Random Effects Model Analysis
to
Test for Instrument Swap Effect

Data Collection: Suzanne Doughten, Heather Wright, ODU

Data Analysis: Mike Lane, ODU

Presenter: Elgin Perry, eperry@chesapeake.net

Presented to: **DIWG, 2/26/2025**

Road Map

1. Review of Instrument Swap Comparison Study Methods (Heather Wright)
2. Description of Mixed Model Statistical Method
3. Review of Results for NH3F, NO23, NO2F, PO4F, SIF
4. Conclusions
5. Discussion of implications for future studies

Traditional Model for Instrument Comparison Study

$$y_{ij} = \mu + \beta_i + \epsilon_{ij} \quad (1)$$

Where: y is concentration, μ is overall mean, β_i is instrument effect $i = 1, 2$, and ϵ_{ij} is independent and identically distributed (iid) random error.

Because data are paired, we might simplify by taking differences:

$$y_{2j} - y_{1j} = d_j = \gamma + \varepsilon_j \quad (2)$$

Where: $\gamma = \beta_2 - \beta_1$, and $\varepsilon_j = \epsilon_{2j} - \epsilon_{1j}$

Statistical comparison made using Paired t-test or Wilcoxon Signed Rank Test.

Random Effects Model for Instrument Comparison Study

$$y_{ijk} = \mu + \beta_i + \delta_{ij} + \epsilon_{ijk} \quad (3)$$

Where: y, μ, β are as before,

Now the random effect is split into two parts $\delta_{ij} + \epsilon_{ijk}$

δ is a random effect caused by calibration

ϵ is a random effect caused by sample to sample variance

Note that all of the y_{ijk} with the same ij have the same calibration effect causing them to be correlated. This is a violation of the iid assumption of the Traditional model.

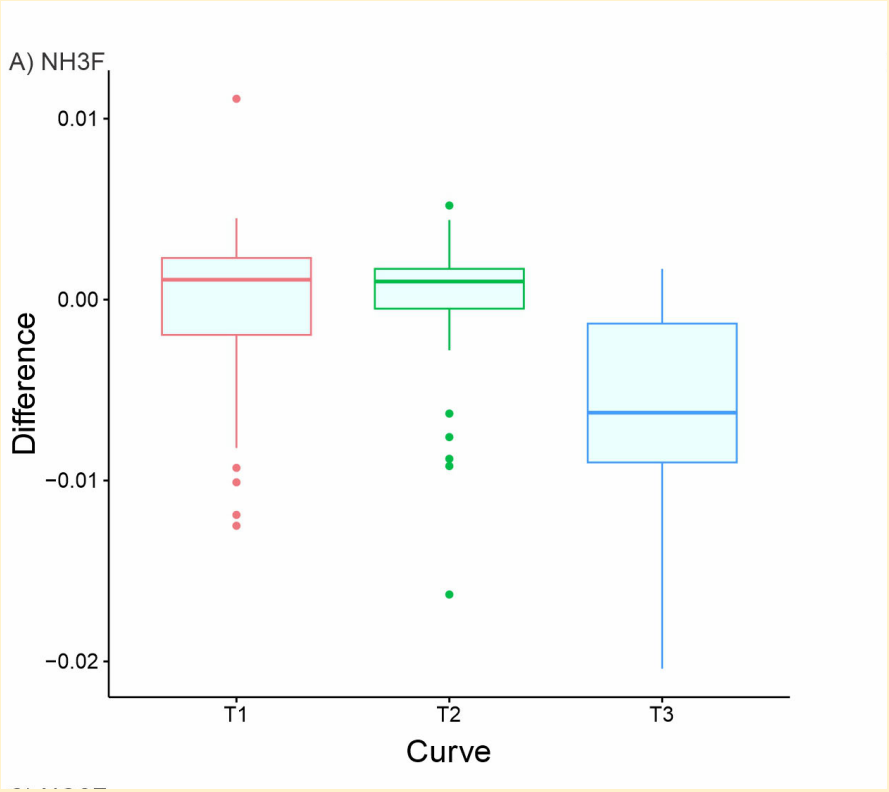
The is a difference version of this model as well

