



BivRegBLS: A one nearly two year- anniversary

R package for **method comparisons**
based on **errors-in-variables**
regressions and **tolerance intervals**

Marion BERGER 

Bernard FRANCO 



20 mins to show...

- **Real Case studies**

- Using **Bivariate Least Squares (BLS)** / **Correlated-BLS (CBLS)** regressions
 - To compare **X and Y variables** when both **have variabilities**
 - Generalization of the Deming regression in (X, Y) and (Mean, Difference) scales

- **Outside of this presentation, we can**

- Convince you that BLS/CBLS regressions are the **best choices**
 - Notably it overcomes pitfalls of Bland-Altman approach
 - See Bernard Francq's thesis and articles, discuss with him
- Show all the features of the package

This is real life



- **Examples of methods comparison from**

Contributors 

CMC

Delphine Attonaty, Hugo Zuin

Microbiology

Hugo Zuin

Bioanalysis

Karine Imberdis

Toxicology

Fanny Windenberger

- **And ... relationship (CMC)**

Hajer Souaifi-Amara

CMC 1 : Mass spectrometry versus UV

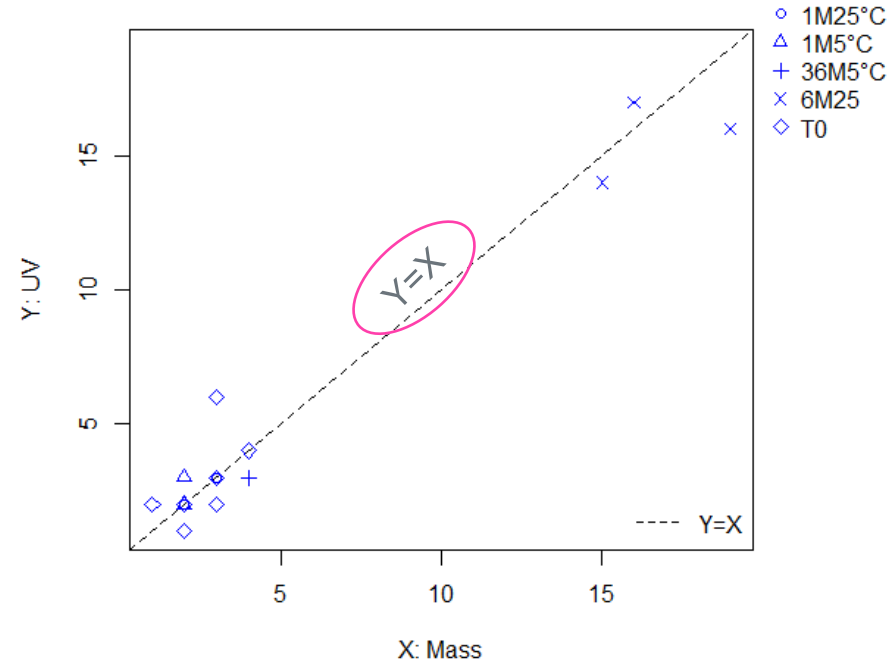
19 samples (batches/timepoints/temperatures combinations)

No replicates

Normal data for peptide measurement

Assess equivalence area

Data plot according to combinations



CMC 1 : Mass spectrometry versus UV

Estimated Variances from other studies

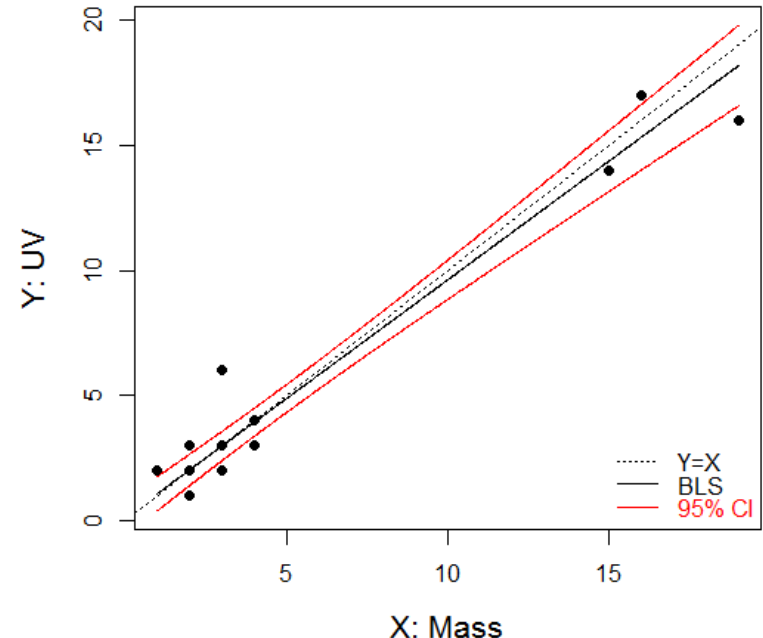
- Mass: method qualification - 2 operators, 6 replicates/operator → $\text{between} + \text{within-variability} = 0.40 + 0.02 = 0.42$
- UV: 11 samples from reference standard → only $\text{within-variability} = 0.07$

```
BLS=BLS(data=MUV,xcol=4,ycol=3,var.x=0.42,var.y=0.07)
```

```
XY.plot(BLS.results = BLS, xname="X: Mass", yname="Y: UV",  
graph.int = c("CI"))
```

CI: Confidence Intervals

BLS Regression



CMC 1 : Mass spectrometry versus UV

Estimated Variances from other studies

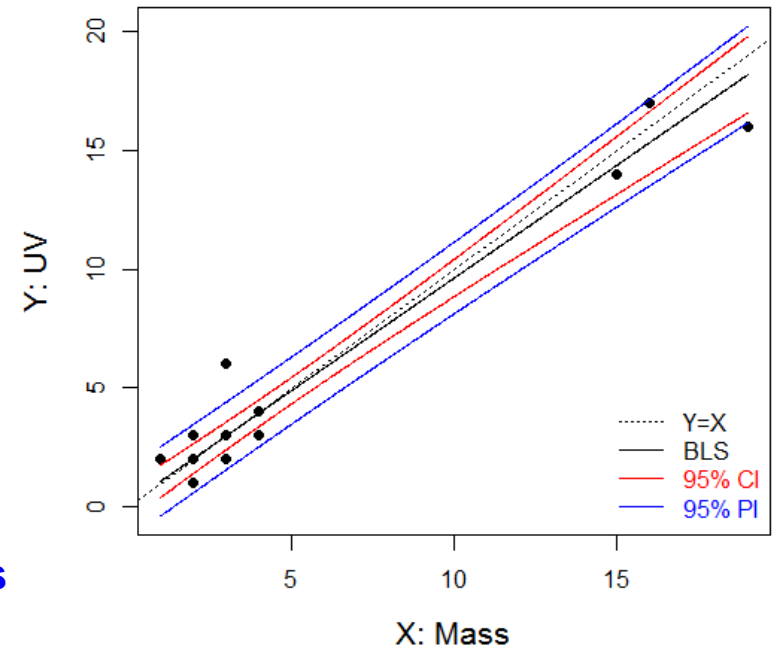
- Mass: method qualification - 2 operators, 6 replicates/operator → within+between-variability=0.02+0.40=0.42
- UV: 11 samples from reference standard → only within-variability = 0.07

BLS=BLS(data=MUV,xcol=4, ycol=3, var.x=0.42, var.y=0.07)

XY.plot (BLS.results = BLS, xname="X: Mass", yname="Y: UV",
graph.int = c("CI","PI"))

