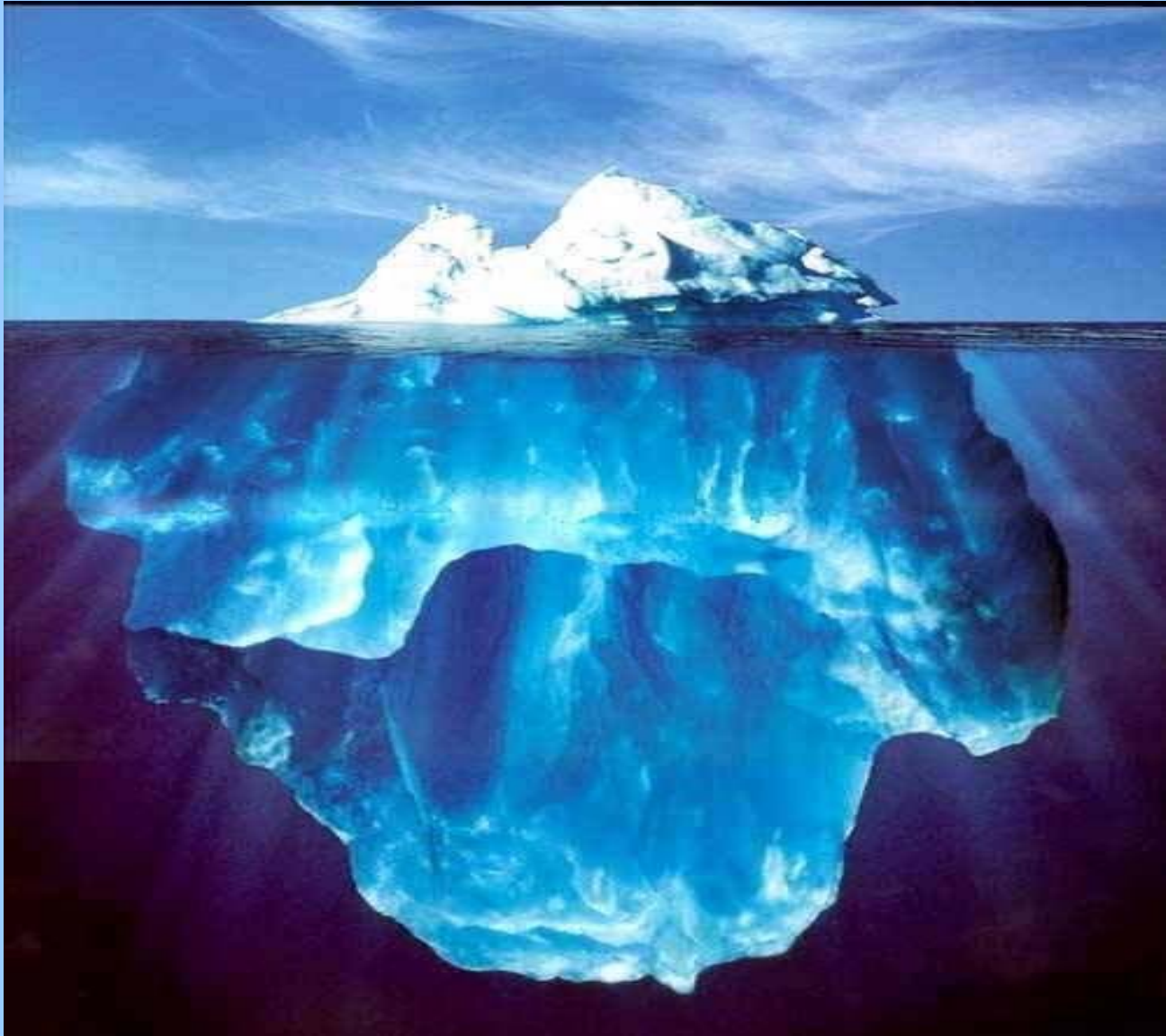


Model Calibration Group 2B

Day 2

Headline: Group 2B Braves Massive Iceberg – few casualties.



Topic: Implementation of the statistical model

- Question 1. When is it justifiable not to include all the data?
- Question 2. What sources of uncertainty/variability to include and how?
- Question 3. How to account for a variety of data collection issues?