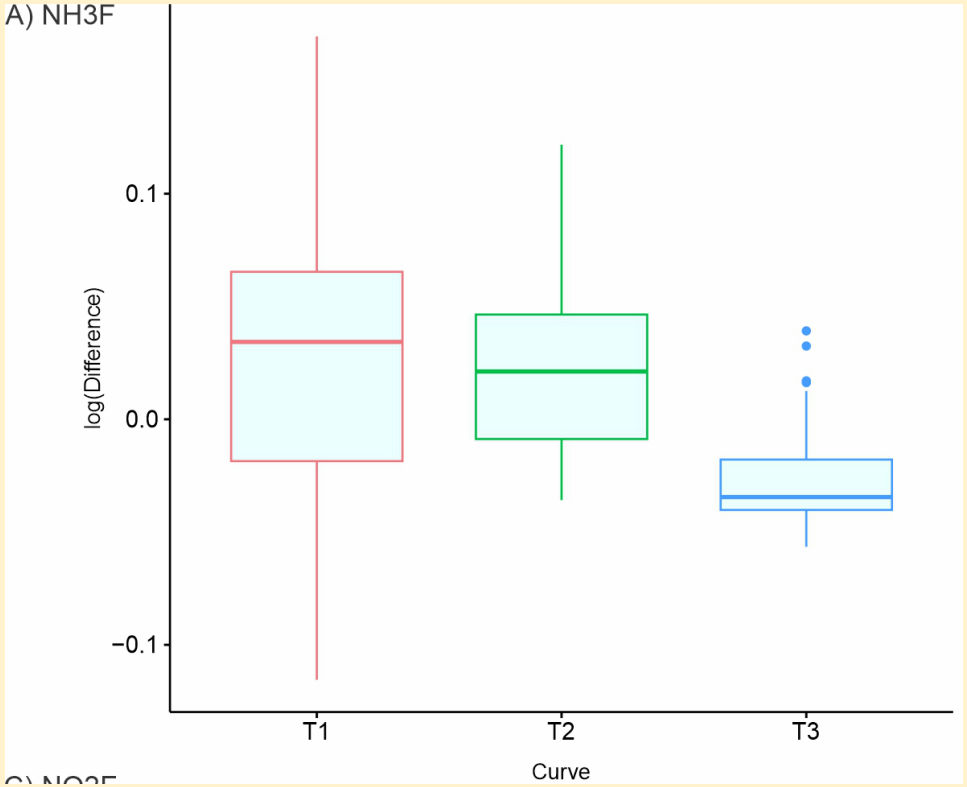
$$y_{2jk} - y_{1jk} = d_{jk} = \gamma + \rho_j + \varepsilon_{jk} \quad (4)$$

Where: $\gamma = \beta_2 - \beta_1$, $\rho_j = \delta_{2j} - \delta_{1j}$ **and** $\varepsilon_{jk} = \epsilon_{2jk} - \epsilon_{1jk}$

In the analyses that follow, the difference is (Lachat (old) – Seal (new))

