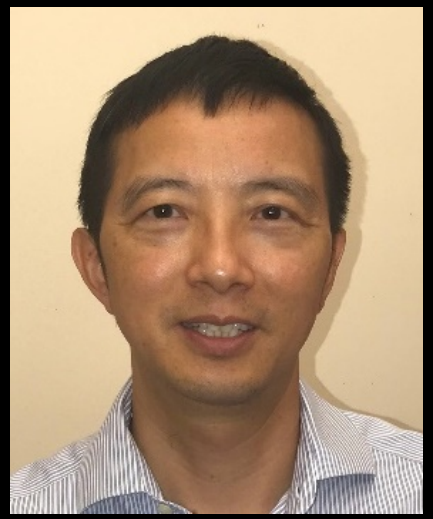
What you are about to witness is real

The participants are **not** actors






They are ACTUAL litigants with a case pending in




THE COURT OF PUBLIC OPINION

All parties have agreed to drop their claims

and have their cases settled

HERE

and have their cases set

HEA

with Judge Ken Kowalski


<0.05
THE
PEOPLE'S
COURT


<0.05
THE
PEOPLE'S
COURT

MODELING WITHOUT BORDERS
ACoP 10

THE CASE OF THE CRISIS IN CONFIDENCE

ACoP 10

JAMES ANTHONY ROGERS

PLAINTIFF

ERIC BURROUGHS JORDIE DEFENDANT

ACCUSED OF: OVERCONFIDENCE

Evidence Exhibit A : Excerpt From Doctor Jordie's Report

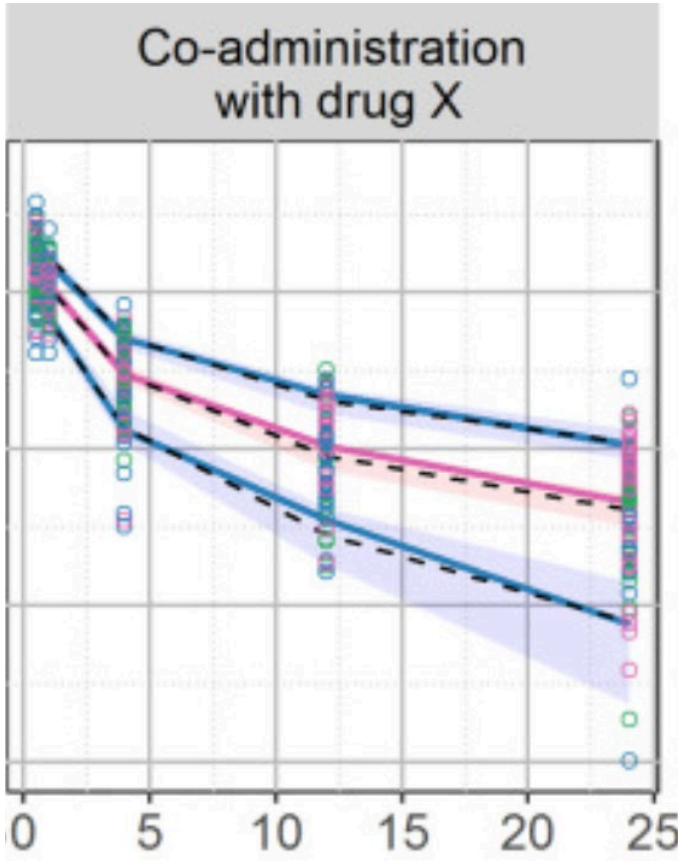
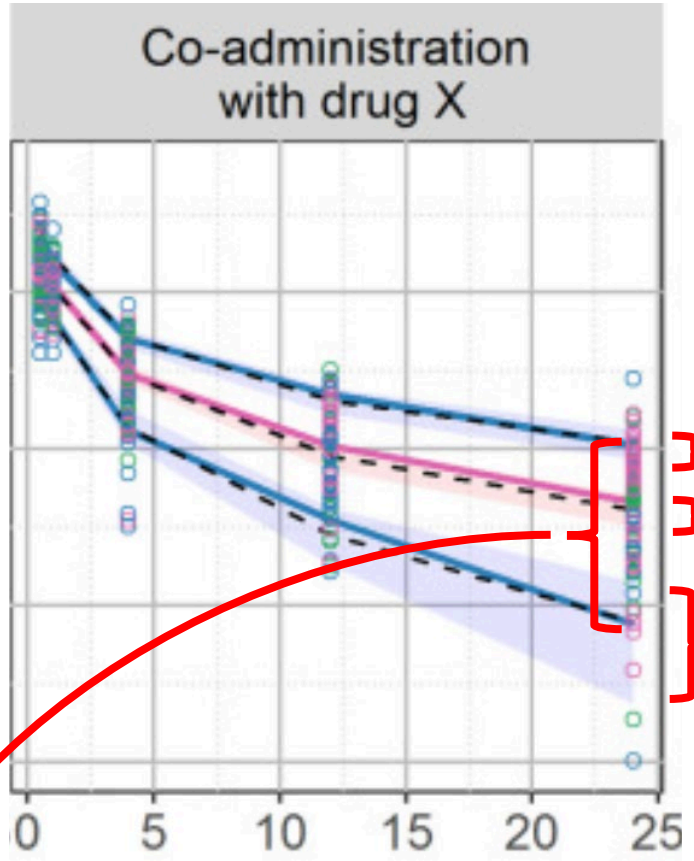


Figure 1: The blue and red lines are the observed percentiles (10, 50, and 90th percentiles), the blue and red ribbons are the corresponding 95% **confidence intervals.** The dashed black lines are predicted percentiles. [As such, the upper and lower dashed lines demarcate 80% **prediction intervals.**] Observations corresponding to the lowest, median, and highest doses are shown in blue, pink, and green, respectively.

