Non Clinical Statistics Conference
October 3-4-5 2018
Maison de la Chimie, Paris

Be
Confident, Predictable or Tolerable
in Method Comparison Studies,

Correlated-Errors-in-Variables regressions in XY and MD Plots

Bernard G FRANCQ
GSK, Belgium
bernard.x.francq@gsk.com

Marion BERGER
Sanofi (Montpellier, France)
marion.berger@sanofi.com

The systolic blood pressure data (mmHg)

Bland, Altman. Measuring agreement in method Comparison studies. Stat Methods Med Res. 1999; 8:135-160

Compute the mean measures

Patient	Sphygometer 'J'			Semi automatic 'S'				
i	J1	J2	J3	S1	S2	S3	J	S
1	100	106	107	122	128	124	104.3	124.7
2	108	110	108	121	127	128	108.7	125.3
…	…	…	…	…	…	…	…	…
i	X_{i1}	X_{i2}	X_{i3}	Y_{i1}	Y_{i2}	Y_{i3}	X_i	Y_i
…	…	…	…	…	…	…	…	…
84	106	98	100	137	135	134	101.3	135.3
85	122	112	112	121	123	128	115.3	124.0

```
library(BivRegBLS); data(SBP)
res = desc.stat(data = SBP, xcol = 2:4, ycol = 8:10)
```

Question: is J (Manual) *equivalent* to S (Automatic) ?

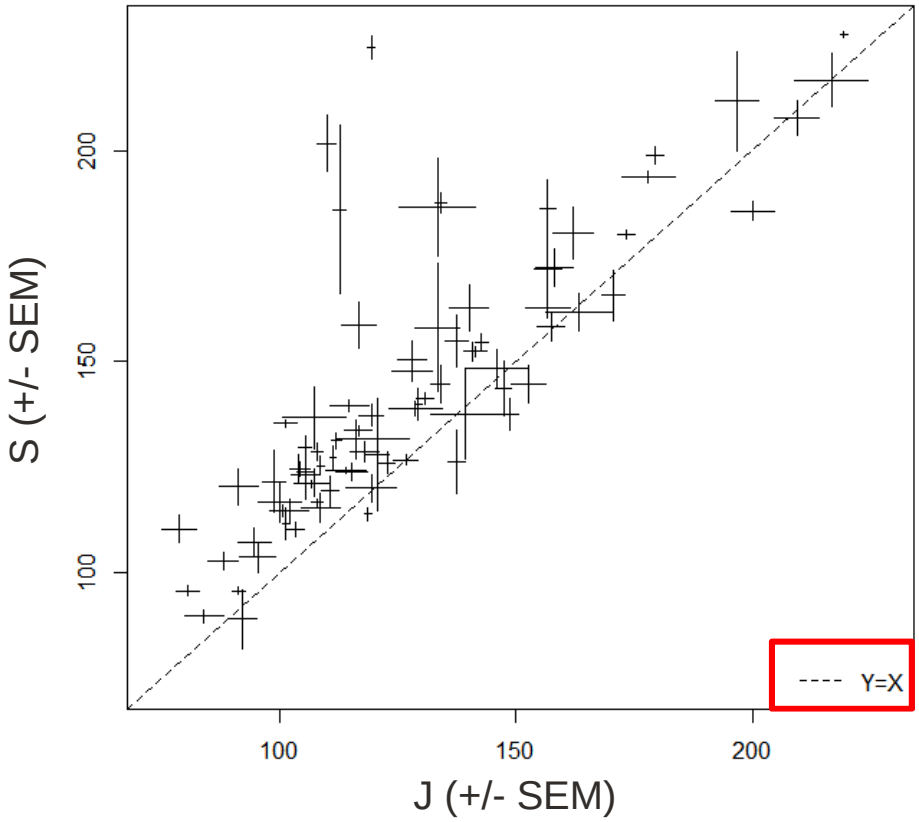
If my systolic blood pressure is measured at 125 mmHg with a manual device, what can we expect from an automatic device?

- ✓ Prediction of new measurements based on observed measures
- ✓ Prediction with or without measuremant uncertainty
- ✓ Prediction of average or individual measurements