Is the Calibration Effect a real thing?
Example: NH3F

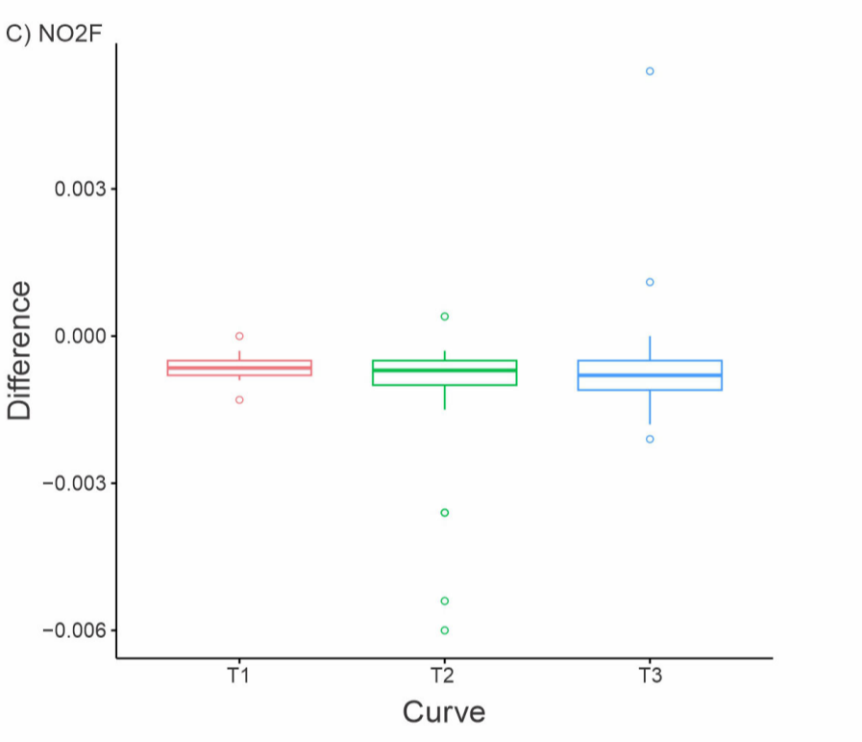


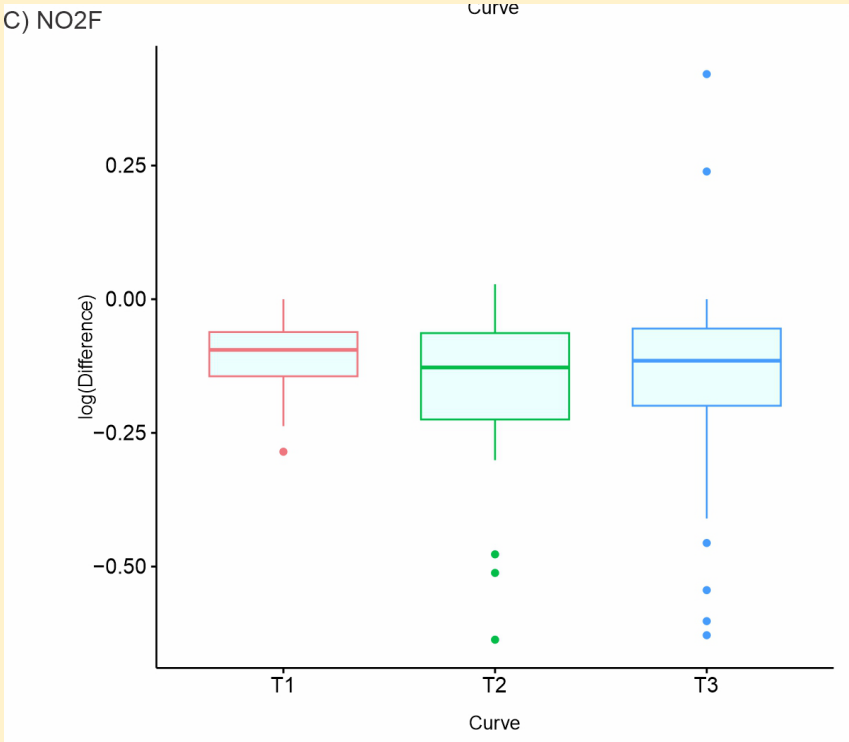


It does seem that the Calibration effect is real.

Shows the importance of averaging over several Calibrations.