CI: Confidence Intervals **PI: Prediction intervals**

BLS Regression



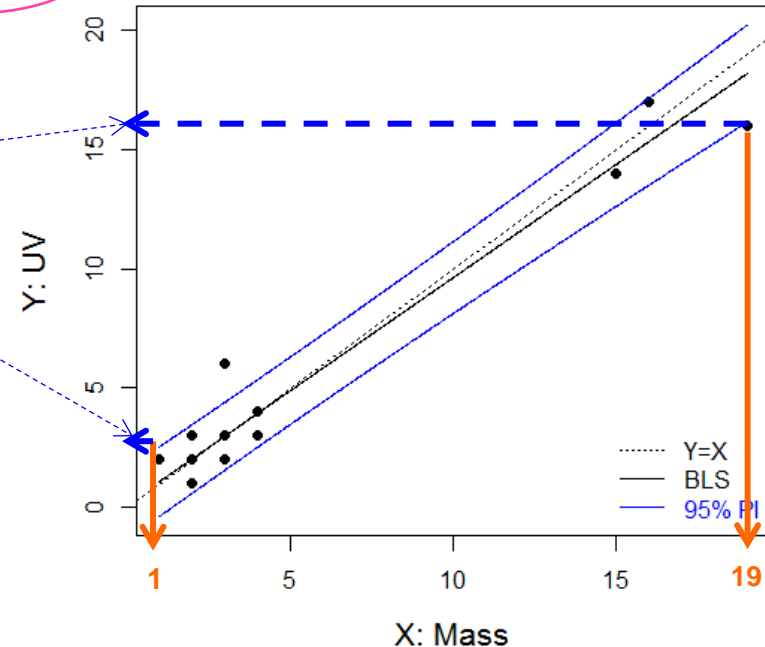
CMC 1 : Mass spectrometry versus UV

Assess equivalence area

BLS=BLS(data=MUV,xcol=4, ycol=3, var.x=0.42, var.y=0.07, xpred=c(1,19))

\$xpred.BLS	X0	Ypred	95% PI Lower	95% PI Upper
xpred 1	1	1.083752	-0.3775545	2.545159
xpred 2	19	18.173283	16.1394950	20.207071

BLS Regression



CMC 1 : Mass spectrometry versus UV

Assess equivalence area

BLS=BLS(data=MUV,xcol=4, ycol=3, var.x=0.42, var.y=0.07, xpred=c(1,19))

\$xpred.BLS			95% PI Lower	95% PI Upper
X0	Ypred			
xpred 1 1	1.083752		-0.3775545	2.545759
xpred 2 19	18.173283		16.1394950	20.207071

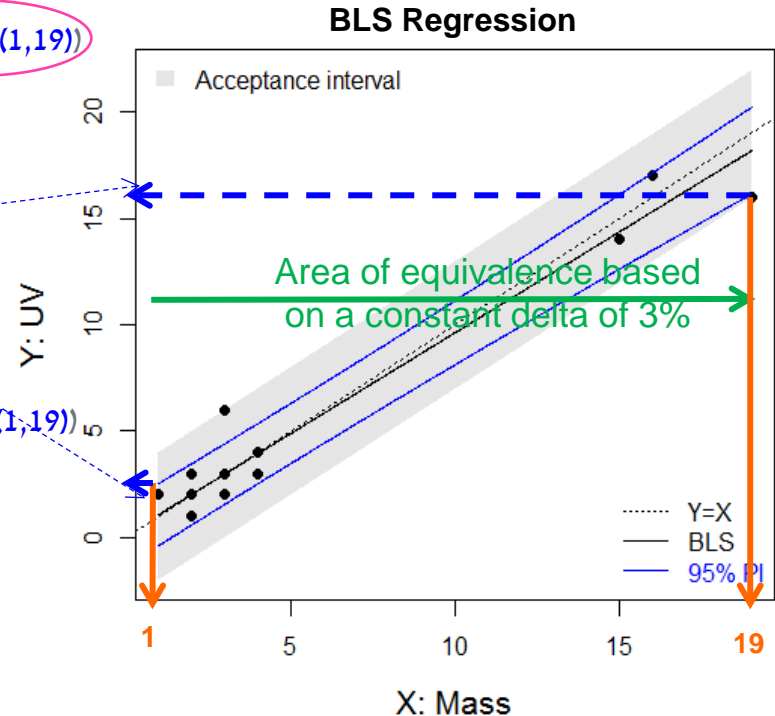
+1.5 (green arrow from 2.545759 to 16.1394950)

-3 (green arrow from 16.1394950 to 1.083752)

BLS=BLS(data=MUV,xcol=4, ycol=3, var.x=0.42, var.y=0.47, xpred=c(1,19))

\$xpred.BLS			95% PI Lower	95% PI Upper
X0	Ypred			
xpred 1 1	1.083752		-0.7572175	3.064399
xpred 2 19	18.173283		15.5312374	20.290048

-3.5 (green arrow from 15.5312374 to 1.083752)



CMC 2 : Break loose assessment

17 samples (batches/timepoints/temperatures combinations)

10 or 36 replicates with method 1

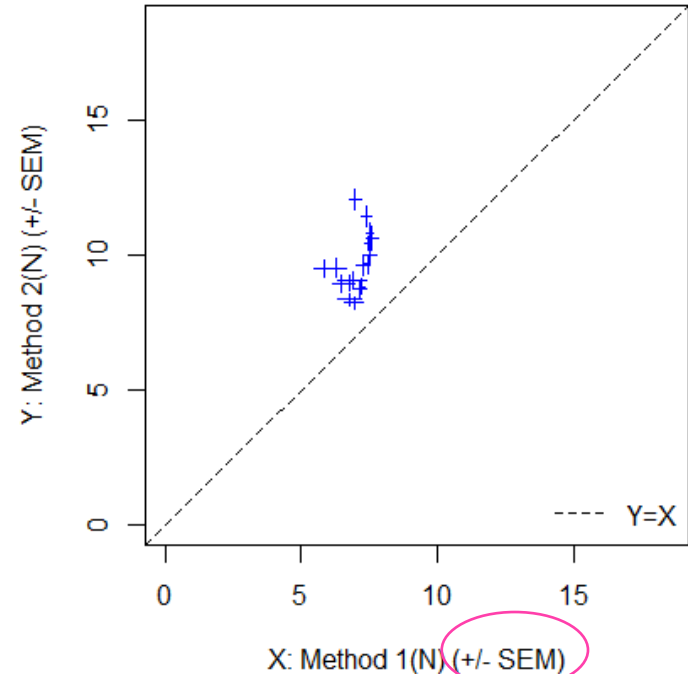
50 replicates with method 2

Normal data

Estimated Variances from replicates

Assess maximum difference between methods

Raw data plot



CMC 2 : Break loose assessment

Assess maximum difference between methods

BivBL = BLS(workBL, ycol = 4:53, xcol = 54:89, pred.level = 0.9, **qx = 10, qy = 50**)

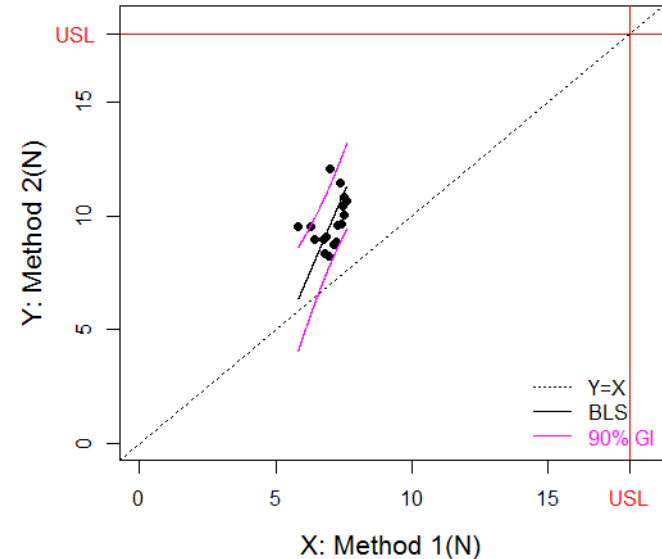
XY.plot (BivBL, xname = "X: Method 1(N)", yname = "Y: Method 2(N)",
graph.int= c("GI"))

GI: Generalized interval

Interval for the mean of q_y values from the mean of q_x values

GI is the appropriate interval here because 10 replicates in routine with method 1 and 50 with method 2