Evidence Exhibit A : Excerpt From Doctor Jordie's Report



This terminology is debatable but I'm not bringing charges on this point (but I encourage your honor to comment on this)

I must most strenuously object to using the term "confidence interval" in this way

Figure 1: The blue and red lines are the observed percentiles (10th, 50th, and 90th percentiles), the blue and red ribbons are the corresponding 95% **confidence intervals.** The dashed black lines are predicted percentiles. [As such, the upper and lower dashed lines demarcate 80% **prediction intervals.**] Observations corresponding to the lowest, median, and highest doses are shown in blue, pink, and green, respectively.

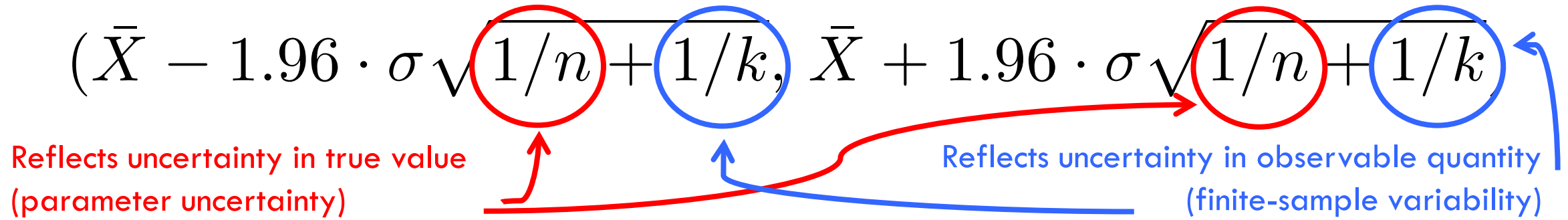
Evidence Exhibit B : Excerpts from Stats Literature

Suppose $X_1, \ldots, X_n \sim \text{Normal}(\mu, \sigma)$, where σ is known.
A 95% confidence interval for **the true mean** μ is:

$$(\bar{X} - 1.96 \cdot \sigma\sqrt{1/n},\ \bar{X} + 1.96 \cdot \sigma\sqrt{1/n})$$

A 95% prediction interval for **the sample mean** $\bar{X}^{(\text{new})}$ of the next k observations is:

$$(\bar{X} - 1.96 \cdot \sigma\sqrt{1/n + 1/k},\ \bar{X} + 1.96 \cdot \sigma\sqrt{1/n + 1/k})$$



Closing Arguments

$$(\bar{X} - 1.96 \cdot \sigma\sqrt{1/n + 1/k},\ \bar{X} + 1.96 \cdot \sigma\sqrt{1/n + 1/k})$$

Reflects uncertainty in true value (parameter uncertainty)

Reflects uncertainty in observable quantity (finite-sample variability)

The simulations underlying Dr. Jordie's VPC did not incorporate anything analogous to this, i.e. they did not incorporate parameter uncertainty

The simulations underlying Dr. Jordie's VPC reflected precisely this kind of predictive variability, in order to define the range of expectation of an observable quantity.

The resulting intervals are therefore analogous to a sort of ideal prediction interval for a statistic, but they really are

<u>nothing like confidence intervals</u> !!!

Doctor Jordie's Defense

CPT: Pharmacometrics & Systems Pharmacology

TUTORIAL

Model Evaluation of Continuous Data Pharmacometric Models: Metrics and Graphics

Abstract

This article represents the first in a series of tutorials on model evaluation in nonlinear mixed effect models (NLMEMs), from the International Society of Pharmacometrics (ISoP) Model Evaluation Group. Numerous tools are available for evaluation of NLMEM, with a particular emphasis on visual assessment. This first basic tutorial focuses on presenting graphical evaluation tools of NLMEM for continuous data. It illustrates graphs for correct or misspecified models, discusses their pros and cons, and recalls the definition of metrics used.