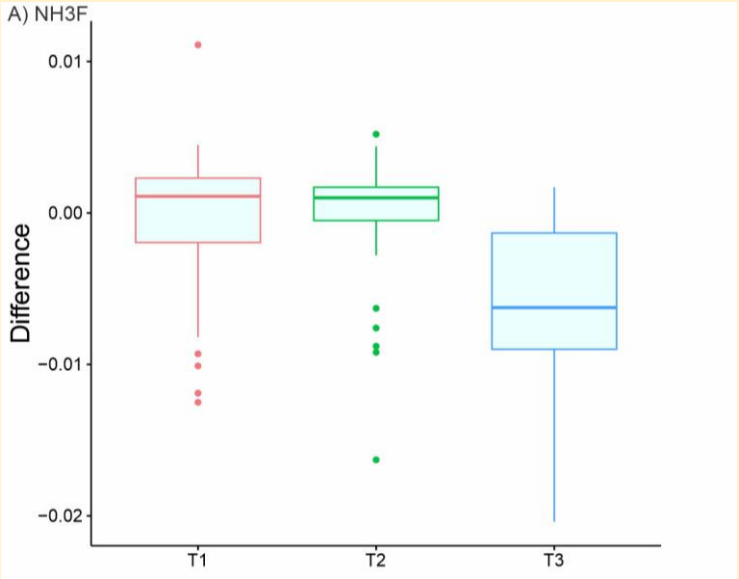
Another Example: NO2F

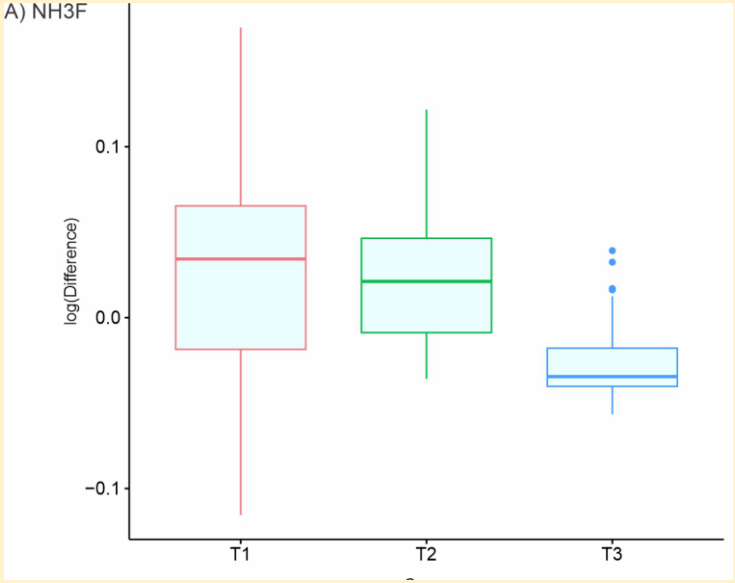




Rule of thumb from literature is to have at least 5 levels of random effect.

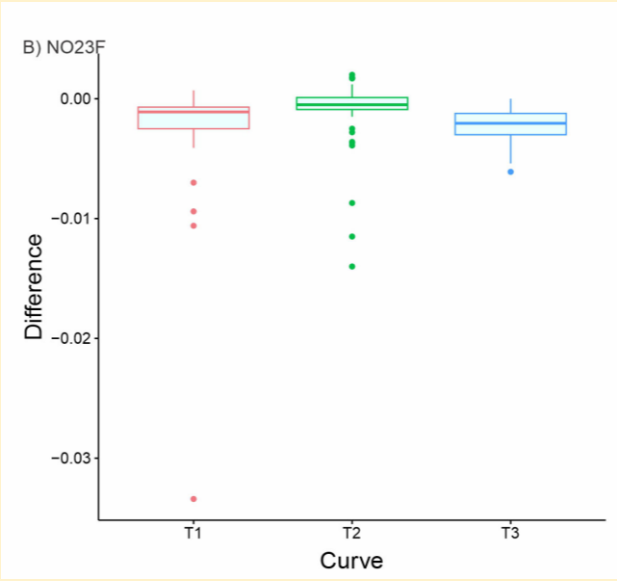
Results: NH3F

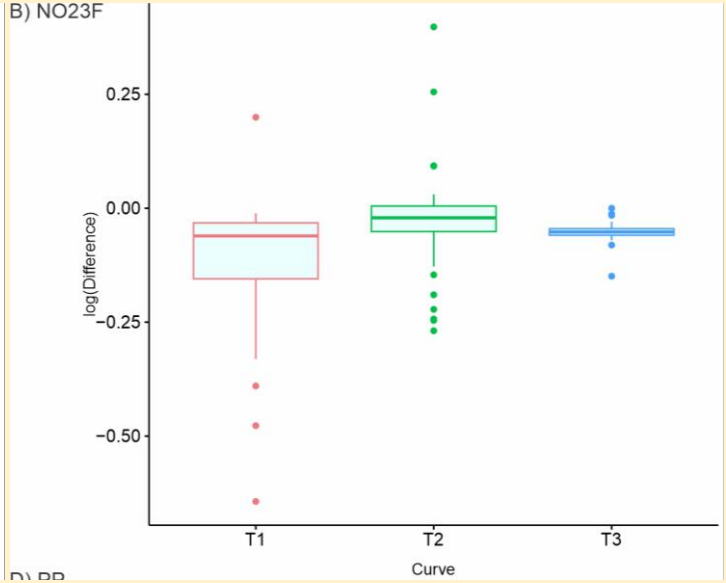




NH3F	Effect Source	Effect Size	p-value
	Calibration	0.0242	<0.0001
	Instrument	0.009445	0.559

