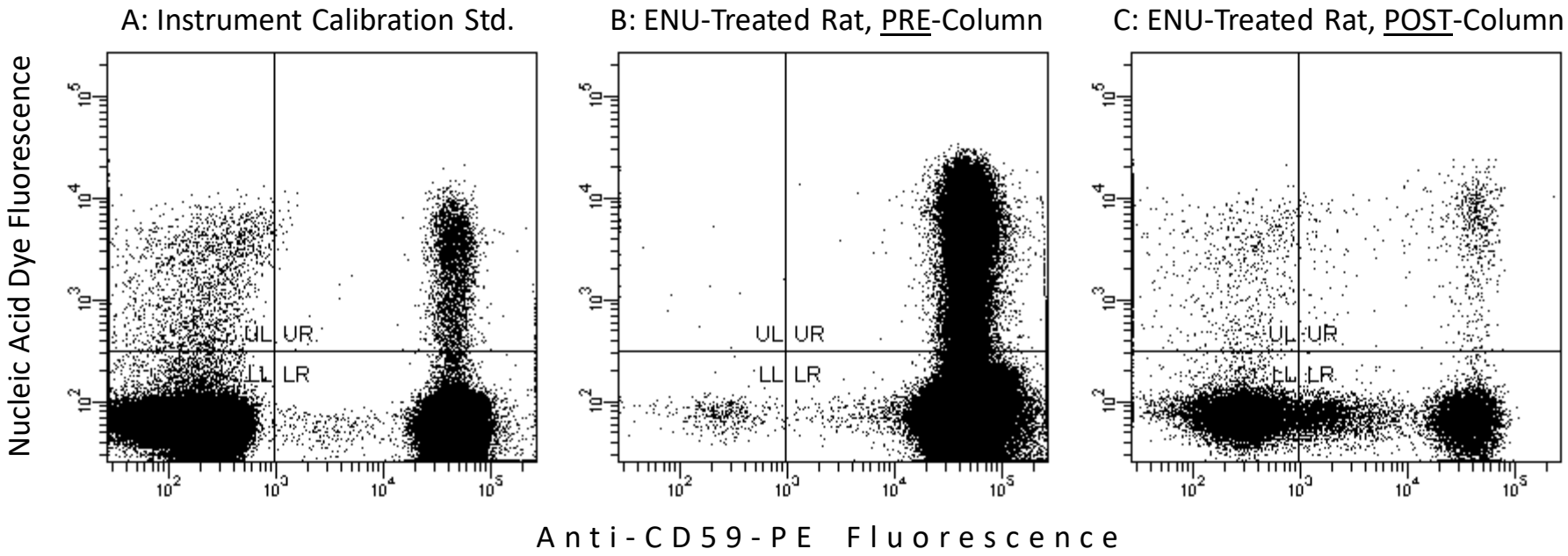
Approaches for Evaluating the Quality of Historical Negative Control Data: *Pig-a* Examples

Stephen D. Dertinger, Ph.D.
Director of Research
Litron Laboratories
sdertinger@litronlabs.com



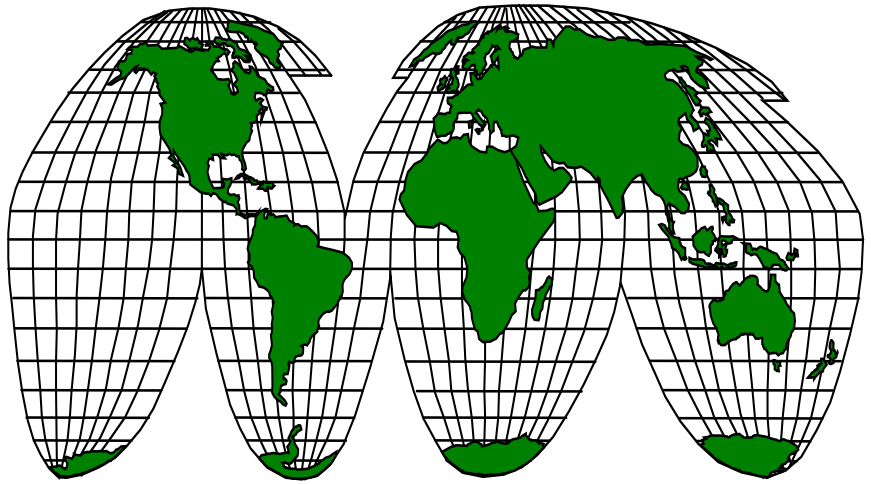
Disclosure

- S.D. works for Litron Laboratories, a company that sells reagent kits and offers testing services based on flow cytometric analysis of genetic toxicology endpoints, including *in vivo Pig-a* gene mutation (MutaFlow®) that will be discussed here



Acknowledgements

- This presentation is based on work accomplished by the 2023 IWGT Workgroup "Statistical Approaches & Data Interpretation"
  - Stephen Dertinger, Litron, chair
  - Dingzhou (Dean) Li, Pfizer, co-chair
  - Carol Beevers, Corteva Agriscience, rapporteur
  - Kristine Witt, NIH-NIEHS
  - Andreas Zeller, Roche
  - Bob Heflich, FDA-NCTR
  - David Lovell, St. George's University of London
  - George Douglas & Andrew Williams, Health Canada
  - Daniel Roberts, ToxSys
  - Robert Smith, Labcorp
  - Yoshifumi Uno, MB Medience
  - Changhui Zhou, InnoStar



Outline

- Use of Historical Control Distributions (HCD) in regulatory genetic toxicology studies
  - Note we'll focus on historical NEGATIVE control distributions for this presentation
- HCD Data Quality Checks
  - Initial considerations
  - Tools for evaluating sources of variation
  - Calculating useful intervals
- Rat *Pig-a* case study: Aristolochic acids I&II (AAs)
- Conclusions: the case for less rigidity/more nuanced use of HCD

Historical Control Distributions

- Genetox OECD Test Guidelines have harmonized their language regarding HCD and their uses
  - One component of demonstrating laboratory proficiency
  - One component of demonstrating study validity
  - ★ One of three assessments made to judge whether a particular study's response data are "clearly negative" or "clearly positive"
    A. Pair-wise test that considers concurrent vehicle/solvent control data
    B. Trend test
    C. Do the study data fall above or below an upper bound limit value derived from HCD?