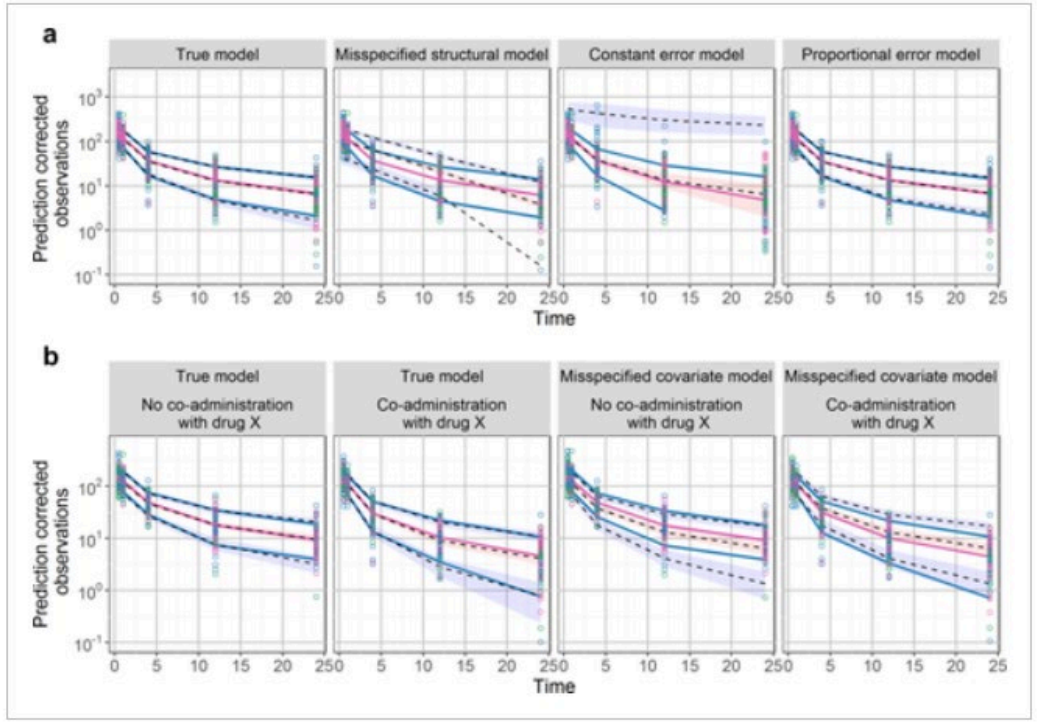


Figure 2 Open in figure viewer | PowerPoint

The prediction-corrected visual predictive checks (pcVPCs) plots of different models. The blue and red lines are the observed percentiles (10, 50, and 90th percentiles) the blue and red ribbons are the corresponding 95% confidence intervals. The dashed black lines are predicted percentiles. Observations corresponding to the lowest, median, and highest doses are shown in blue, pink

CPT Pharmacometrics Syst. Pharmacol. (2017) **6**, 87–109; doi:10.1002/psp4.12161; published online 24 November 2016.

Case of the Crisis in Confidence
Basic Terminology

- ***Population*** – The entire collection of experimental units
  - In pharmacometrics the experimental unit is typically the individual subject (e.g., healthy volunteer or patient)
- ***Sample*** – A subset of experimental units from the population
  - The sample should be random for proper statistical inference
- ***Parameter*** – A fixed number that represents some distributional characteristic of the population
  - E.g., the population mean, population percentile (median, 10^{th} and 90^{th})
- ***Statistic*** – An estimate of the parameter from a random sample
  - A statistic is a random variable (e.g., sample percentile) that varies from sample to sample (i.e., sampling variation)

Case of the Crisis in Confidence

Statistical Intervals

- ***Confidence Interval (CI)*** – Used to make inference about a parameter (e.g., population mean or percentile)
  - Reflects uncertainty in the parameter
- ***Prediction Interval (PI)*** – Used to make inference about a future value of a statistic (e.g., sample mean or percentile)
  - Reflects uncertainty in both the parameter as well as the sampling variation for the statistic

Note: Statistical intervals make inference about repeated experiments used to quantify the uncertainty in the estimate of the population parameter (CI) or future value of a sample statistic (PI).

Case of the Crisis in Confidence
Parameter Uncertainty

- ***Parameter Uncertainty*** – Can be thought of as the trial-to-trial variation in the parameter estimates
  - If we were to fit the same model to different sets of trial data based on the same trial design we would get different estimates of the parameters for each trial

Note: To make proper inference for repeated experiments (trials) the statistical intervals must also reflect parameter uncertainty (i.e., trial-to-trial variation).

Case of the Crisis Confidence
Law of Large Numbers

- As the sample size (N) increases toward infinity, the sample statistic will converge to its corresponding population parameter
  - E.g., as N $\rightarrow\infty$ the sample mean ($\bar{X}$) converges to the population mean (μ)
    - $\mu = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} X_i$

Case of the Crisis in Confidence
Stochastic Simulations to Predict Parameters and Statistics

- Suppose we simulate responses for say **N=100** subjects for each of K=1000 simulated trials from the final model parameter estimates
  - For each trial we calculate the sample mean or median across the N=100 subjects
  - The sample mean or median would vary from trial to trial reflecting the sampling variation
- Now suppose we simulate responses for say **N=10,000** subjects for each of K=1000 simulated trials from the final model parameter estimates
  - For each trial we calculate the sample mean or median across the N=10,000 subjects
  - For each trial the sample mean or median should converge to the same value (i.e., the population parameter) assuming N=10,000 subjects is sufficient for the Law of Large Numbers to hold

Case of the Crisis in Confidence

Stochastic Simulations for VPCs

- Stochastic simulations for VPCs use the observed sample sizes in the dataset used to develop the model
- These observed sample sizes are typically too small for the statistic to converge to the population parameter
- For an internal VPC where the final parameter estimates are used for each simulated trial, the interval formed by the 5^{th} and 95^{th} percentiles across the K=1000 simulated trials reflect the sampling variation of the statistic (median or 10^{th} and 90^{th} percentiles)
  - Note the resulting interval cannot be a confidence interval since it reflects sampling variation but not parameter uncertainty (since it uses the same parameter estimates for each simulated trial)
  - Note the resulting interval would be a valid prediction interval if it included parameter uncertainty

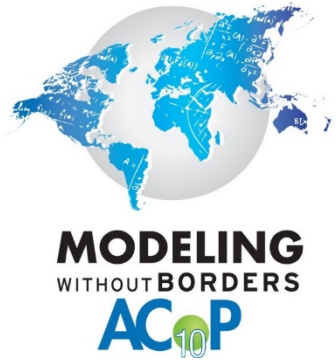
Gerald Hahn (1970). Statistical Intervals for a Normal Population, Part I: Tables, Examples, and Applications. *Journal of Quality Technology*, 2:115 - 125.

Gerald Hahn (1970). Statistical Intervals for a Normal Population, Part II: Formulas, Assumptions, Some Derivations. Journal of Quality Technology, 2:195 – 206.