XY and MD plot – replicated data

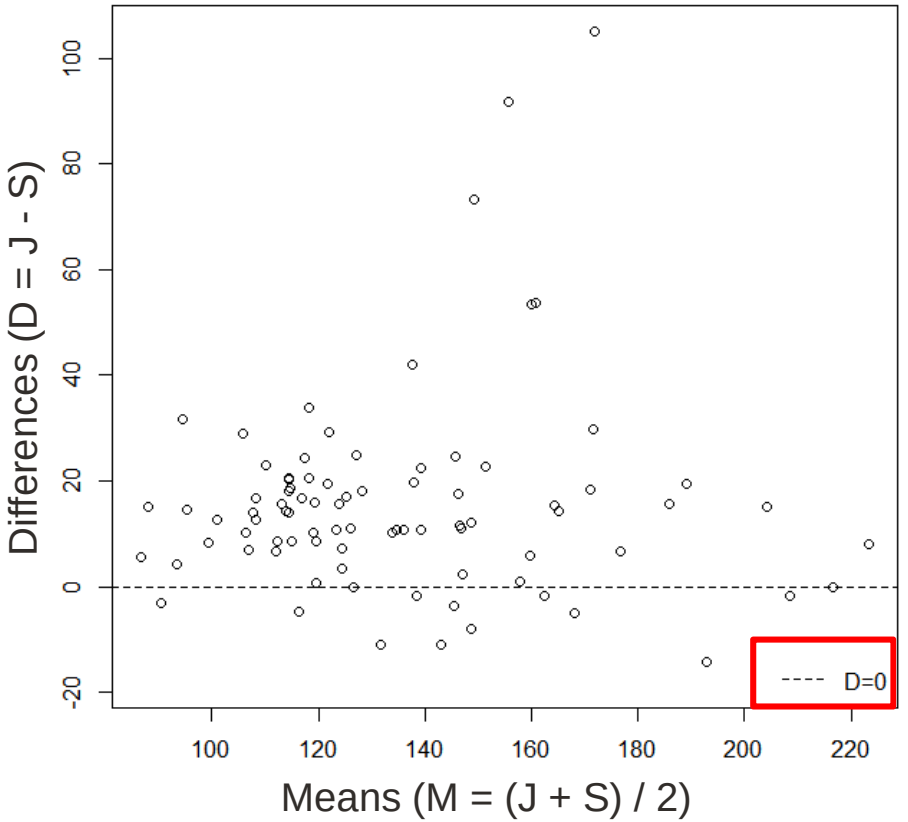
The classical XY plot (with additional error bars)

raw.plot(res)



The MD plot (Bland-Altman)

raw.plot(res, graph = "MD.means")



Question

Which statistical interval to use around the regression line?

Confidence Intervals (CI) vs Prediction Intervals (PI)

CI and PI are usually introduced separately in the statistical literature, i.e. OLS:

$$\hat{Y} \pm t_{0.975,n-1} S \sqrt{\frac{1}{n} + \frac{(X - \bar{X})^2}{S_{XX}}} \qquad \hat{Y} \pm t_{0.975,n-1} S \sqrt{1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{S_{XX}}}$$

When n → ∞ a CI collapses to the point estimate
a PI moves closer to the true quantiles

The concepts of 'confidence' and 'prediction' are actually similar

prediction interval
= confidence interval
or a future observation

confidence interval
= prediction interval for a mean

$$\Longrightarrow \quad \hat{Y} \pm t_{0.975,n-1} S \sqrt{\frac{\mathbf{1}}{\boldsymbol{q}} + \frac{1}{n} + \frac{(X - \bar{X})^2}{S_{XX}}}$$

GI = CI for $q = \infty$
GI = PI for $q = 1$

Why using Generalized Interval (GI)?

- Predict the mean of q new measurement (and not a single value)
- Release a batch based on a mean value
- Very useful when the number of replicates in the study design is different than the one during the process

Prediction Interval (PI) on replicated data **for single measure**

Solution given by Francq and Govaerts (*Statistics in Medicine, 2016*)

Prediction Intervals for BLS (XY plot) or CBLS (MD plot) regression

- Hyperbolic PI
- For $n_X = n_Y$ or $n_X \neq n_Y$ (number of replicates)
- Models based on average data but predict singles measures
- Excellent coverage probabilities

Generalized Intervals in errors-in-variables regressions

Solution given by Francq, Berger et al. (*BivRegBLS, R package 2017*)

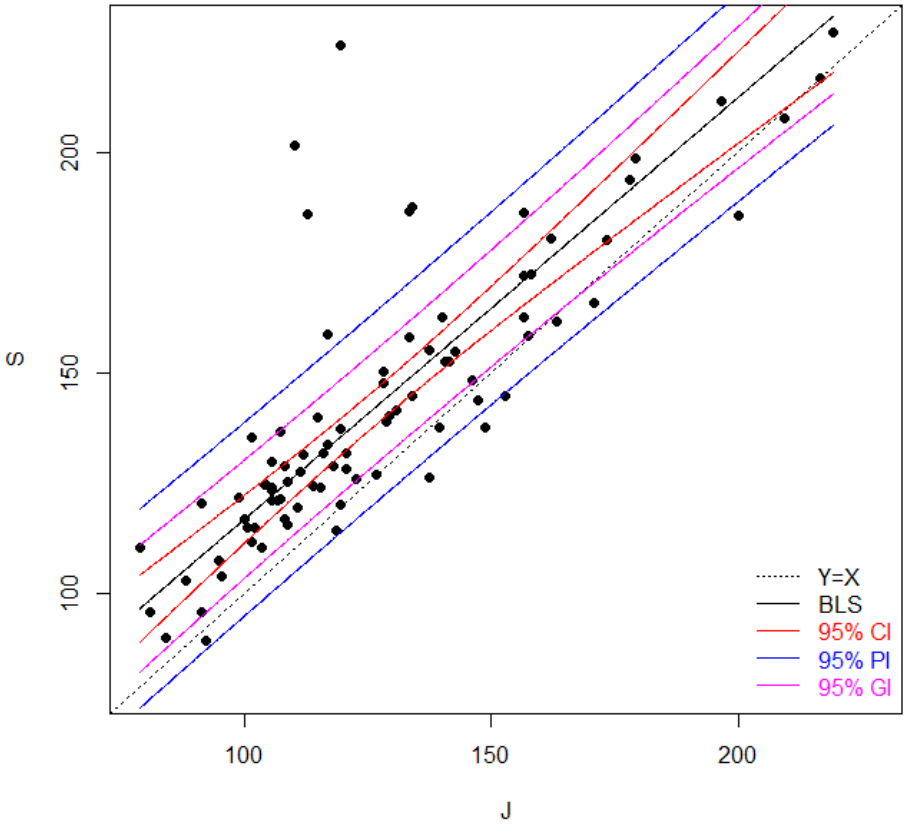
- Extends and includes the CI and PI
- Use 2 q parameters to tune the interval → q_X and q_Y