Break Loose - BLS Regression

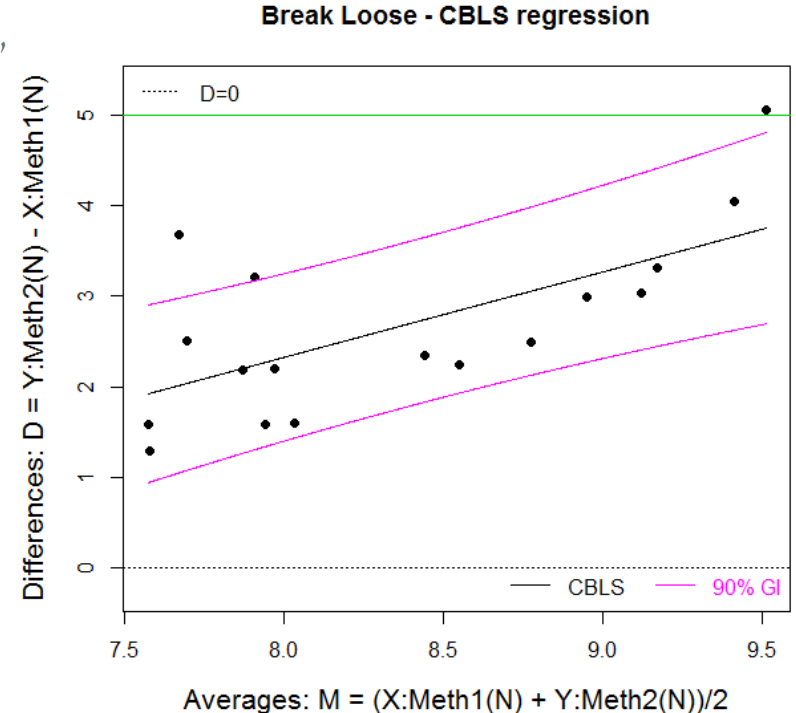


CMC 2 : Break loose assessment

Assess maximum difference between methods using **CBLS***

CBivBL = CBLS(workBL, ycol = 4: 53, xcol = 54: 89, **pred.level = 0.9**,
qx = 10, qy = 50)

90% of the differences between mean calculated on 10 syringes
(method 1) and mean calculated on 50 syringes (method 2)
< 5 Newtons



* Correlated BLS

Microbiology : Counting particles $\geq 10\mu\text{m}$ in drug powder

11 products

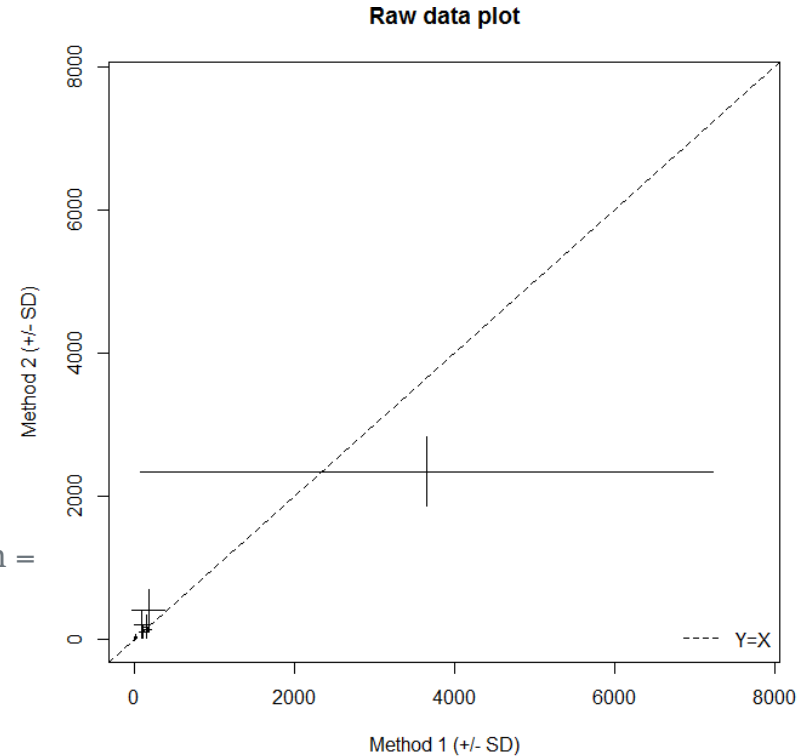
3 replicates for X and Y

Log-normal data

Estimated Variances (from replicates)

Assess conversion factor

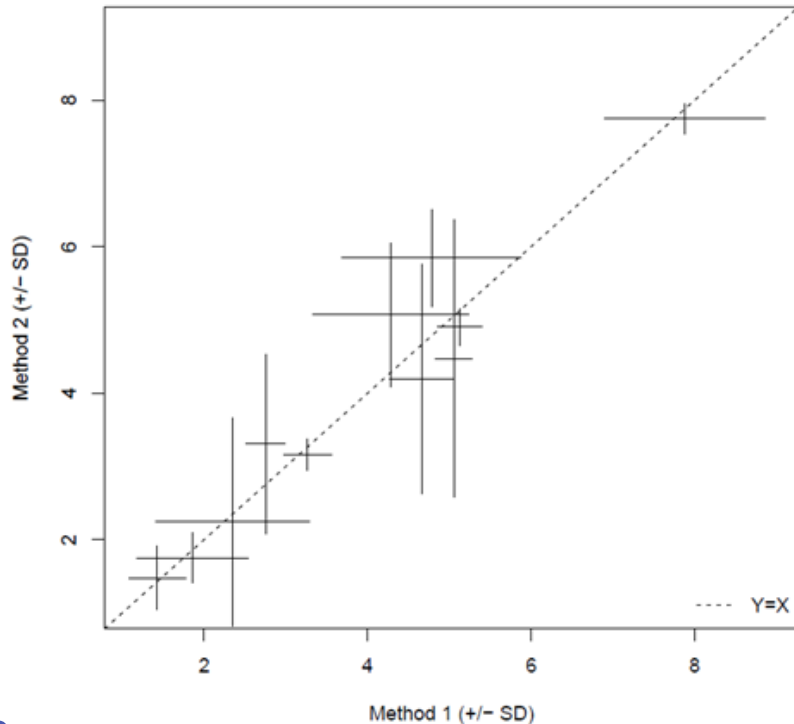
```
raw.plot(desc.stat(dt, xcol = 2: 4, ycol = 5: 7), graph = "XY.bar.SD", main =  
"Raw data plot", xname = "Method 1", yname = "Method 2")
```



Microbiology : Counting particles $\geq 10\mu\text{m}$ in drug powder

After Log-transformation

Log-transformed data plot



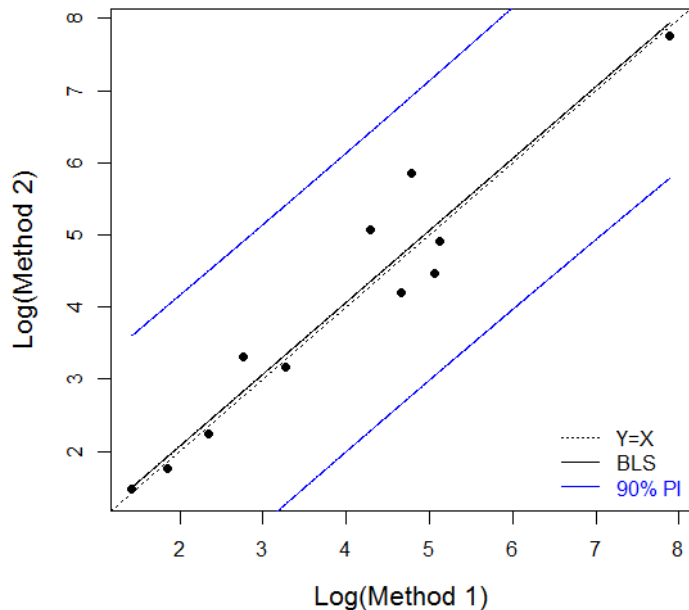
No obvious pattern for assuming heteroscedasticity