I'll be focusing on the use of HCD for "**Criterion C**" purposes; as stated in a 2016 OECD publication, this allows for the consideration of "**biological relevance**"

HCD Data Quality Checks

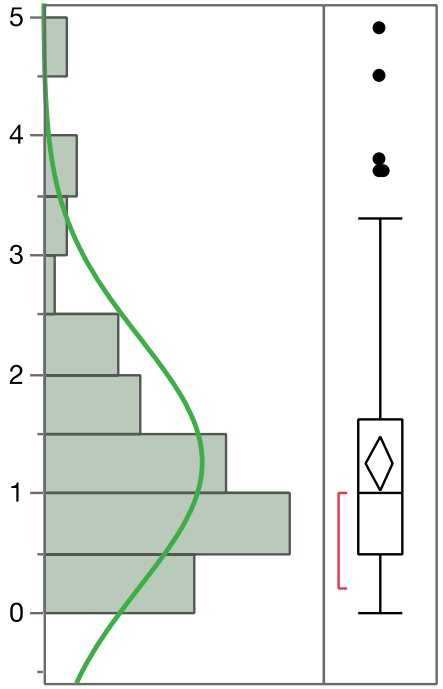
- HCD = rat mutant phenotype reticulocytes (MUT RET)*
  - 13 studies over 14 month period of time
  - N = 78 Crl:CD rats, 1/2 males and 1/2 females
- To simulate an assay that has drifted to an “out of control” status, some of the analyses use the **78 actual** rat MN-RET frequencies *plus* **12 simulated** values
  - Six that are 10-fold higher and six that are 10-fold lower than actual values

*from Dertinger et al., Environ. Mol. Mutagen. 60 (2019) 704-739; available upon request

Useful to First Evaluate Data Distribution

- Do the data approximate normal distribution?
  - Some assessment tools (e.g., Control Charts) and interval calculations (e.g., Tolerance Intervals) **assume normality**; transform as necessary if you intend to make use of these methods