Integration of Pharmacometric and Statistical Analyses Using Clinical Trial Simulations to Enhance Quantitative Decision Making in Clinical Drug Development – October 24, 2019 (1 – 5 pm)

- Why should we integrate pharmacometric and statistical analyses?
- Learn how to formulate quantitative decision rules using confidence interval criteria.
- **Understand the distinction between confidence intervals and prediction intervals and how to perform stochastic simulation procedures to construct such statistical intervals.**
- Learn how to apply clinical trial simulation procedures to evaluate various probability metrics to support study design recommendations and quantitative decision-making.


GUILTY



THE CASE OF THE PHARMACOMETRICIAN's PITIFUL PLANNING

FRANCE MENTRE

PLAINTIFF

JIN JIN

DEFENDANT

ACCUSED OF:

DREADFUL DESIGN

Table: Phase 3 Population PK analysis of drug FIM

	Basic Model		Gene covariate model		
	Estimate	rse%	Estimate	rse%	p-value
ka (h^{-1})	71	281%	145	425%	
CL (L/h)	1.1	18%	1.3	19%	
β_CL_gene			-49%	55%	0.07
V (L)	19.3	27%	17.1	20%	
ω_ka	155%	129%	191%	60%	
ω_CL	109%	14%	106%	13%	
ω_V	78%	70%	53%	43%	
σ_add (mg/L)	1.3	39%	1.4	34%	
σ_prop	23%	24%	23%	22%	

Study design and assumptions

- Phase 3 trial for drug FIM for Seurat's disease
- 100 patients included in the treatment group
  - 50 patients were included in the pop PK analysis
- 100 mg BID
- 2 samples measured at day 10
  - ½ patients 0 (trough), 1h ; ½ patients 0 (trough), 4h
- An objective of the popPK analysis was to estimate the effect of slow metabolizers on clearance
  - Expected proportion of slow metabolizers (mutant gene): 20%
  - Expected effect: 50% decrease of CL
- The expected interpatient variability on CL is very large 100%
  - From a previous population PK analysis in Phase 2

- Using any design software you could have evaluated the following 4 designs
  - N = 100 vs 50
  - Sampling design: n= 3 (0, 1, 4) vs n=2 (half (0,1) ; half (0,4))

N	n	rse CL	rse V	rse ω_CL	rse ω_V	Power covariate	NSN 80%
100	3	12%	18%	16%	45%	77%	108
100	2	12%	27%	27%	250%	76%	111
50	3	17%	25%	22%	63%	48%	108
50	2	17%	38%	38%	354%	47%	111

➢ A better design would have been 3 samples in all 100 patients to estimate adequately the parameters and have a power of 77% of showing the covariate effect

- The primary objective of the Phase 3 trial is to evaluate efficacy and safety in the target patient population.
- PK is a secondary objective.

- **Number of samples per subject:**
  - Patients with Seurat disease suffer from pain and immobility, and PK sampling is a painful procedure for them
  - 2 sampling per visit was maximally feasible
  - Unnecessary patient burden and inconvenience may lead to slow trial recruitment and higher risk of patient drop-out.

- **Number of subjects:**
  - Study protocol intended to collect PK data in all patients
  - 50 of the 100 patients did not have any measured drug concentration
    - 10 dropped off before D10
    - 20 did not have PK due to unsuccessful sampling
    - 20 had damaged PK samples
  - Many clinical sites in multiple countries inexperienced with PK sampling

- 4 designs
  - N = 100 vs 50
  - Sampling design: n= 3 (0, 1, 4) vs n=2 (half (0,1) ; half (0,4))

N	n	rse CL	rse V	rse ω_CL	rse ω_V	Power covariate	NSN 80%
100	3	12%	18%	16%	45%	77%	108
100	2	12%	27%	27%	250%	76%	111
50	3	17%	25%	22%	63%	48%	108
50	2	17%	38%	38%	354%	47%	111