→ Assume variances ratio ($\text{VAR}(Y) / \text{VAR}(X)$) constant

Microbiology : Counting particles $\geq 10\mu\text{m}$ in drug powder

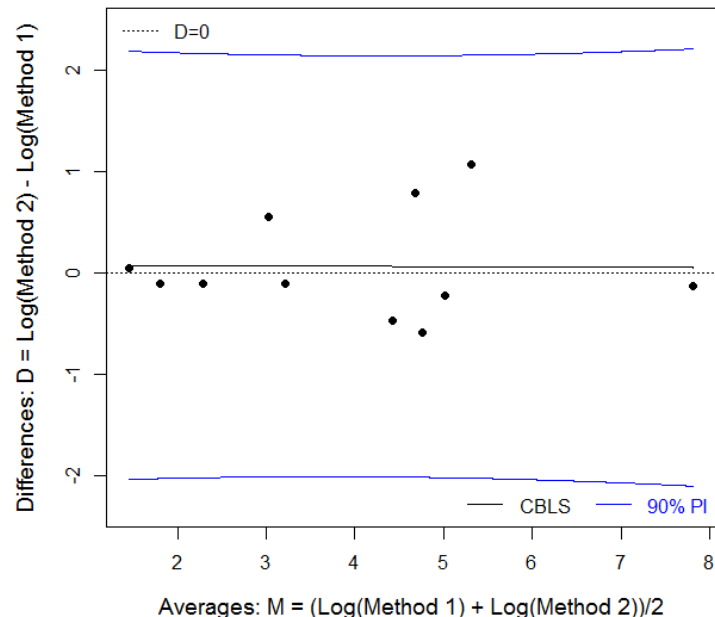
BLS

Particle $\geq 10\mu\text{m}$ – BLS regression plot



CBLs

Particle $\geq 10\mu\text{m}$ – CBLs regression plot



PI is the appropriate interval because only 1 replicate in routine

Microbiology : Counting particles $\geq 10\mu\text{m}$ in drug powder

Assess conversion factor

$\text{CbivBL} = \text{CBLs}(\text{dt}, \text{xc} = 2:4, \text{yc} = 5:7, \text{pred.level} = 0.9)$

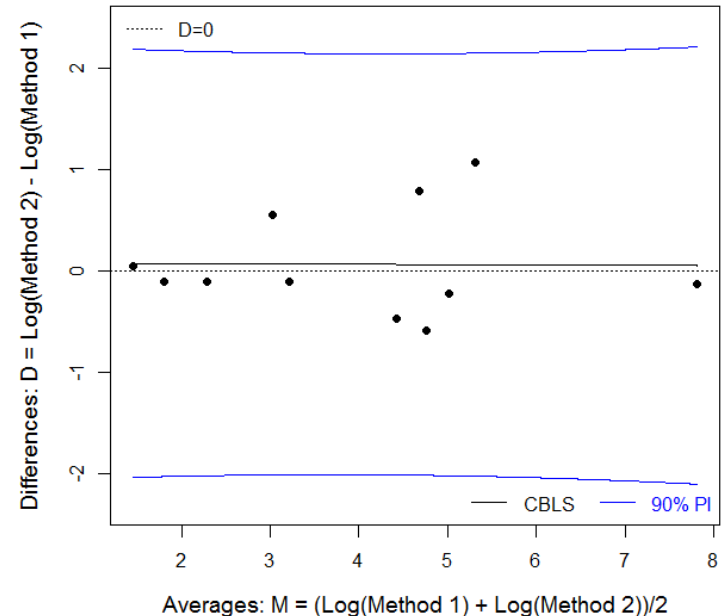
$\text{MD.plot}(\text{CbivBL}, \text{xname} = \text{"Log(Method 1)"}, \text{yname} = \text{"Log(Method 2)"}, \text{graph.int} = \text{c("PI")})$

$\text{Log}(\text{Method 2}) \in \text{Log}(\text{Method 1}) \pm 2.2$ at maximum

$$\text{Method 2} \in \left[\frac{\text{Method 1}}{9.03}; \text{Method 1} * 9.03 \right]$$

CBLs

Particle $\geq 10\mu\text{m}$ – CBLs regression plot



Microbiology : Counting particles $\geq 10\mu\text{m}$ in drug powder

Assume constant bias – Let's compare Agreement and Tolerance intervals

```
$Intervals.horiz.lines
```

	Mean differences	Lower	Upper
AI	0.06663185	-1.778585	1.911848
b TI	0.06663185	-2.019284	2.152548
bg TI	0.06663185	-2.225907	2.359171

← \equiv PI

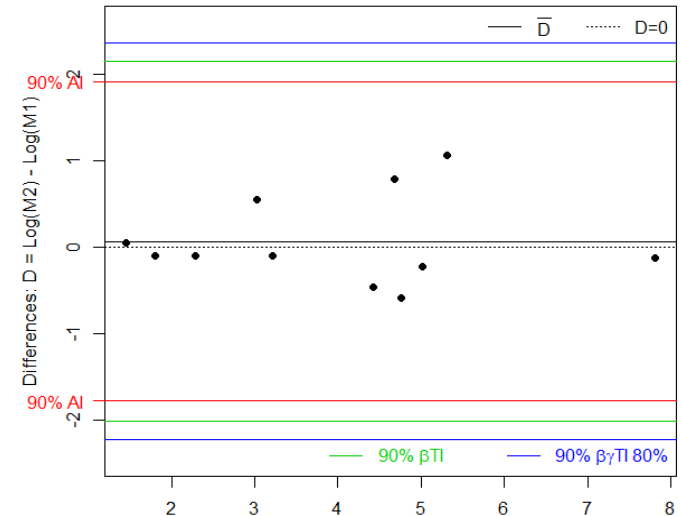
```

Meaning
90% Agreement interval
90% Tolerance interval
90% Tolerance interval (with 80% confidence level)
    
```

MDBL = MD.horiz.lines(dt, xcol = 2:4, ycol = 5:7, pred.level = 0.90,
TI.conf.level = 0.8)

MD.plot(results=MDBL, main = "MD plot", xname = "Log(M1)", yname =
 "Log(M2)", graph.horiz.int = c("bTI", "bgTI"))

Particle $\geq 10\mu\text{m}$ – MD plot



Averages: $M = (\text{Log}(M1) + \text{Log}(M2))/2$
 NCS2018, Oct 3-5 16

Risk of erroneously concluding equivalence with
 Bland-Altman approach

Bioanalysis : Blood measurement in humans

452 samples :

11 subjects

measured on 3 doses (cross-over)

at days 1 and 7

on 6-7 time points

No replicates

Log-normal data

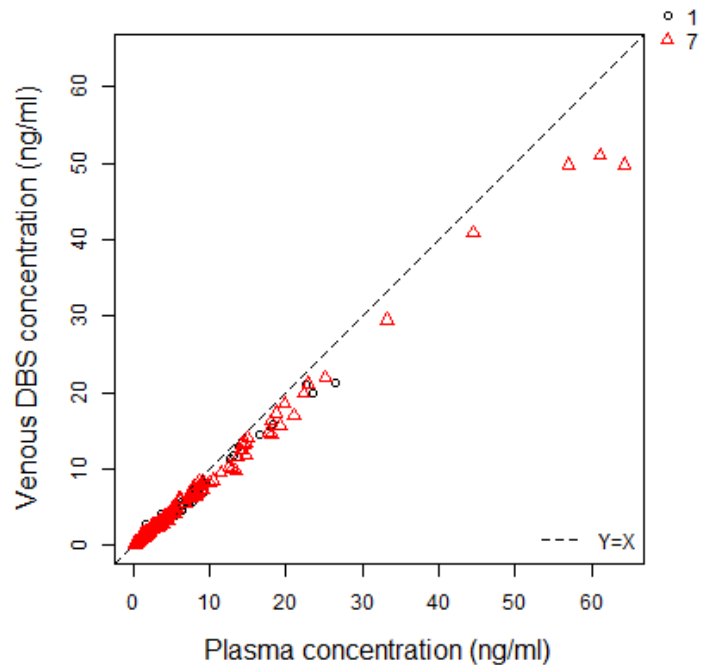
Estimated Variances from method validation studies

6 runs of 3 replicates at 4 concentrations

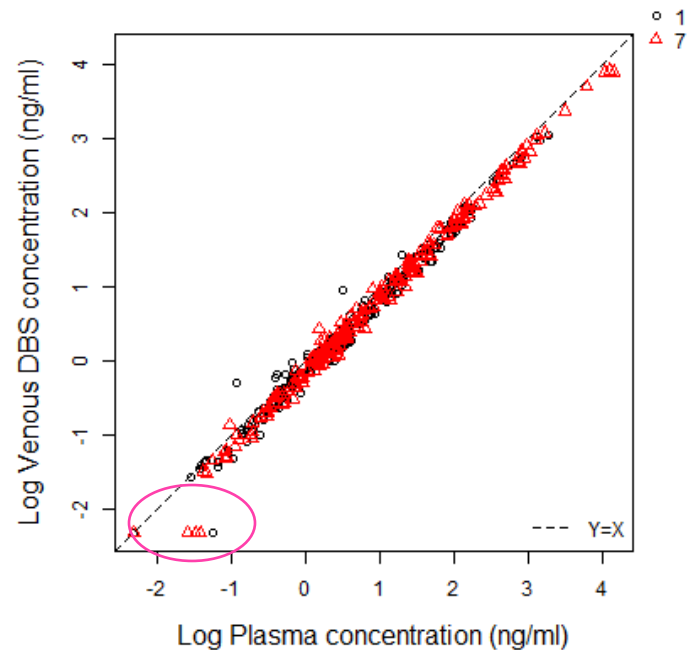
Acceptance criteria: $Y = X \pm 20\%$ of X