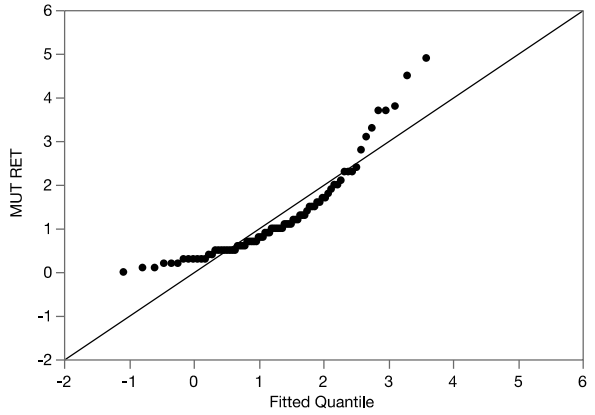
No. MUT RET per Million RET



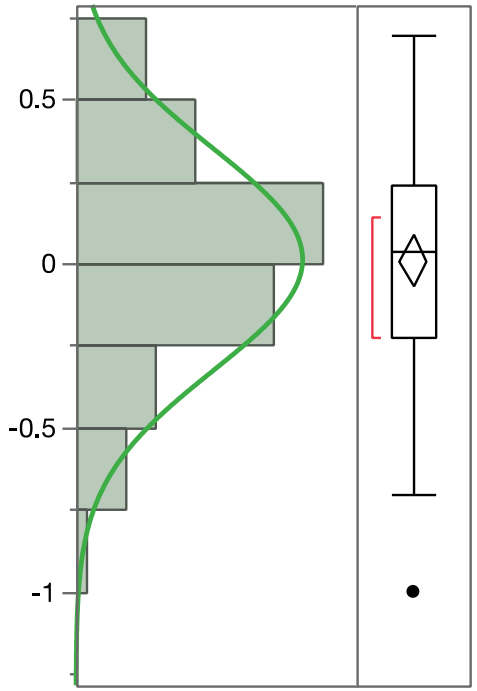
Goodness-of-Fit Test

	W	Prob<W
Shapiro-Wilk	0.8477235	<.0001*

	A2	Simulated p-Value
Anderson-Darling	3.5667791	<.0001*



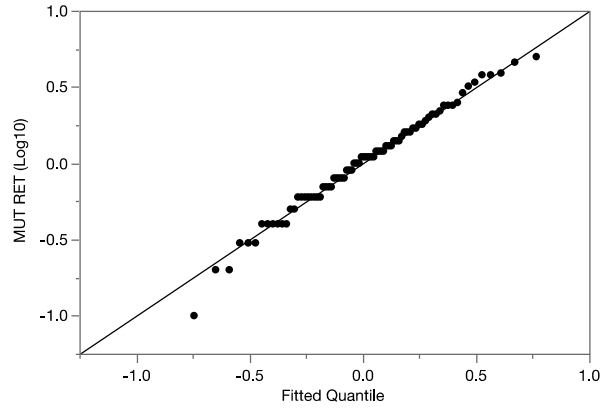
Log10 No. MUT RET per Million RET



Goodness-of-Fit Test

	W	Prob<W
Shapiro-Wilk	0.9878685	0.6722

	A2	Simulated p-Value
Anderson-Darling	0.2172292	0.8200

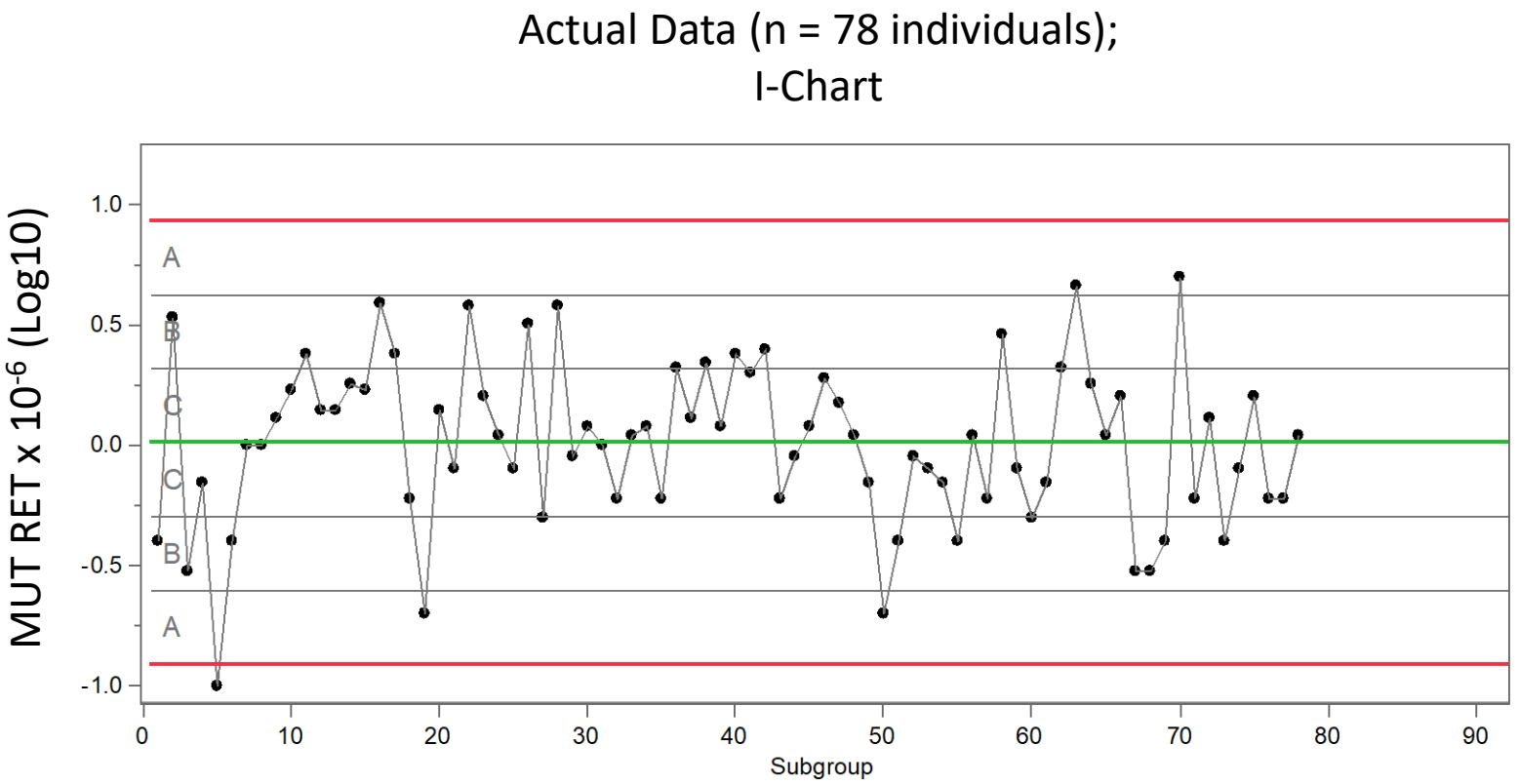


HCD Quality Checks

- There are a **number of useful approaches** for evaluating the quality of historical control data
- We'll look at each of the following, in turn:
  - Qualitative & semi-quantitative assessments
  - Methods used in the fields of manufacturing, process control
    - Control charts, with or without Nelson Rules
    - Stability Index
  - Variance Component Estimates [e.g., REstricted Maximum Likelihood (REML) analyses, Anova]

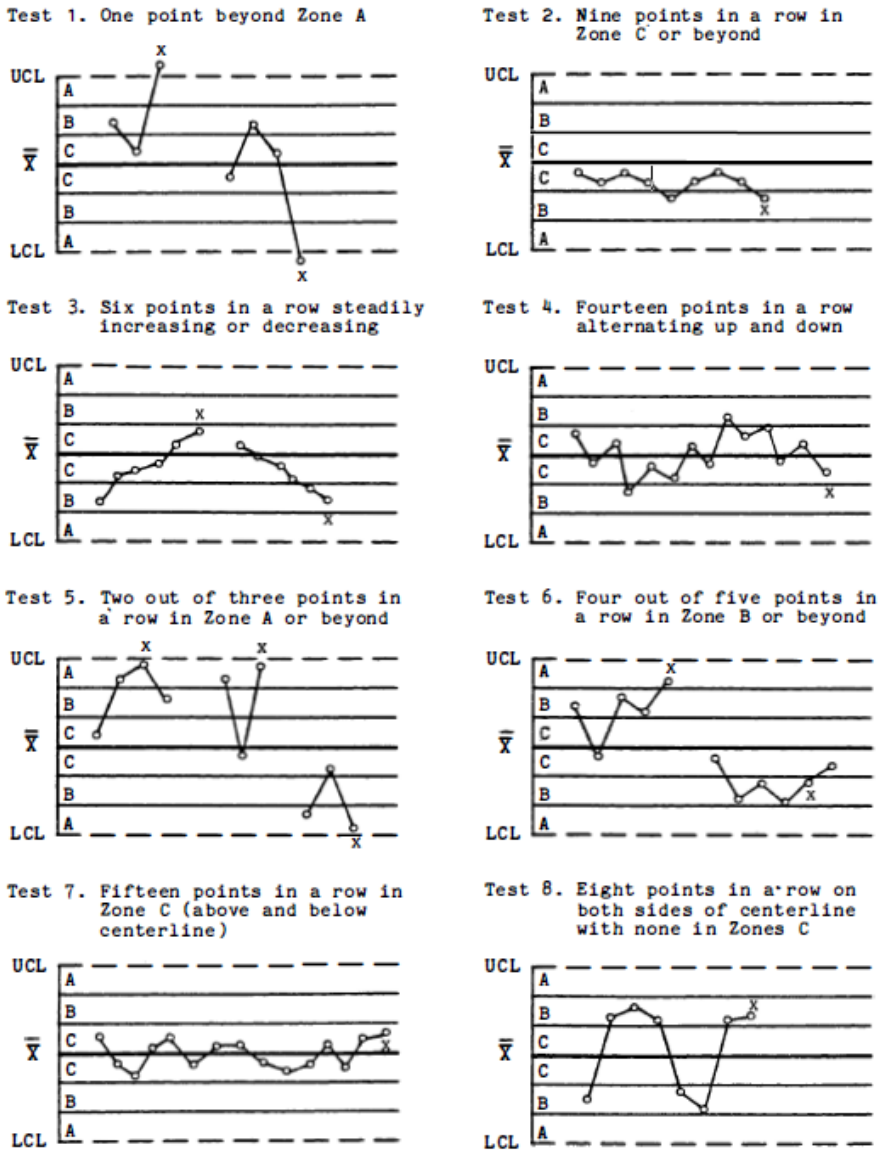
Qualitative & Semi-Qualitative Assessments

- Are the data consistent with published results from proficient labs?
- Is the level of variation across samples within a study and across studies comparable to published results from proficient labs?
- Is there obvious drift with respect to time?
- Control charts can help with these qualitative & semi-quantitative assessments