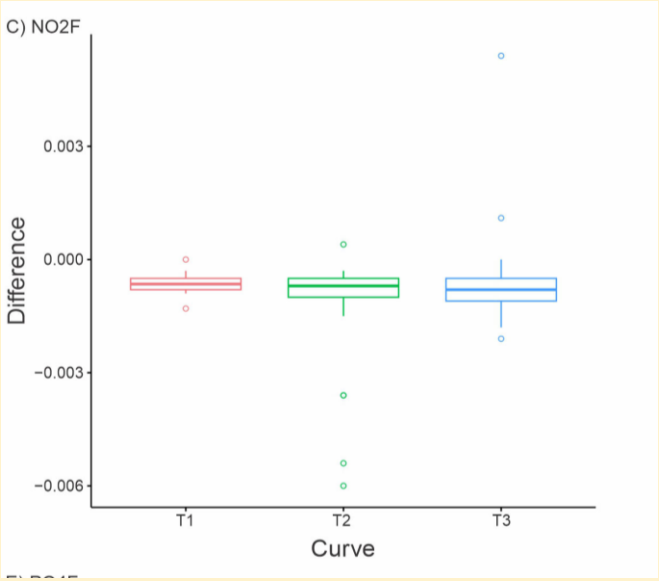
NO23F

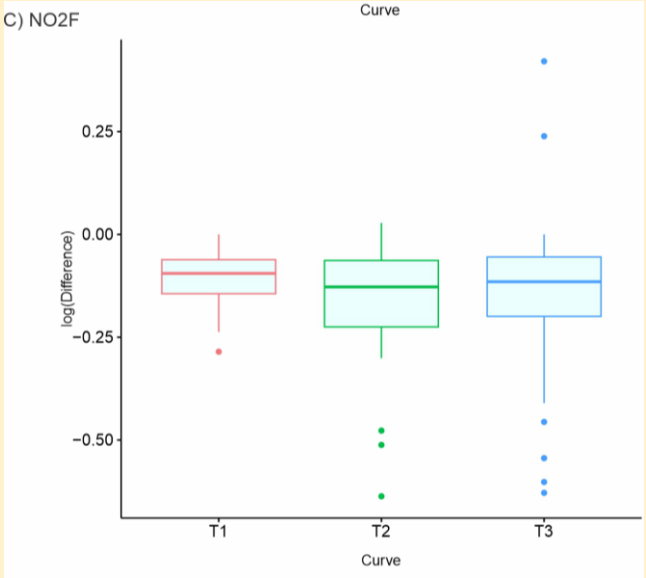




NO23F	Effect Source	Effect Size	p-value
	Calibration	0.0313	0.0029
	Instrument	-0.06419	0.0475