Bioanalysis : Blood measurement in humans

Raw data according to days



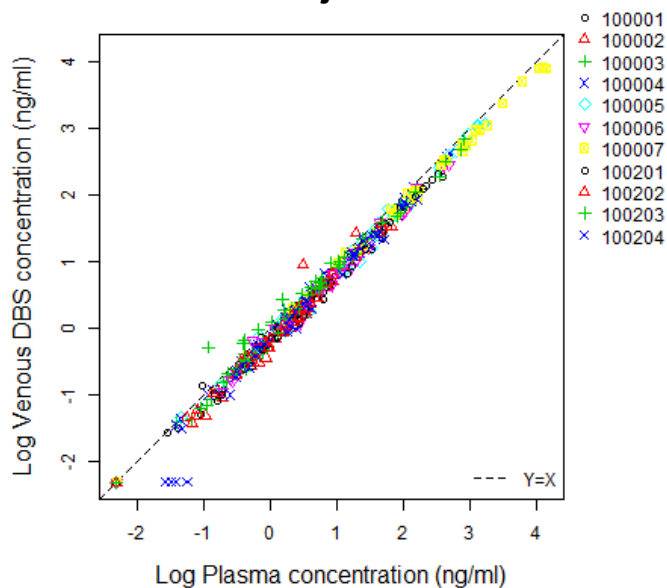
Log-transformed data according to days



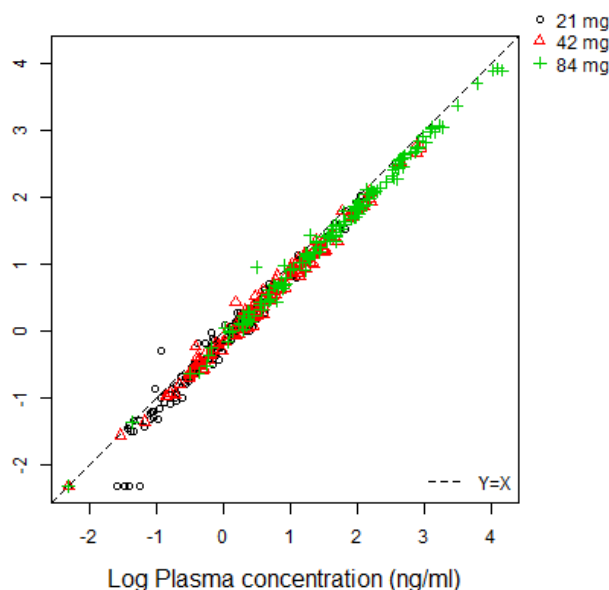
Bioanalysis : Blood measurement in humans

Log-transformed data according to ...

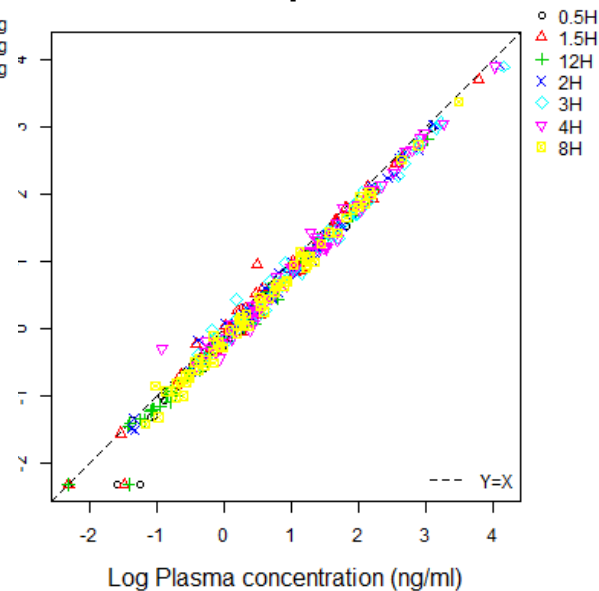
Subjects



Doses



Time points



Bioanalysis : Blood measurement in humans

Estimated Variances from validation studies for log concentrations

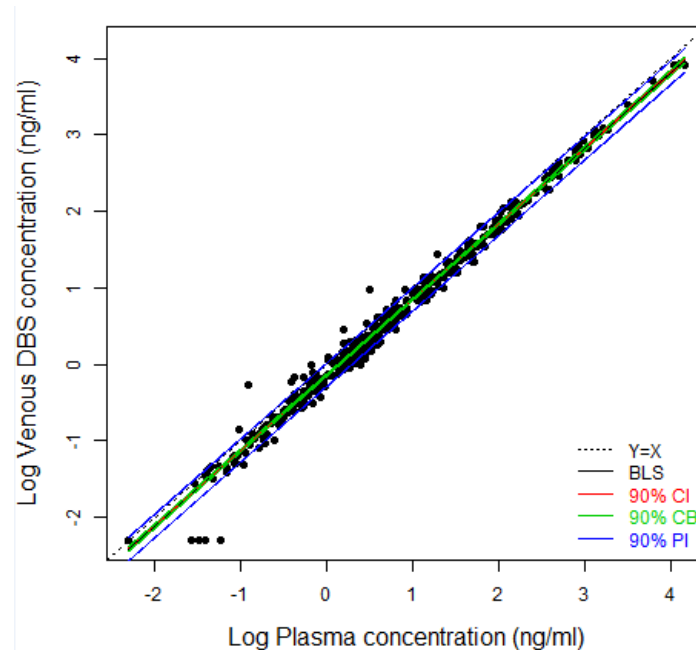
Plasma variance = 0.0032

Venous DBS variance = 0.0054

BLS slope and intercept estimates

BLS\$Estimates

Parameter	H0	Estimate	Std Error	Lower 90%CI	Upper 90%CI	pvalue
Intercept	0	-0.1416	0.0066	-0.1525	-0.1306	0.0000
Slope	1	0.9874	0.0046	0.9798	0.9950	0.0063