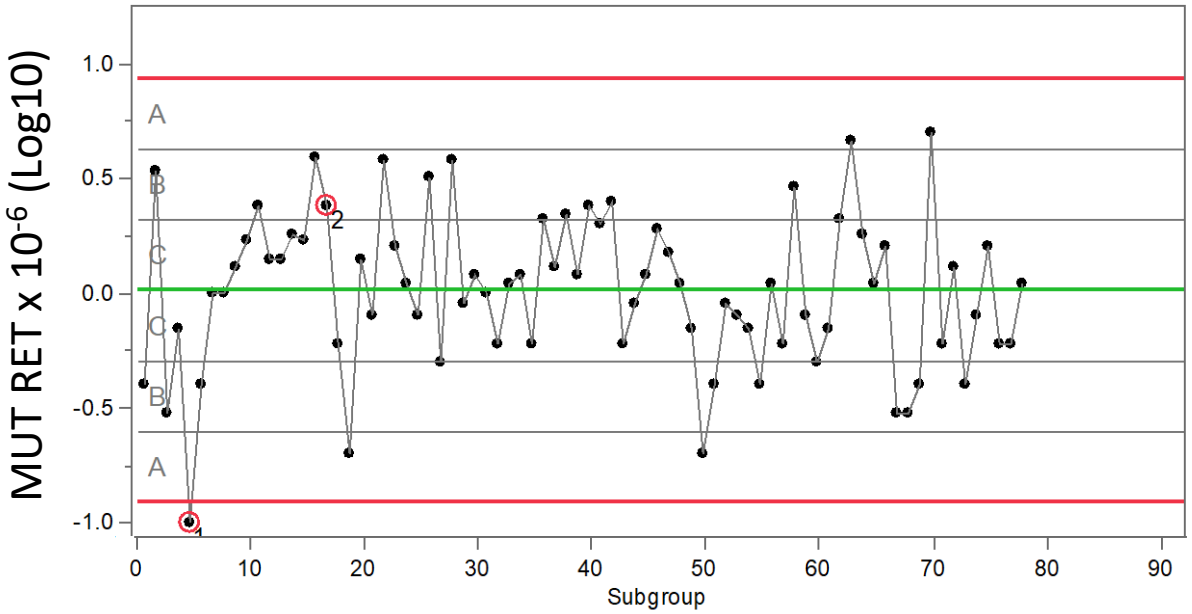
Control Charts with Nelson Rules, from Wiki…

- Nelson rules are a method in process control for determining whether some measured variable is out of control (unpredictable versus consistent)
- First published by Lloyd Nelson in the Journal of Quality Technology, 1984
- The rules are applied to a control chart on which the magnitude of some variable is plotted against time

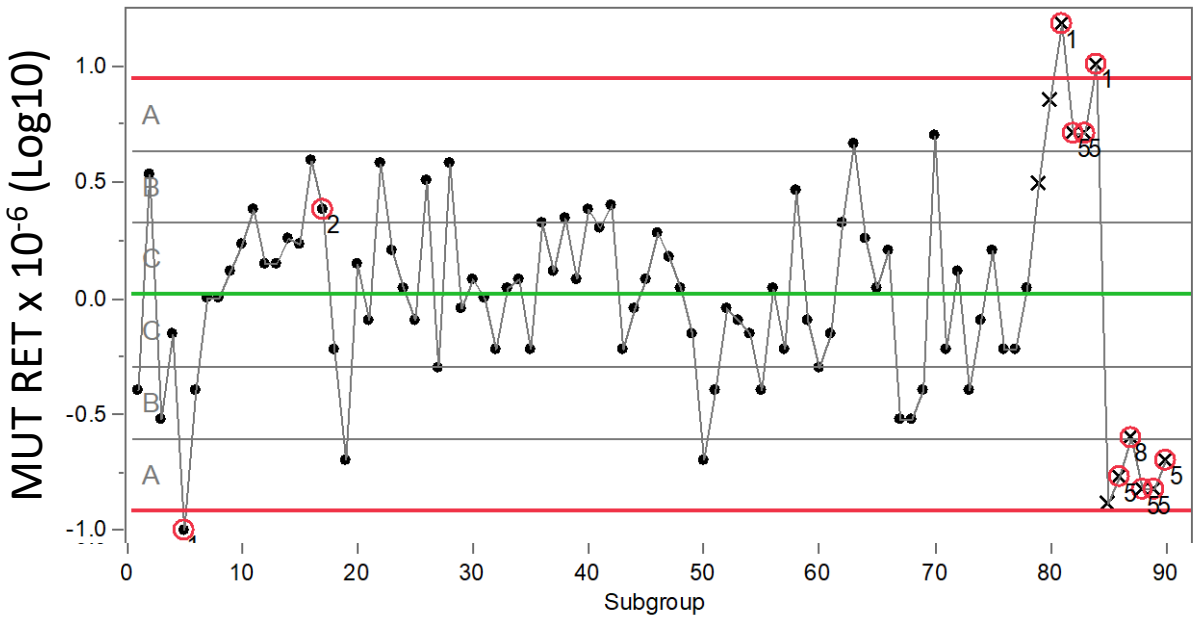


Control Charts with Nelson Rules, cont.

Actual Data (n=78);
Process appears to be "**under control**,"
relatively few Nelson Rules violations



Actual Data + 12 Simulated Samples;
Process has drifted to "**out of control**"
status, many Nelson Rules violations

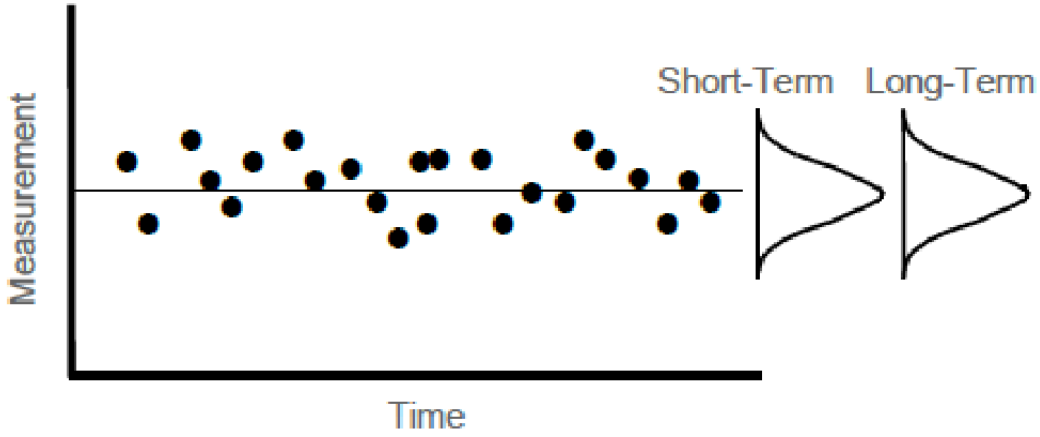


Stability Index

- Manufacturing and Process Control disciplines have developed a variety of tools for evaluating the stability (conversely, the variability) of a process
- One simple metric that might be leveraged for evaluating historical negative control data is the “Stability Index”
- Stability Index = Long-Term Sigma/Short-Term Sigma; close to 1.0 is evidence of stability, i.e., low variability