GUILTY


NOT
GUILTY

MODELING WITHOUT BORDERS
ACoP 10

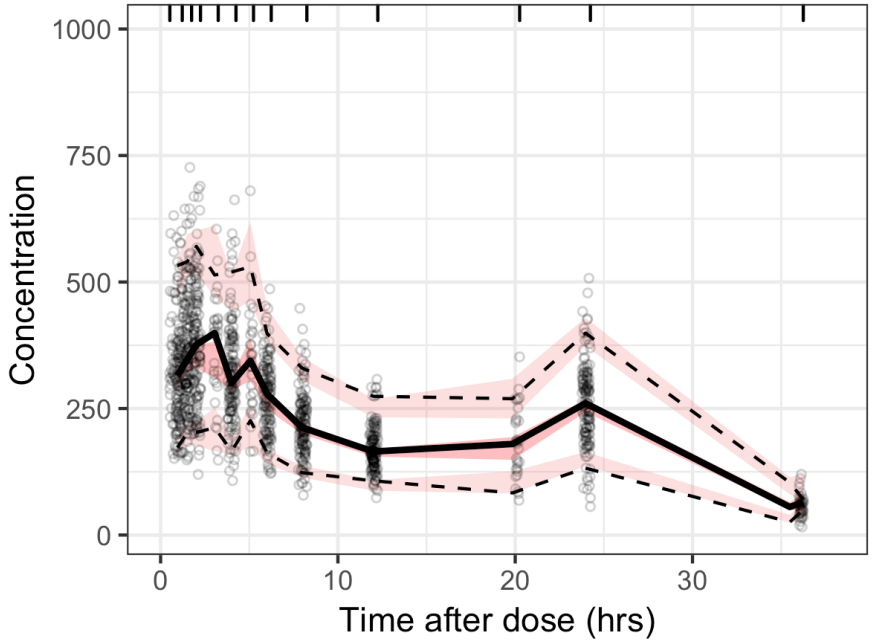
THE CASE OF THE EXTRAPOLATED PREDICTION

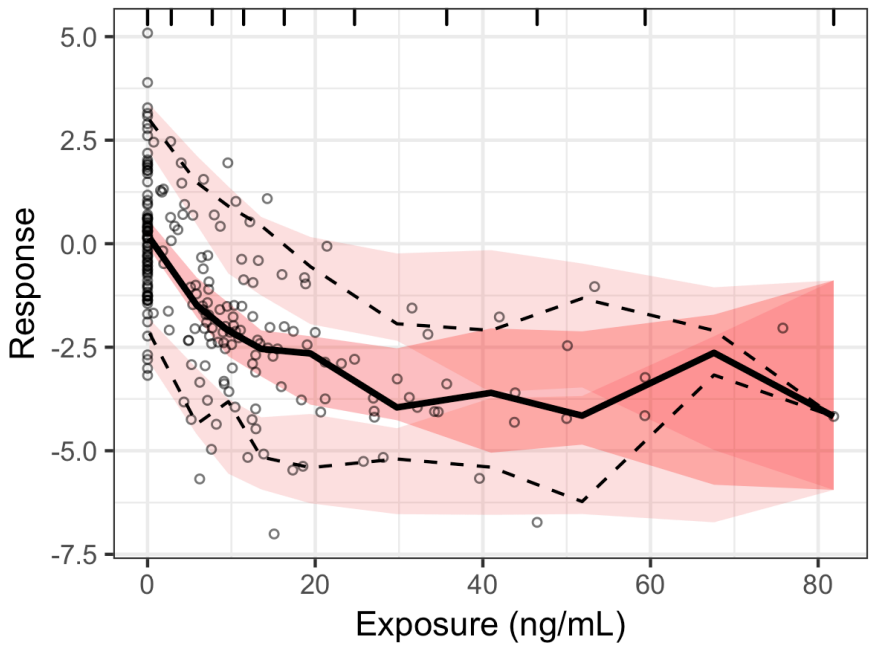
LEI NIE
PLAINTIFF

CHAO LIU
DEFENDANT

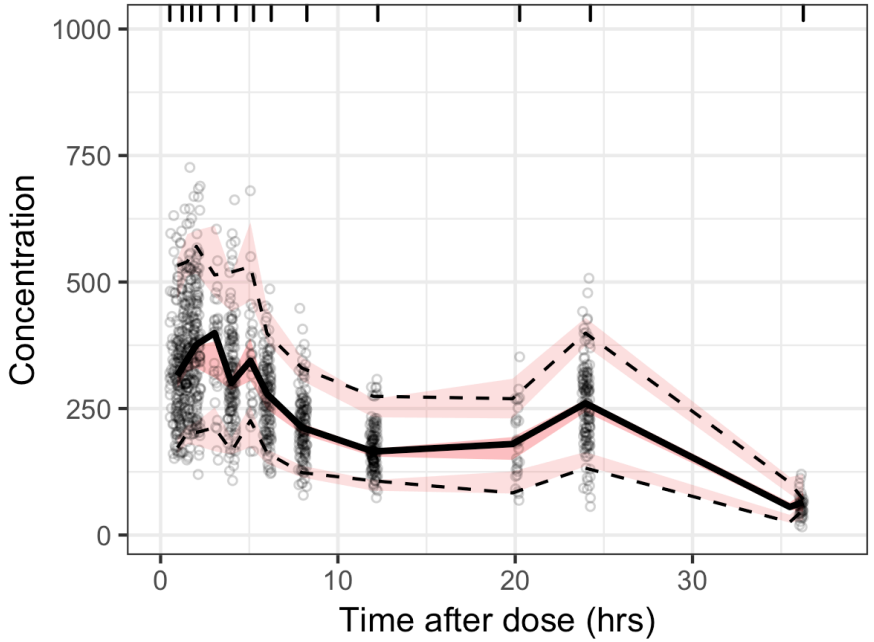
ACCUSED OF:
EGREGIOUS EXTRAPOLATION

FORMULATION X

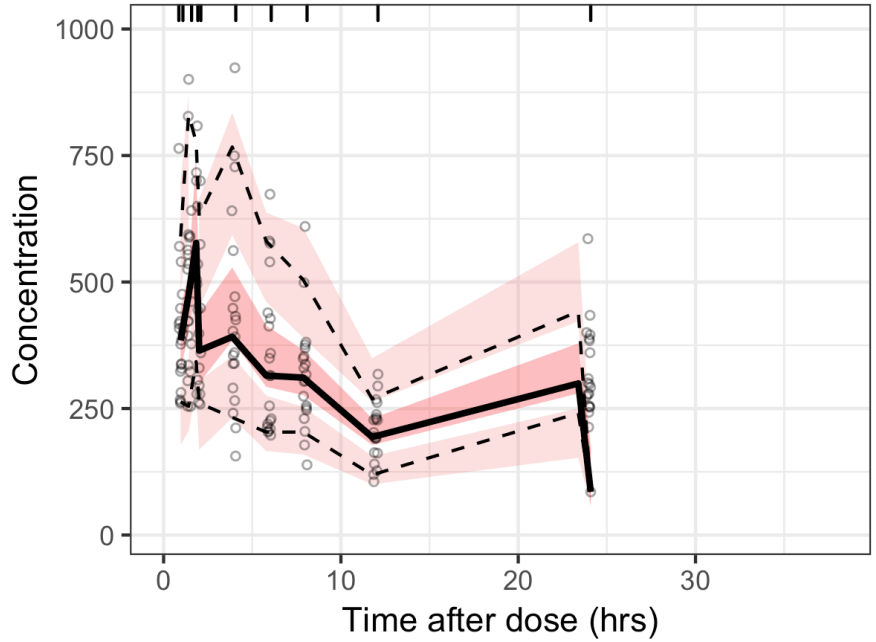


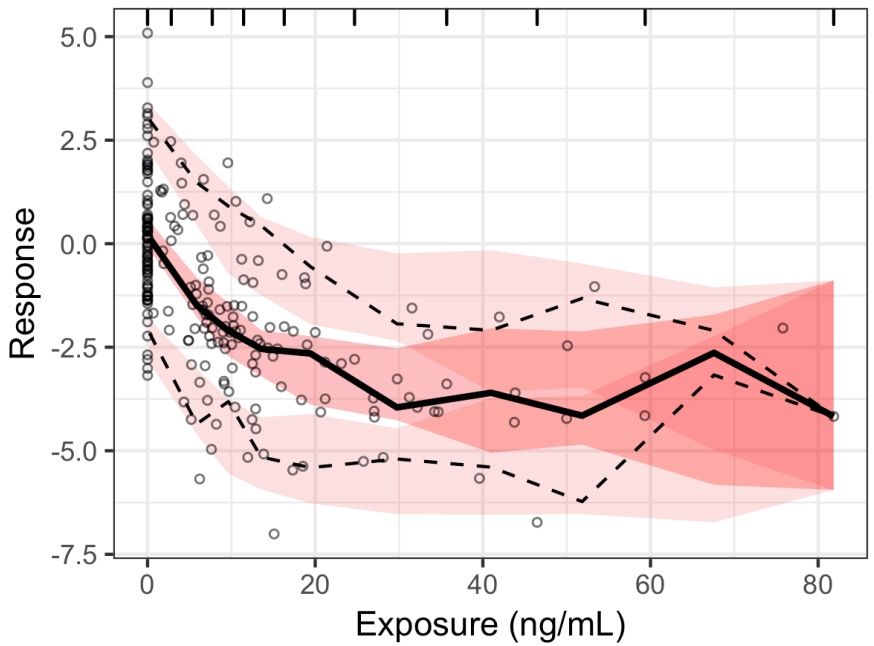


FORMULATION X



FORMULATION Y





?

Case of the Extrapolated Prediction
Empirical Models and Extrapolation

Hahn, J. (1977). Hazards of extrapolation in regression analysis. *JQT* 9:159-165.

> *"Extrapolation of a fitted regression equation beyond the range of the given data can lead to seriously biased estimates if the assumed relationship does not hold in the region of extrapolation."*
>
> *"…extrapolation cannot be supported based on statistical grounds alone…"*

Case of the Extrapolated Prediction
Mechanistic Models and Extrapolation

Box, G.E.P., Hunter, W.G., and Hunter, J.S. (1978). *Statistics for Experimenters*, Wiley, NY.

"Mechanistic models can provide a basis for extrapolation."

"…a well-tested mechanistic model does more than just graduate the data. It confirms our scientific understanding of the system…"

"This better basis for extrapolation is provided because it is the mechanism not a mere empirical curve that is being applied…and this mechanism is based on a partially verified understanding of the system itself.

"…the mechanism may change, so unchecked extrapolation is never safe. Thus, even a mechanistic model should preferably be used only to suggest regions where further experimentation might be fruitful."

Case of the Extrapolated Prediction
Learning Trials

Sheiner, L.B. (1997). Learning versus confirming in clinical drug development. *CPT* 61:275 – 291.