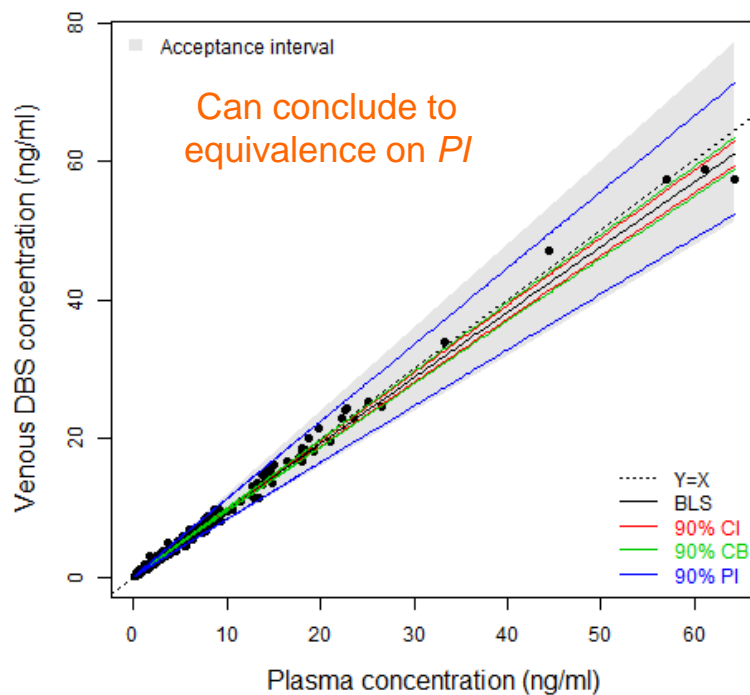
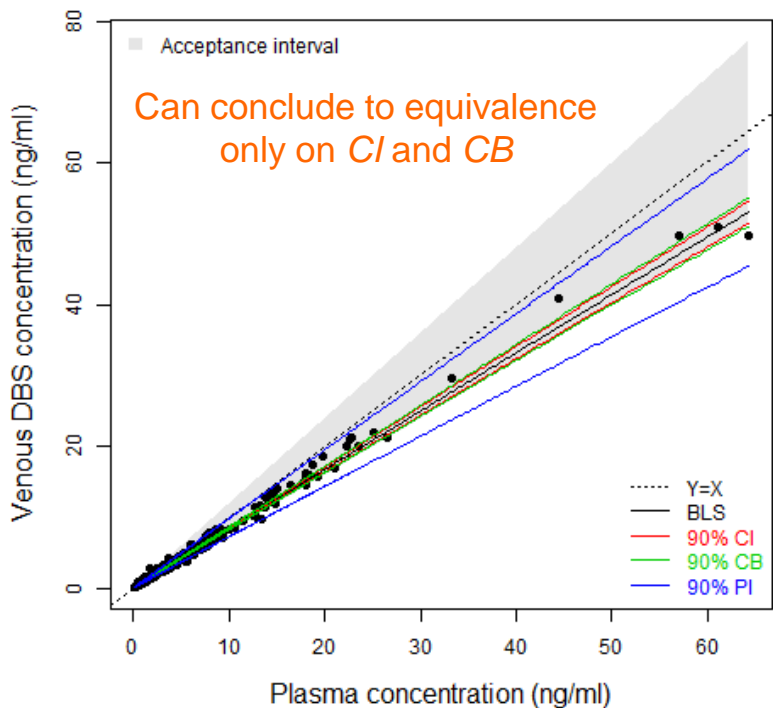


CB: Confidence Bounds

Bioanalysis : Blood measurement in humans

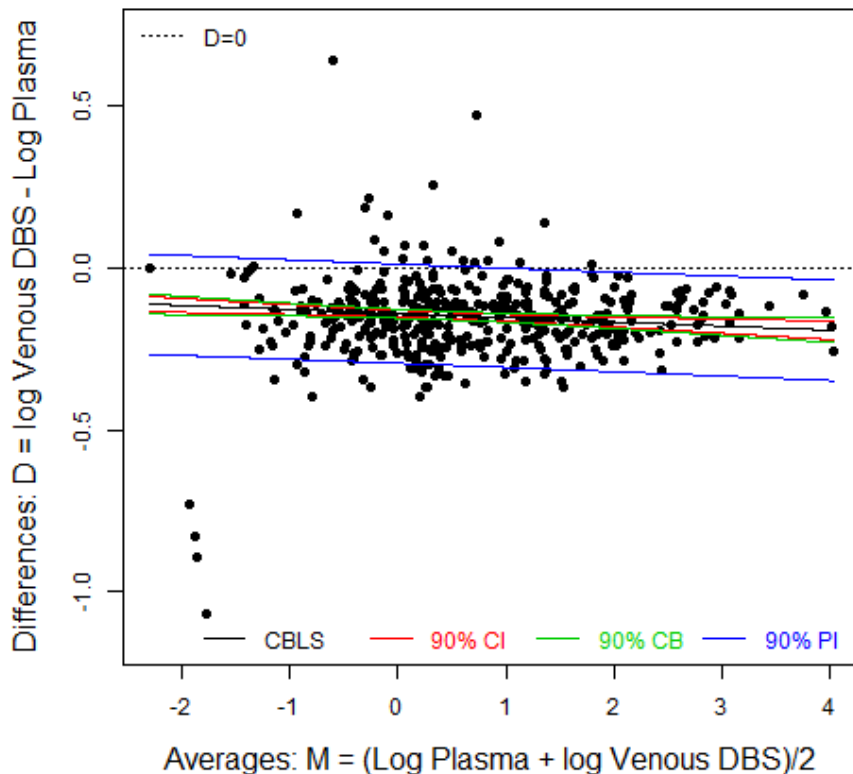
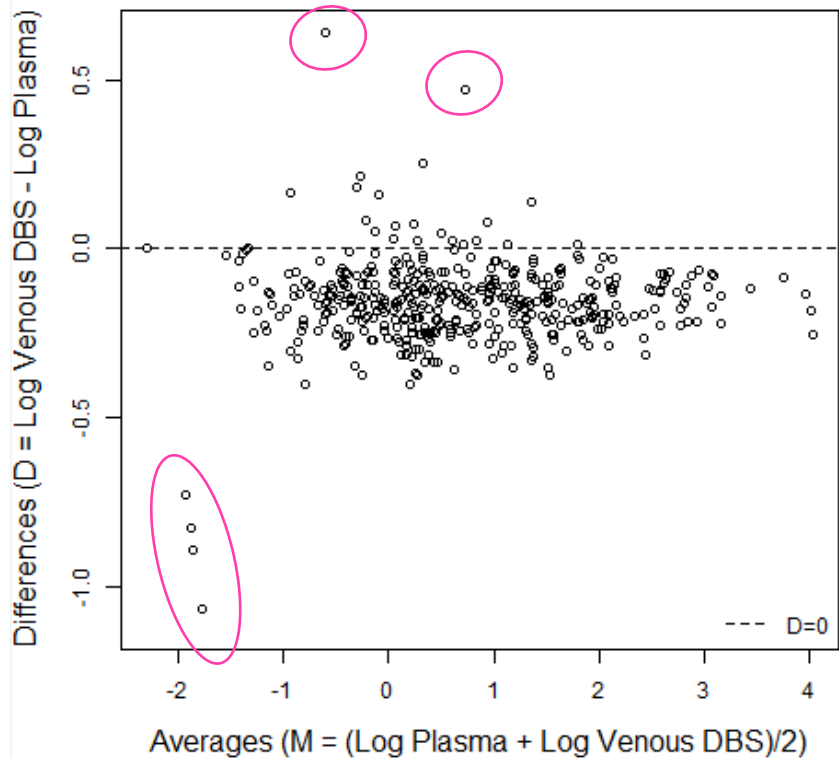
BLS regression with **anti-log**. Equivalence criteria: 20% of X

Let's correct for the intercept bias...



Bioanalysis : Blood measurement in humans

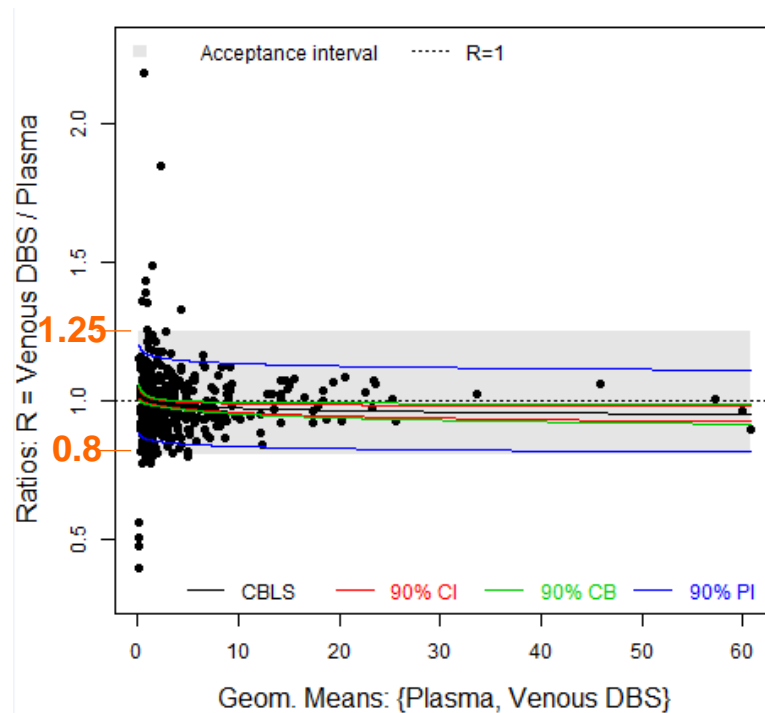
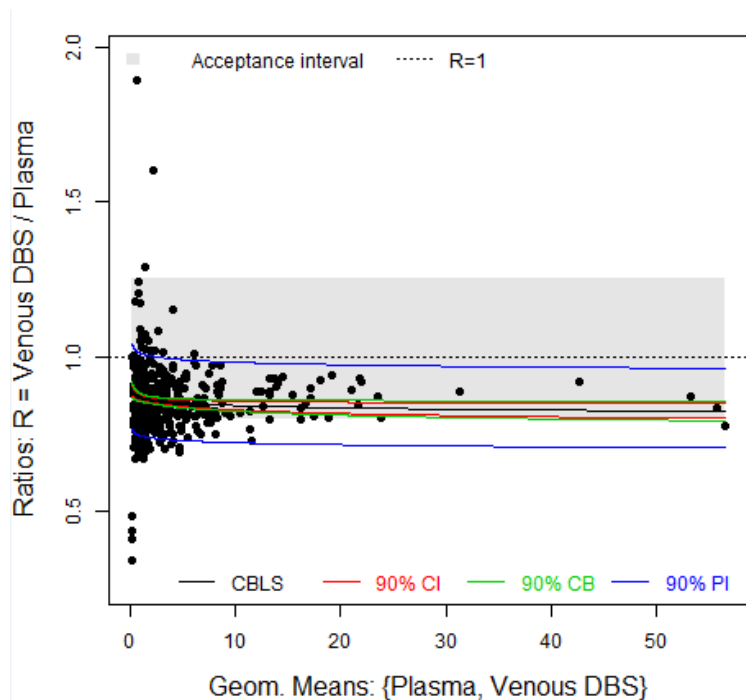
With CBLS regression