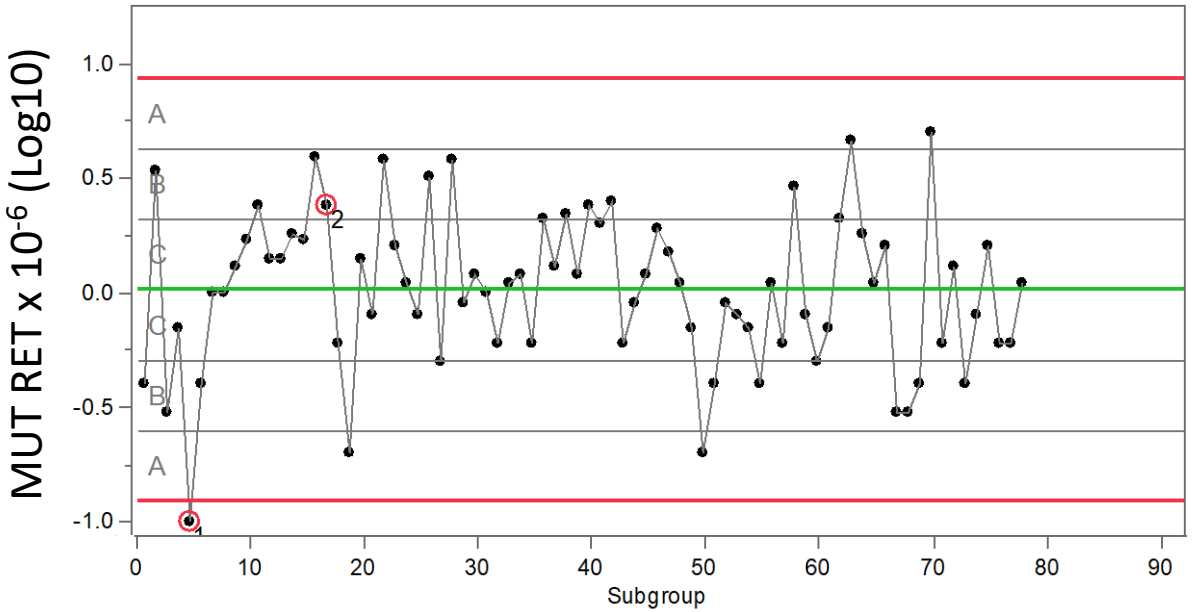
Low Stability, High Variability
(index value much greater than 1)

High Stability, Low Variability
(index value close to 1)

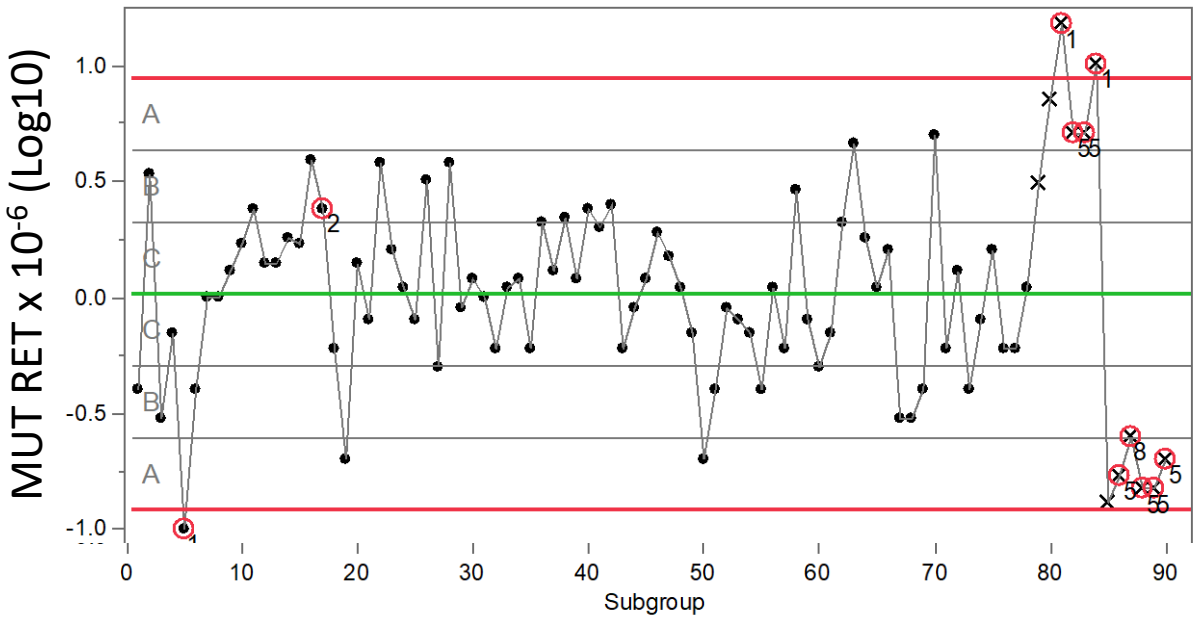


Stability Index, cont.