> *"Reliable assumptions about the form of the relationship between regimen, prognostic factors, and outcomes are therefore needed to interpolate and extrapolate between and beyond the isolated points that are studied, and these can only come from previous scientific knowledge."*

> *"Under an enlightened drug development plan, the "hypotheses" that learning trials generate will be tested in later more rigorous confirmatory trials."*


CASE
DISMISSED

MODELING WITHOUT BORDERS
AC10P

CASE OF THE NOT SO CONFIRMATORY CONCLUSION

BRIAN SMITH
PLAINTIFF

NAG CHEMUTURI

DEFENDANT

ACCUSED OF:

SNEEZING AT SIGNIFICANCE

100 Examinations Each With 80% Power and 5% Significance Level

True Probability		Out of 100	Significance Level / Power	Number of Claims	False Claims
5%	No Effect	95	.05	4.75	4.75/8.75 = 54%
	Effect	5	.8	4	
50%	No Effect	50	.05	2.5	2.5/42.5 = 5.9%
	Effect	50	.8	40	
80%	No Effect	20	.05	1	1/65 = 1.5%
	Effect	80	.8	64	

Case of the Not So Confirmatory Conclusion
Learning Versus Confirming

- ***Learning*** – The focus is on estimation and prediction

  "…analysis of a learning trial, estimating the response surface requires that scant observations at many points on the surface be somehow linked to yield a coherent picture." Sheiner (1997)

- ***Confirming*** – The focus is on hypothesis testing

  "…for the simplest confirmatory designs and sharpest null hypotheses, virtually no assumptions whatever are needed. No (unproved) a priori assumptions means unequivocal conclusions. This is the great strength of a well-designed and executed confirmatory study: when the null hypothesis is rejected, the meaning is clear and unequivocal." Sheiner (1997)

Sheiner, L.B. (1997). Learning versus confirming in clinical drug development. *Clin Pharmacol Ther* 61:275 – 291.