Bioanalysis : Blood measurement in humans

CBLS regression with **anti-log**. Equivalence criteria 20% of X.

Using corrected data...



Toxicology : Hemoglobin measurement in mice

36 animals

No replicates

Normal data

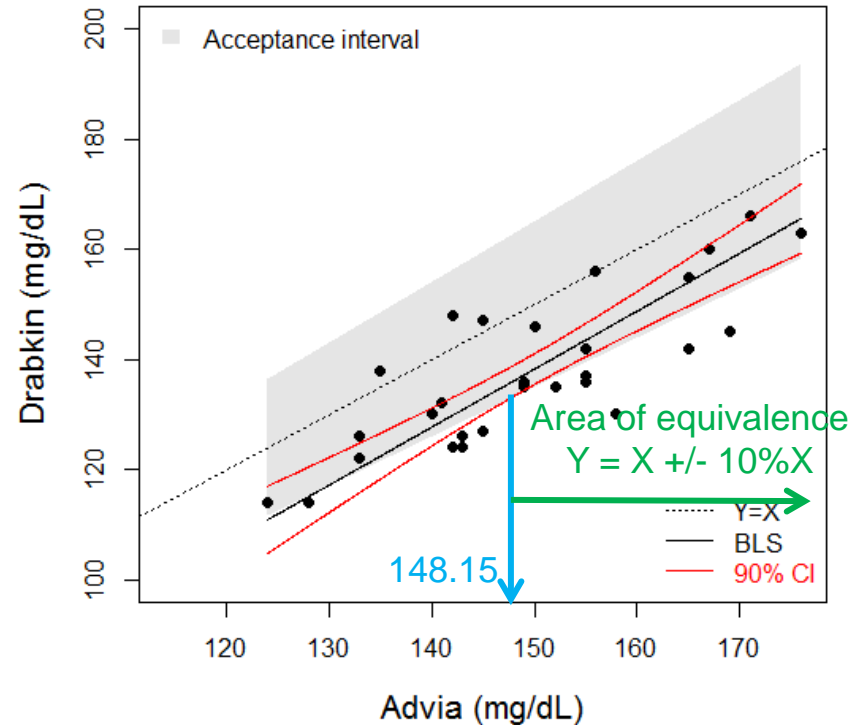
Unknown Variances

→ Try various ratios $\lambda_{XY} = \text{Var}Y/\text{Var}X$

→ Get CI or CB

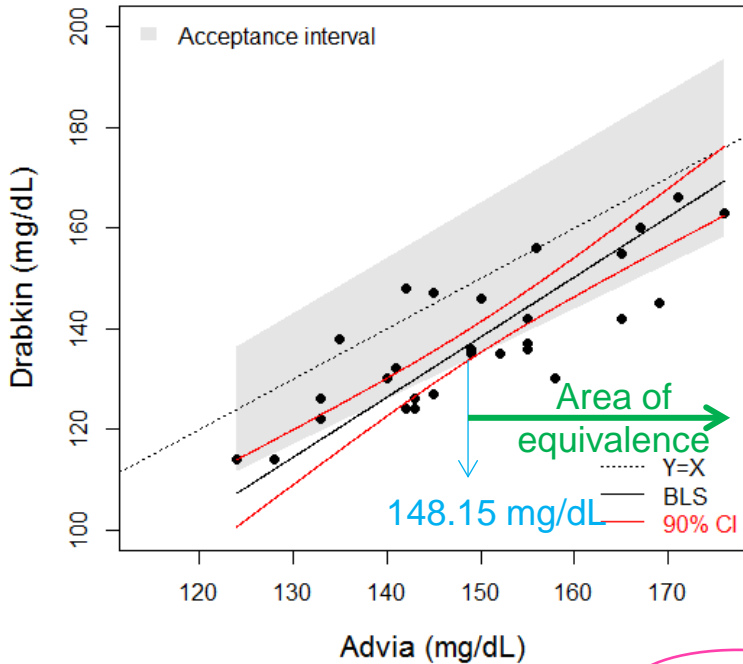
```
resBLS=BLS(data=drabkin, xcol=2, ycol=3, ratio.var=1,  
conf.level=0.90)
```

```
XY.plot(res.BLS, xname = "Advia (mg/dL)", yname = "Drabkin (mg/dL)",  
accept.int = 10, accept.int.perc=TRUE, graph.int = c("CI"), include.int=T)
```

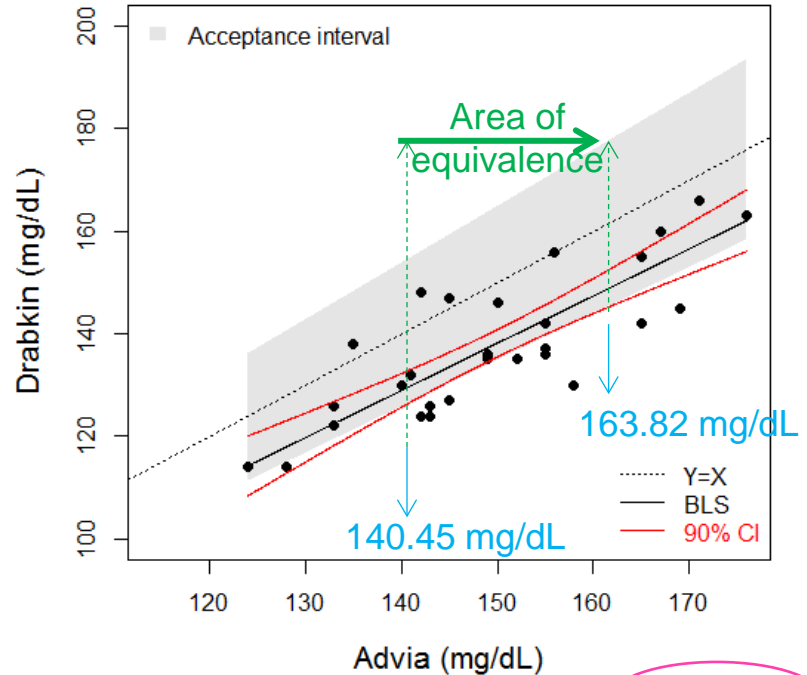


Toxicology : Hemoglobin measurement in mice

Test extreme variability ratios to assess impact on equivalence area



resBLS=BLS(data=drabkin, xcol=2, ycol=3, ratio.var=0.25, conf.level=0.90)