Actual Data (n=78);
Stability Index = **1.10**



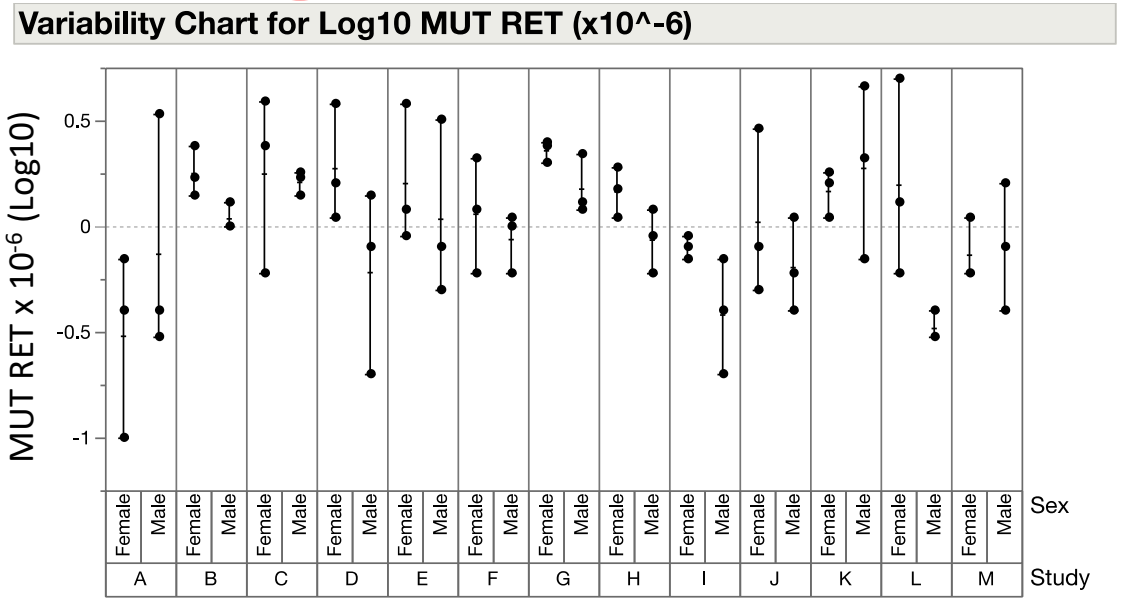
Actual Data + 12 Simulated Samples;
Stability Index = **1.40**



Sources of Variation

- Variance Component Estimates via REsidual Maximum Likelihood (REML), Anova, & Bayesian models can be useful for quantifying sources of variation

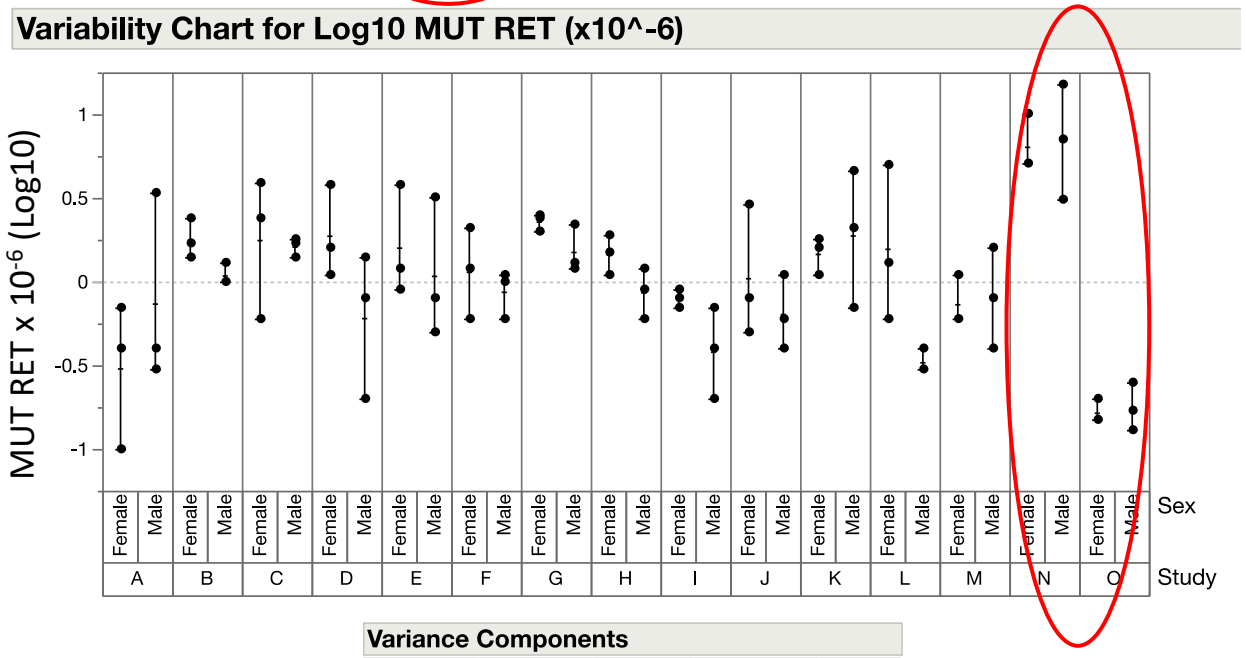
Actual Data (n=78);
Animal-to-animal variation dominates,
71% of total variation observed



Variance Components

Component	Var Component	% of Total	20 40 60 80	Sqrt(Var Comp)
Study	0.01781277	14.8		0.13346
Sex	0.01058523	8.8		0.10288
Study*Sex	0.00657102	5.4		0.08106
Within	0.08560173	71.0		0.29258
Total	0.12057075	100.0		0.34723

Actual Data + 12 "Out of Control" Samples;
Now inter-study variation dominates,
53% of total variation



Variance Components

Component	Var Component	% of Total	20 40 60 80	Sqrt(Var Comp)
Study	0.10452253	52.9		0.32330
Sex	0.00722481	3.7		0.08500
Study*Sex	0.00578724	2.9		0.07607
Within	0.07996240	40.5		0.28278
Total	0.19749698	100.0		0.44441

Calculation of Intervals

- Intervals that describe the distribution of the historical control data are useful for a variety of purposes
- BUT… it is premature to calculate and utilize intervals for the purposes described in genetic toxicology OECD TGs **until/unless an assay has been found to be "under control"**
  - Qualitative assessments
  - Control charts (could be supplemented with Nelson Rules, Stability Index)
  - Variance Component Estimates (e.g., REML, Nested Anova, Bayesian)
  - Etc.
- The following slide describes several **less appropriate** and several **more appropriate** means of calculating intervals for Criterion C purposes

Calculation of Intervals for Criterion C Purposes

- Usually not appropriate
  - Range
  - Confidence interval
  - 3 sigma “Control Limit”
- Appropriate
  - Quantile (does not assume normal distribution)
  - 2 sigma “Warning Limit”
  - Prediction interval (by default this assumes normal distribution; some software allows you to select non-parametric calculation that does not assume normality)
  - Tolerance interval (by default this assumes normal distribution; some software allows you to select non-parametric calculation that does not assume normality)

Rat *Pig-a* Case Study: AAs I & II*

- Crl:CD rats exposed to vehicle or each of three dose levels of AAs (I & II mixture)
- N = 6 per treatment group, 3 males and 3 females
- Treatment = oral gavage, 3 days per week x 3 consecutive weeks
- Blood collected on day 21 (relative to day 1 start of Tx)
- In the interest of time, we'll focus on one measurement: MUT RET
  - Log10 transformation