NO2F

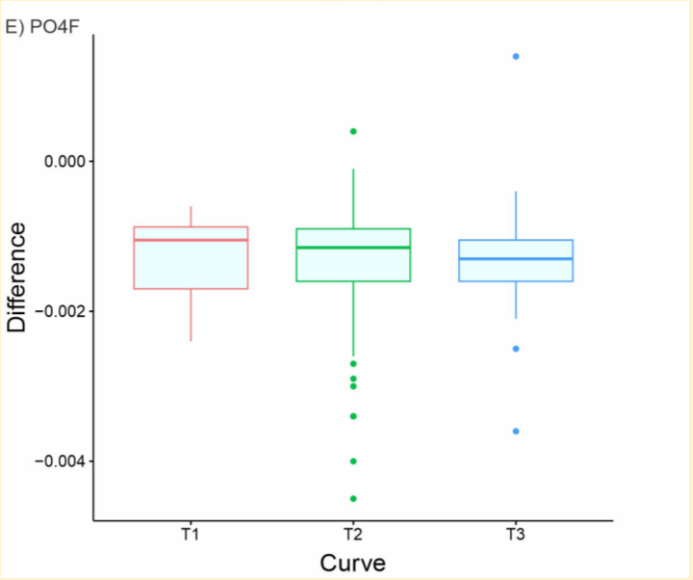


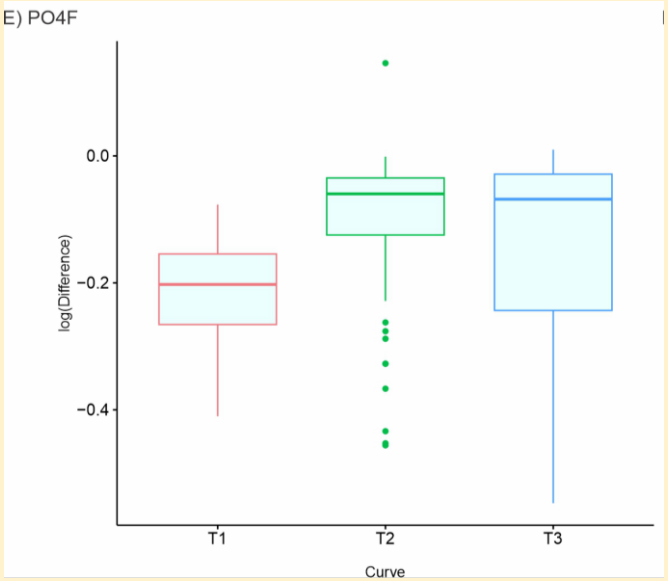


NO2F	Effect Source	Effect Size	p-value
	Calibration	0	1
	Instrument	-0.14873	<0.0001

analysis failed: boundary (singular) fit

PO4F

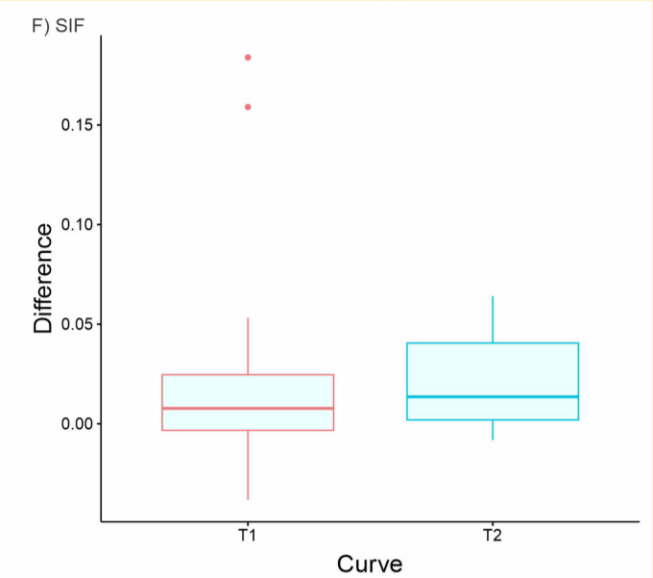


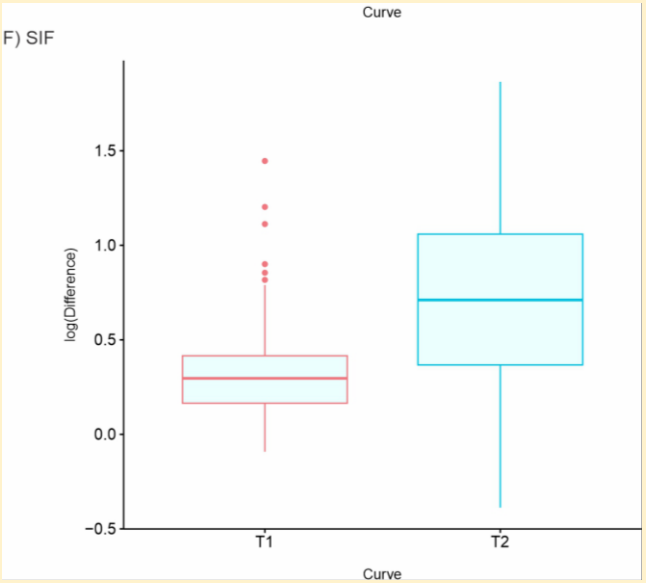


PO4F	Effect Source	Effect Size	p-value
	Calibration	0	1
	Instrument	-0.0013	<0.0001

analysis failed: “boundary (singular) fit:”

SIF





SIF	Effect Source	Effect Size	p-value
	Calibration	0.1897	0.0002
	Instrument	0.575	0.0548

Conclusions:

1. **Calibration effects seem real**
2. **Averaging over calibrations is needed to get an unbiased instrument comparison**
3. **At least 5 calibrations are recommended (a lot of work).**

Discussion:

From Suzanne Doughten:

> **“I think one thing that is getting lost in the DIW group is that these comparison studies are for the CBP data analysist so that if there are trends in the data they can decipher if they are due to an instrument change or environmental conditions. We are not trying to prove the instrumentation is valid, the manufacturer has already done that. I don't know if we want to bring up why we do these studies.”**