Question 3. How to account for a variety of data collection issues?

**Different experimental designs
(longitudinal, serial sacrifice, closed chamber, number of data points)**

**Availability of individual vs.
aggregated data.**

Two issues here:

1) Statistical considerations are different for data from different study designs

e.g., correlations in data due to serial sampling from the same individual versus different individuals at different time points.

2) How to incorporate data from different studies into the model development.

Question 3. How to account for a variety of data collection issues?

- First need to characterize the statistical nature of the data from different experimental designs.
- Need to explicitly state assumptions about serial correlation if the data are from the same animals multiple time points (e.g., blood concentrations) versus independent data from different animals at different time points (destructive sampling, tissue concentrations).
- Models should be calibrated using the structure of the experimental protocol appropriately; failure to do this could lead to bias in the estimates (parameters that are incorrect, variance estimates that are inflated or underestimated, predictions that are wrong). There are no short-cuts here.

How is this mix of data addressed in different statistical approaches (Frequentist, Bayesian)

- Requires statistical expertise/consult. Consequences of inappropriately accounting for data structure include biasing distributions, variance, estimates, predictions, etc.
- Characterizing this bias is technically possible, but practically very difficult. Generally more information is available to estimate the mean than to estimate serial correlation.
- One way to address serial correlation is with the error term. It is OK to use mixed data but the error structure needs to be explicitly stated in terms of whether it represents inter- or intra- individual variability (or something else).
- In all cases, data sets should be available for subsequent model review and evaluation.

How is this mix of data addressed..... (continued)

- Another way to address serial correlation is in a Bayesian analysis - using priors for individual parameters for each test subject (for which there must be multiple data points) and generating population variances from the posteriors.
- Serial correlation is somewhat implicitly assumed in a PBPK model, i.e., each step in estimating a value for a state variable is based on the previous value.
- In calibrating a PBPK model, there may be a mix of data/time points, some from the same animal over time, and some from different animals at different time points (i.e., with different values for their initial state).
- If mixed data are too divergent, it may be difficult to have good fits or to estimate reasonable parameter values.

Question 4. What statistical method(s) are the most appropriate/practical?

Multiple data sets and sequential analysis. Is it useful?

- It is better to analyze all data at once, but there are circumstances when sequential analysis can be useful:
 - If there are time constraints
 - To better define priors (Bayesian) or starting points for likelihood estimates in a subsequent set.
 - Caveat: in the Bayesian analysis, evaluate the differences in a sequential approach compared to combined sets. Do outcomes differ dependent upon the sequence?
 - When application of the model calibrated against different subsets of the data, provide insight into differences amongst the subsets.

When sequential analysis can be useful (continued)...

- To sequentially estimate parameters that are time sequenced in a physiologically/biochemical pathway.
 - data address different processes in the pathway and the data are independent (e.g., metabolism then glutathione depletion).
- Because of limitations in the algorithm(s) available or in the user's capabilities. If the latter, however, advice is to consult statistical expertise.

Topic: How to evaluate model performance (deviance between predictions and data)?

Question 1. Absolute (“is this good enough?”)

- Evaluate goodness of fit, specifying approach.
- In some cases, there is qualitative evidence supporting constraints on a prediction (e.g. dose-metrics cannot exceed a reasonable physiological range or must fall into a practical range of variation to be useful).
 - In these cases, performance should be evaluated against qualitative expectations as well as the more quantitative, classical goodness-of-fit to the available data.

How to evaluate model performance.....

Question 2. Relative (“is A better than B?”), while accounting for degree of model parsimony.

- Evaluate differences in the application of the different models (e.g. differences in the dosimetry estimates).
- Evaluate the magnitude of the difference in the predictions and their variances.
- If no difference, the question is moot.

If the difference is large (large being a qualitative judgment), then...

- If there are data to resolve the better model/estimate, then do so.
- If two models are equal with respect to the data that are used to produce the models, and they have different underlying assumptions in their structure, then...
 - If time/resources allow, develop data to test the model assumptions i.e., the rigorous, quantitative, scientific approach.
 - If time/resources are constrained, choice of model may be based on a more qualitative evaluation of the scientific support for the assumptions using tools such as decision analysis, expert opinion, and other qualitative information. Policy and historical use considerations may influence this decision.