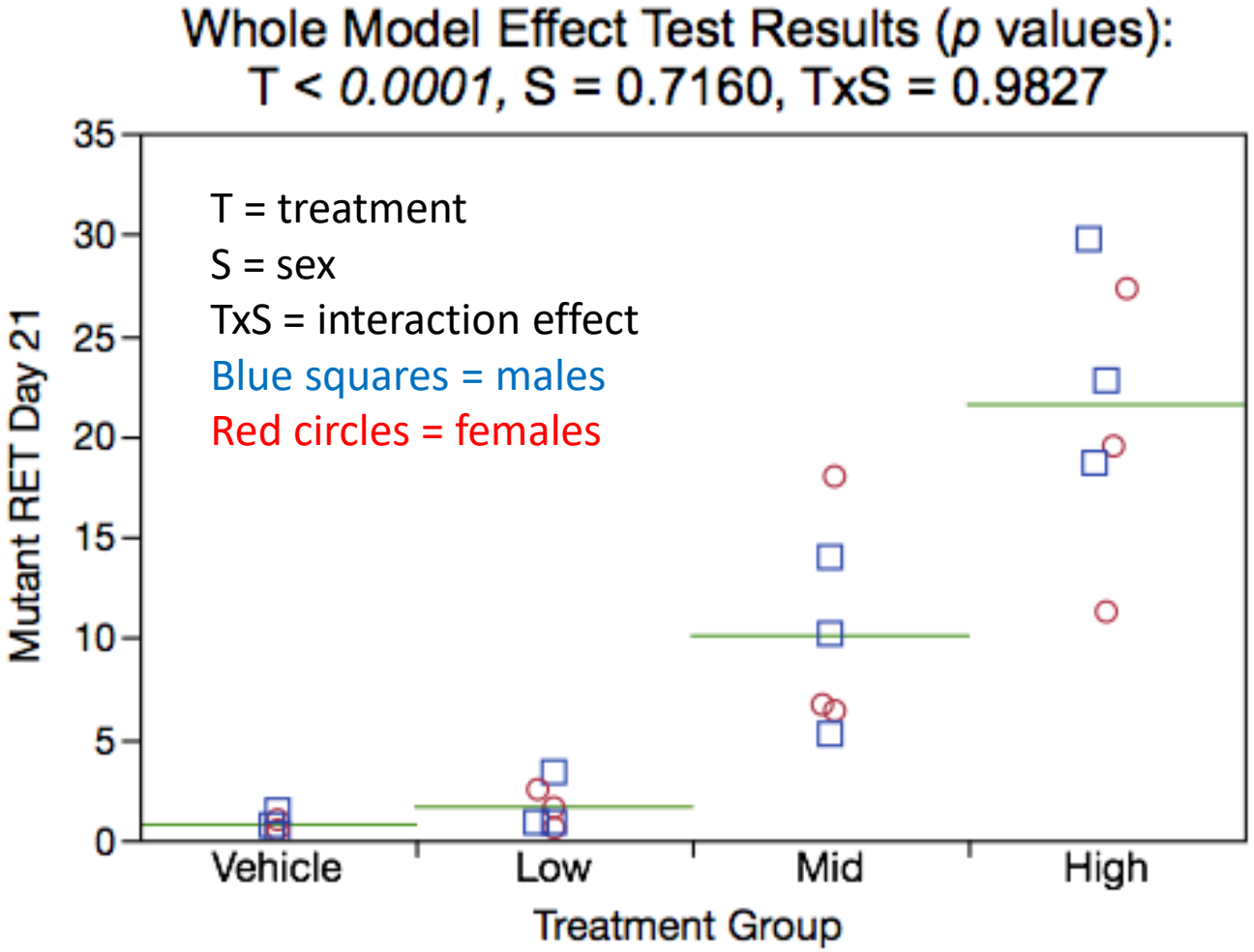
*data from Dertinger et al., Environ. Mol. Mutagen. 60 (2019) 704-739; available upon request

Rat *Pig-a* Case Study: AAs I & II

<u>Criterion A. Pair-wise tests</u>
Treatment had a statistically significant effect (ANOVA)

Mid and High dose groups are significantly elevated compared to concurrent control (Dunnett's)

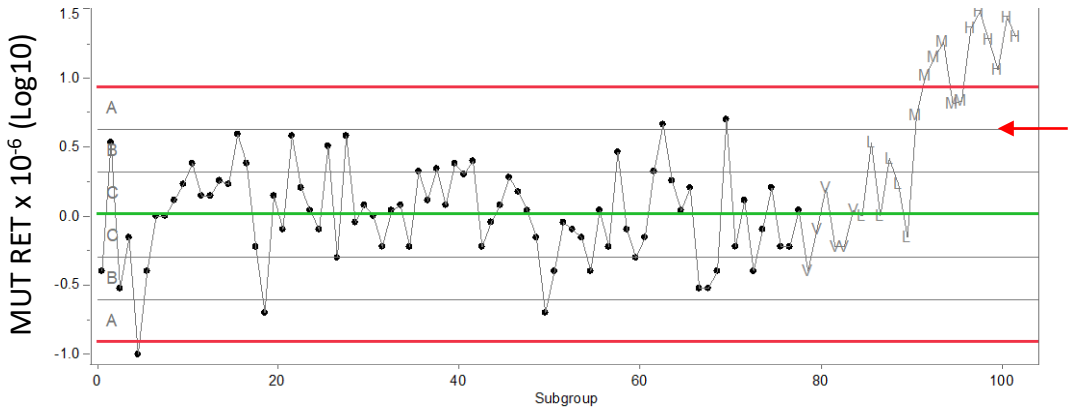
<u>Criterion B. Trend test</u>
Increased dose levels are associated with significantly increased frequencies of MUT RET (Regression)