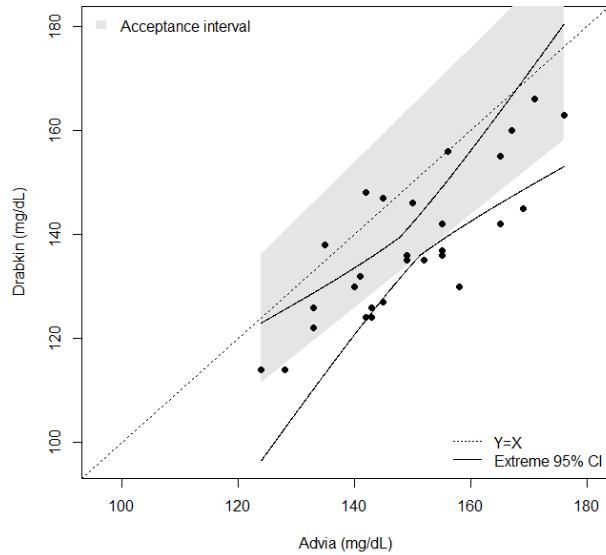


resBLS=BLS(data=drabkin, xcol=2, ycol=3, ratio.var=4, conf.level=0.90)

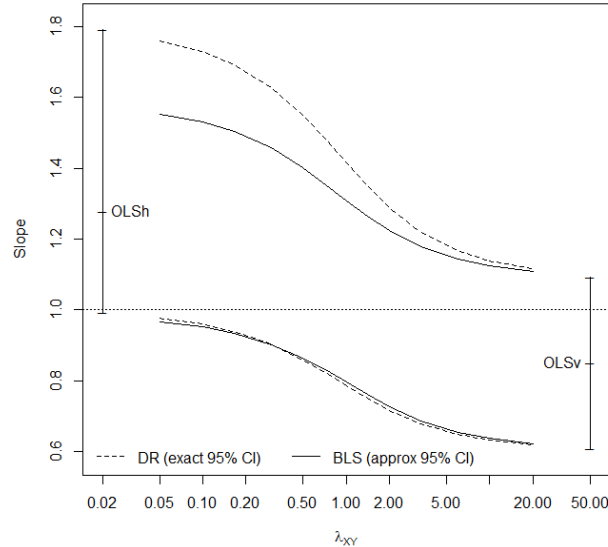
Toxicology : Hemoglobin measurement in mice

Check variability ratios λ_{XY} impact on

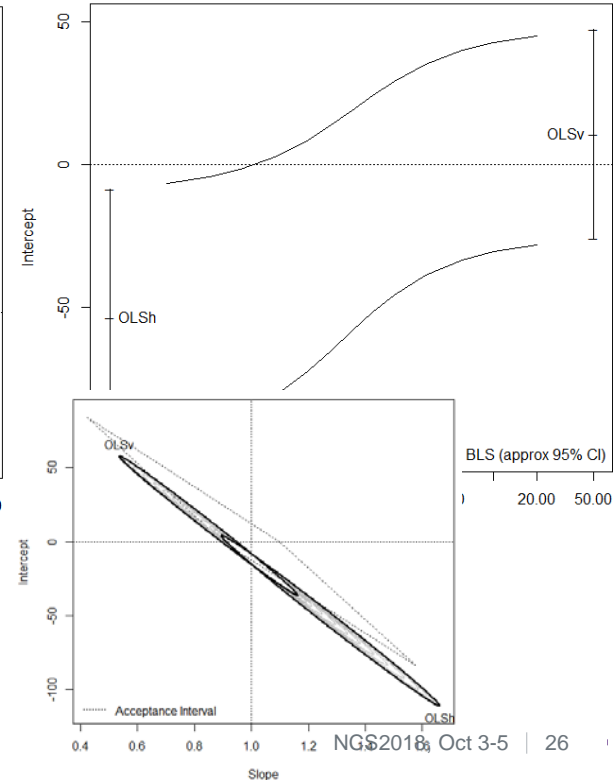
BLS's extreme CI



Slope's CI



Intercept's CI



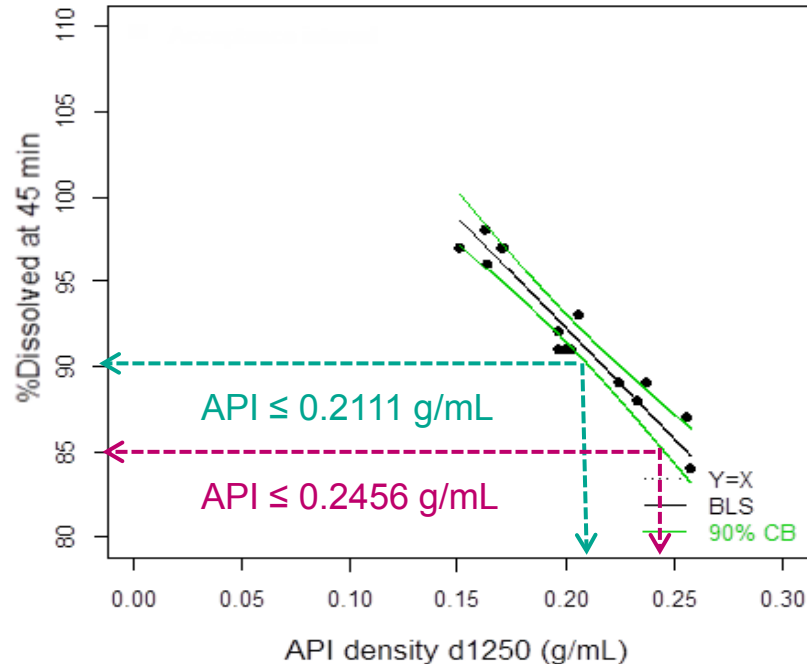
FullCI=FullCIs.XY(data = drabkin, xcol = 2, ycol = 3, conf.level = 0.95, npoints = 1000, nlambda = 13)

GraphFullCIs.XY(FullCIs = FullCI, xname = "Advia (mg/dL)", yname = "Drabkin (mg/dL)", graph = "all", accept.int = 10, accept.int.perc = TRUE, include.H0 = TRUE, include.int = TRUE)

CMC 3 : Just get a relationship

Regressing API density versus % dissolved at 45min

17 batches, **no replicates** but **variance ratio estimated from 2012-2013-2014 batches**



A few words on the BivRegBLS to finish



Fruit of a partnership



- Fully validated
- Very easy to use
 - 18 functions, many arguments for customization
- Handle unbalanced data
- Small to large sample size
- Variances' homogeneity or heterogeneity within methods and between methods
- BLS-CBLS techniques include:
 - OLSv, OLS_h,
 - Orthogonal Regression,
 - Geometric Mean Regression,
 - Deming Regression
- BLS (X,Y) is identical to CBLS (M,D)
- “Horizontal” Tolerance intervals are **robust to outliers**

How to get access to BivRegBLS (published on CRAN in January 2017)

- **Download R software from CRAN (Comprehensive R Archive Network) website:**
 - <https://CRAN.R-project.org/package=BivRegBLS>

- **Download BivRegBLS package by typing in R**
 - `install.packages('BivRegBLS')`

THANK YOU



References

- JM Bland, DG Altman. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8:135–160, 1999.
- B Carstensen. *Comparing Clinical Measurement Methods: A Practical Guide*. Wiley: Chichester, UK, 2010.
- Tan CY, Iglewicz B. Measurement-methods comparisons and linear statistical relationship. *Technometrics*, 1999; 41(3):192-201.
- Francq BG, Govaerts BB. How to regress and predict in a Bland-Altman plot? Review and contribution based on tolerance intervals and correlated-errors-in-variables models. *Statistics in Medicine*, 2016; 35:2328-2358.
- Francq BG, Govaerts BB. Measurement methods comparison with errors-in-variables regressions. From horizontal to vertical OLS regression, review and new perspectives. *Chemometrics and Intelligent Laboratory Systems*, 2014; 134:123-139.
- Francq BG, Govaerts BB. Hyperbolic confidence bands of errors-in-variables regression lines applied to method comparison studies. *Journal de la Societe Francaise de Statistique* 2014; 155(1):23-45.
- Francq BG. Errors-in-variables regressions to assess equivalence in method comparison studies. Ph.D. Thesis, Universite Catholique de Louvain, Institute of Statistics, Biostatistics and Actuarial science, Louvain-la-Neuve, Belgium, 2013.

This work was sponsored by Sanofi and the University of Glasgow.

Bernard G Francq is currently an employee of the GSK group of companies. Marion Berger is an employee of Sanofi. Whilst the authors are employees of GSK and Sanofi, there is no business relationship between the two companies, the authors worked on the study in a personal capacity.