Case of the Not So Confirmatory Conclusion
Covariate Model Building

- Covariate model building is essentially a **learning** activity and is exploratory in nature
  - The focus is on estimation and prediction and how patients/subjects across a wide range of demographic and prognostic factors are related
  - The primary goal of stepwise covariate modeling (SCM) is to obtain a final parsimonious model with reduced prediction error relative to other covariate models that could be considered
- Covariate model building procedures (such as SCM) do not lend themselves easily to confirmatory conclusions of statistical significance

Case of the Not So Confirmatory Conclusion

Statistical Significance

- Statistical significance is inherently a confirmatory statement
  - A p-value associated with a statistical hypothesis test is a measure of the strength of the evidence for the effect
- The validity of a p-value requires prespecification of the model(s) and hypothesis test(s) to prevent bias

  Edwards, D. (1999). On model prespecification in confirmatory randomized studies. *Stat in Med* 18:771 – 785.
- Multiplicity of testing further complicates the assessment of statistical significance
  - A few "statistically significant" covariate effects from 50+ likelihood ratio tests (LRTs) that achieve a certain magnitude of the LRT statistic do not convey the same strength of evidence as a single hypothesis test from a prespecified model that achieves this same magnitude of the LRT statistic
  - The false positive rate increases with the number of hypothesis tests performed

  Li,G. et. al. (2017). An introduction to multiplicity issues in clinical trials: the what, why, when and how. Int J of Epidemiology 46:746 – 755.

Case of the Not So Confirmatory Conclusion

Full Covariate Modeling Approach

- A full covariate modeling (FCM) approach may be more appropriate for dealing with multiplicity of testing:

  "The FPR (false positive rate) for the FCM approach dramatically increases with the number of covariates. The chance of incorrectly selecting ≥1 seemingly clinically relevant covariates can be increased from 5% to a 40 – 70% range for 10 – 20 covariates." Xu et. al. (2018)

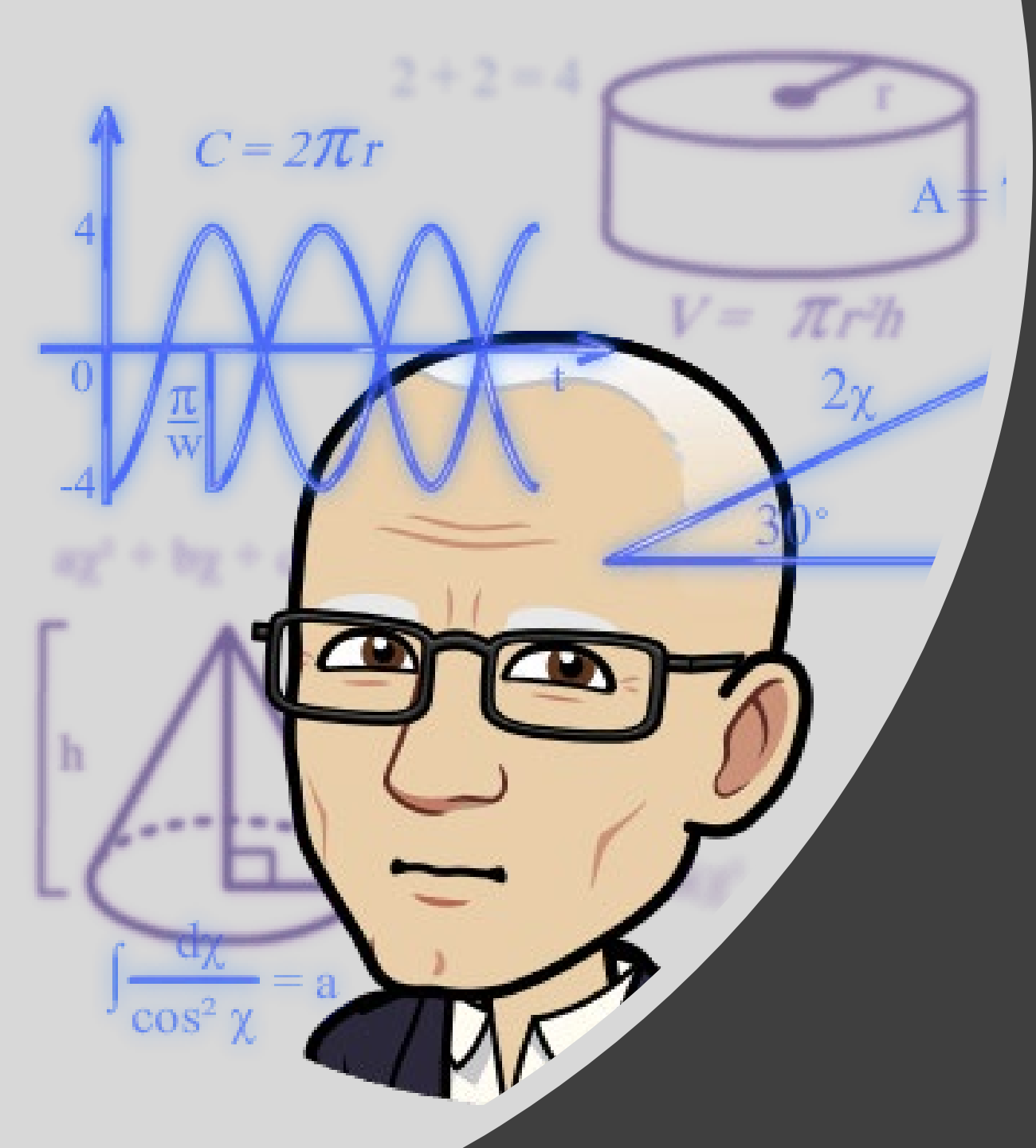
  "The SCI (simultaneous confidence intervals) approach may provide appropriate control of the family-wise FPR…at 5% or 10%..." Xu et. al. (2018)

  Xu, X.S., Yuan, M., Zhu, H., Yang, Y., Wang, H., Zhou, H., Xu, J., Zhang, L., and Pinheiro, J. (2018). Full covariate modelling approach in population pharmacokinetics: understanding the underlying hypothesis tests and implications of multiplicity. Br J of Clin Pharmacol 84:1525 – 1534.