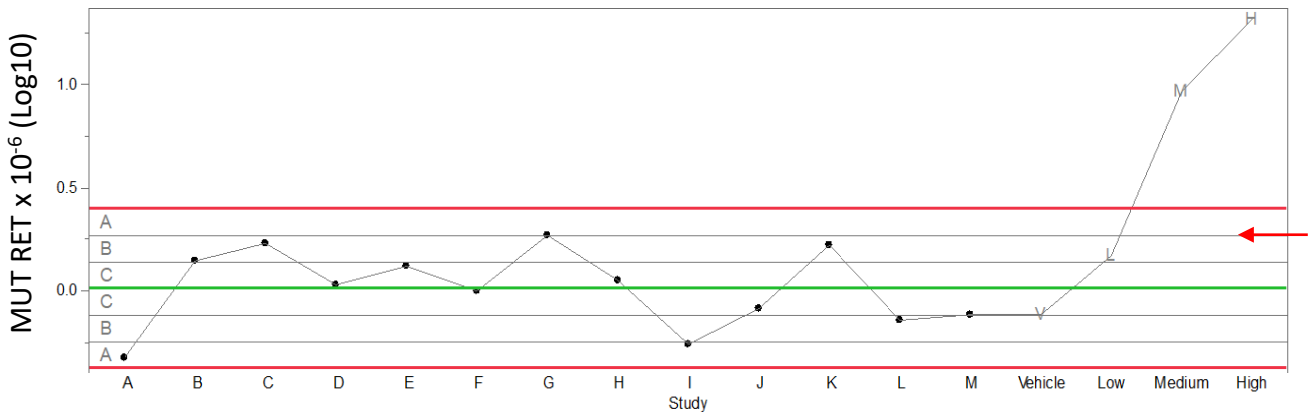
Rat *Pig-a* Case Study: AAs I & II

- Regarding Criterion C: one of the IWGT Workgroup's themes is that "**like should be compared to like**"
  - Individual animals in a particular study should only be compared to historical control-derived intervals that were also based on individual animals
  - Likewise, treatment group mean data in a particular study should only be compared to historical control-derived intervals that were also based on study means

I-Chart: Only Med and High AAs Tx group animals exceed Warning Limit ←



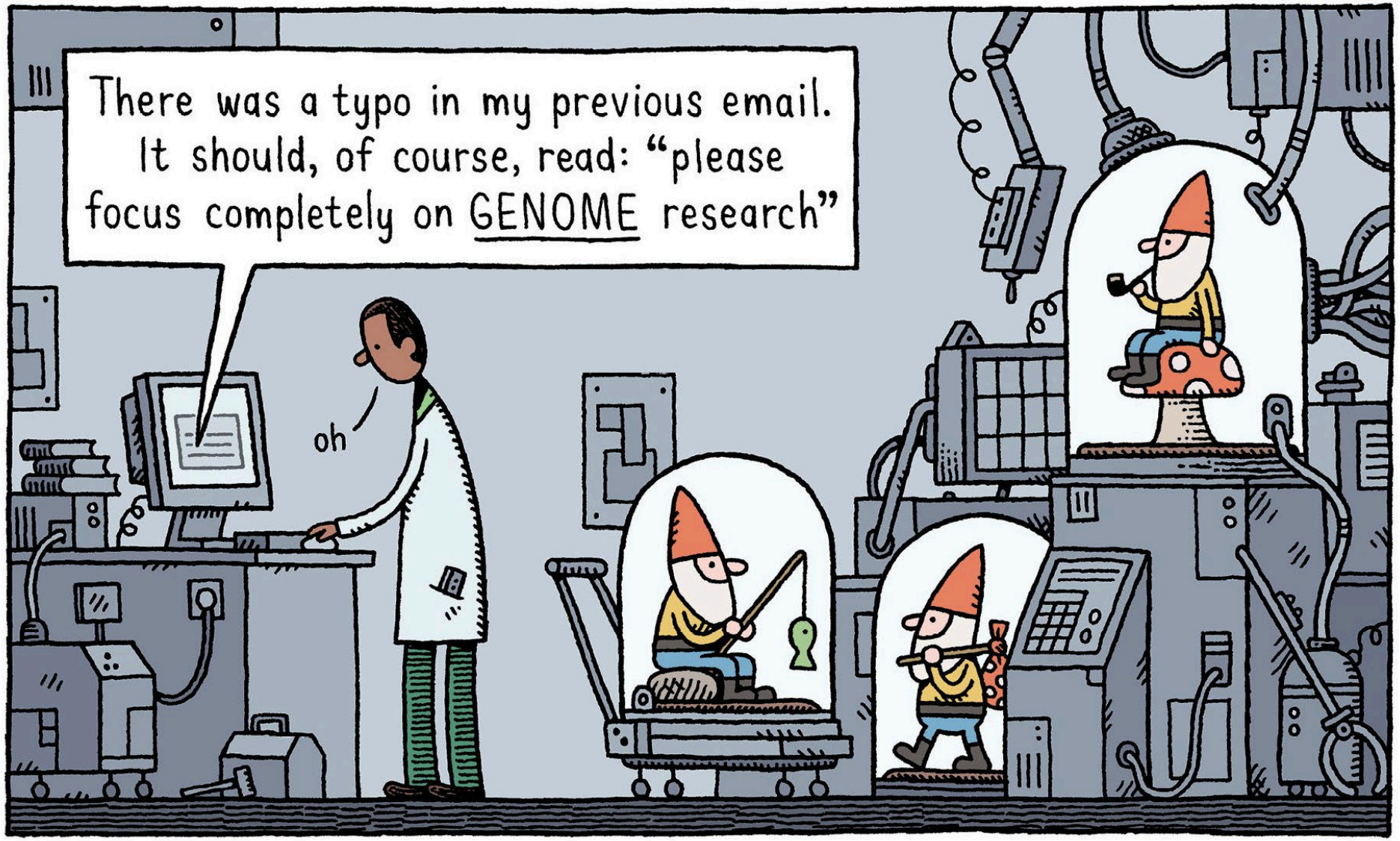
X-Bar Chart: Only Med and High AAs Tx group means exceed Warning Limit ←



Conclusions

- Regulatory genetic toxicology and OECD Test Guidelines make use HCD in several manners
- Important to assess the quality of HCD, and a number of qualitative, semi-quantitative, and quantitative approaches can be employed
- Rather than rigidly applying Criterion C, more nuance should be employed:
  - does the HCD describe inter-animal variation? If so, go ahead and apply criterion C;
  - otherwise do not place much weight on it
- There will need to be more detailed reporting of HCD and the type(s) of quality assessments undertaken for all regulatory study stakeholders to gain confidence in their use
- More work is necessary, in the meanwhile look for:
  - IWGT Workgroup Report, currently “in press,” available at Environ. Mol. Mutagen.
  - HESI-GTTC HCD survey results
  - note that OECD currently considering taking this subject on (Fall SPSF?)

Thank you for your attention! Questions?



Cartoon by Tom Gauld

in Department of Mind-Blowing Theories, published by Drawn & Quarterly, 2020