Predict the mean of q_Y measures by Y from the mean of q_X measures by X

- **GI = PI if $q_X = q_Y = 1$**
- **GI = CI if $q_X = q_Y = \infty$**

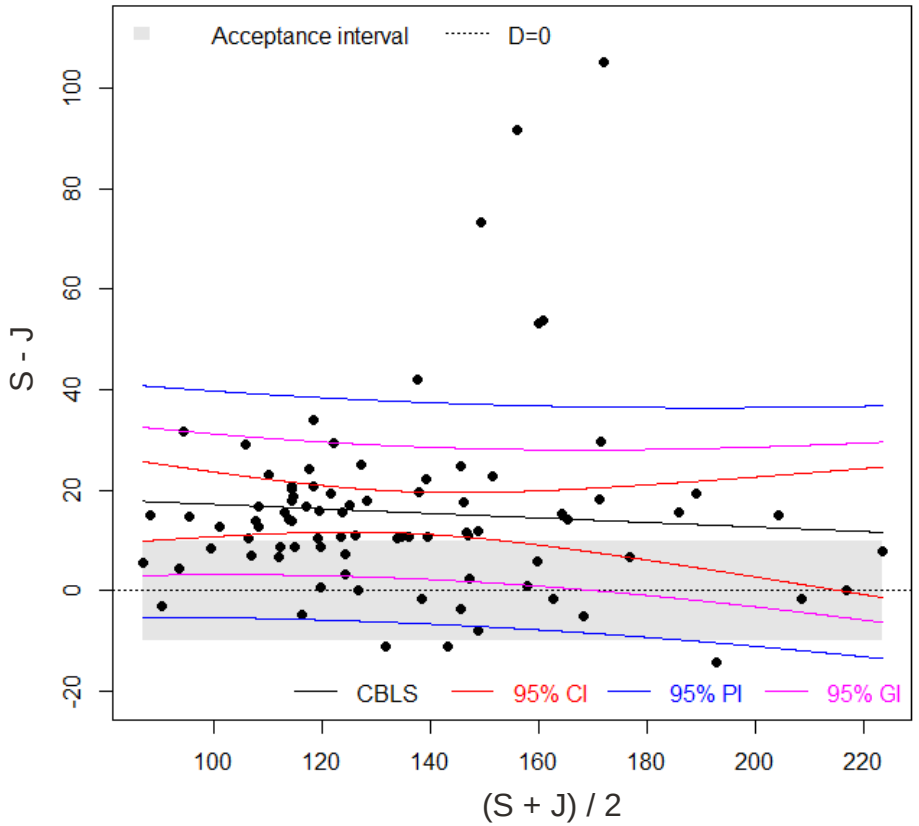
Generalized Intervals (GI)

BLS regression in a XY plot



res.BLS = BLS (data = SBP, …, qx = 3, qy = 3)
XY.plot (res.BLS, …, graph.int = c("CI", "PI", "GI"))

CBLS regression in a MD plot



res.CBLS = CBLS (data = SBP, …, qx = 3, qy = 3)
MD.plot (res.CBLS, graph.int = c("CI", "PI", "GI"), accept.int = 10)

- ✓ CI: Confidence Interval (for a mean = prediction without error)
- ✓ PI: Prediction Interval (for a single future value)
- ✓ **GI: Generalized prediction Interval for the mean of $q_X = q_Y = 3$ future value**

Conclusion

- BLS is the most general regression in a XY plot
- CBLS is the most general regression in a MD plot

with BivRegBLS:

- You can choose between many intervals according to your objective
- The Generalized Interval (GI) includes the CI (Confidence) and PI (Prediction)
- The GI can be tuned with 2 parameters, q_X and q_Y in errors-in-variables with BLS regression in XY plot or the CBLS in MD plot

:-) :-) :-) :-) :-) :-) :-) :-) :-) :-) :-)

:-) ***Give us your feedback*** **:-)**

:-) :-) :-) :-) :-) :-) :-) :-) :-) :-) :-)

References

- Francq BG, Berger M. BivRegBLS: Tolerance Intervals and Errors-in-Variables Regressions in Method Comparison Studies, 2017; R package version 1.0.0.
- 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 Française 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.

Conflict of interest statement

The authors declare the following interests: Bernard G Francq is 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.