Kowalski, K.G. (2015). My career as a pharmacometrician and commentary on the overlap between statistics and pharmacometrics in drug development. *Stat in Biopharm* Res 7:148 – 159.

Kowalski, K.G. (2019). Integration of pharmacometric and statistical analyses using clinical trial simulations to enhance quantitative decision making in clinical drug development. *Stat in Biopharm* Res 11:85 – 103.


GUILTY

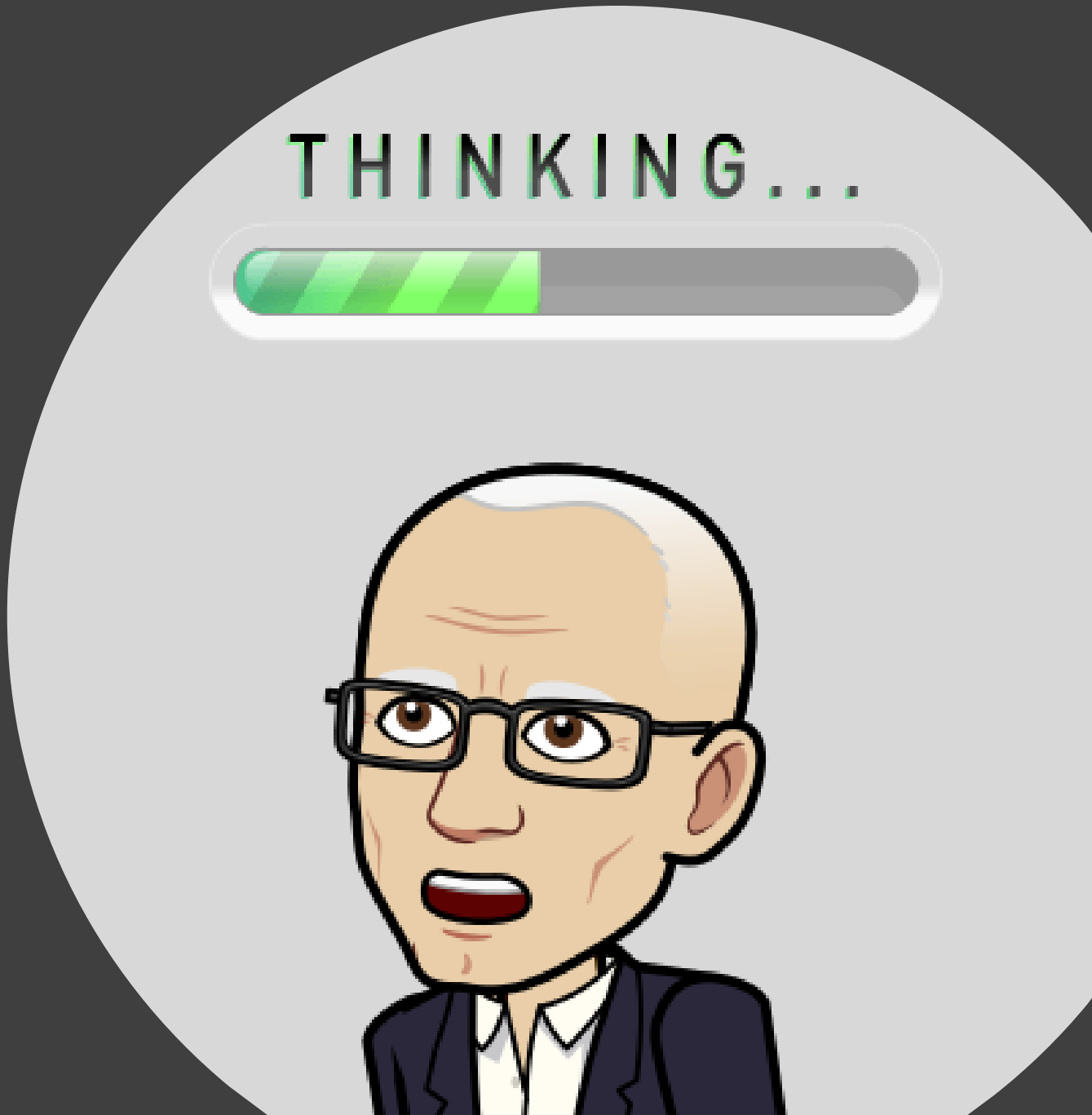


The key is that all analyses require **sound scientific judgement.**

The results of the analysis have to be put into context.

It is the results of the analysis ***along with scientific judgement*** that leads to conclusions that are meaningful. This in part, defines the discipline of Pharmacometrics

I hereby rule Judge Kowalski's ruling……



OVERRULED!

GUILTY!


<0.05
THE
PEOPLE'S
COURT

CAST

(various roles)	JONATHAN FRENCH
(various roles)	STACEY TANNENBAUM
Judge Ken Kowalski	KEN KOWALSKI
Jim Rogers	JIM ROGERS
Eric Jordie	ERIC JORDIE
France Mentre	FRANCE MENTRE
Jin Jin	JIN JIN
Jury Foreman	MATT ZIERHUT
Lei Nei	LEI NIE
Chao Liu	CHAO LIU
Brian Smith	BRIAN SMITH
Nag Chemuturi	NAG CHEMUTURI
Supreme Court Justice	BRIAN CORRIGAN

A production of the

Statistics and Pharmacometrics Interest Group (SxP)