Practical aspects of method comparison studies

Bendix Carstensen
Steno Diabetes Center
& Department of Biostatistics, University of Copenhagen
bxc@steno.dk
www.biostat.ku.dk/~bxc

Lyle Gurrin
School of Population Health, University of Melbourne
lgurrin@unimelb.edu.au
http://www.epi.unimelb.edu.au/about/staff/gurrin-lyle

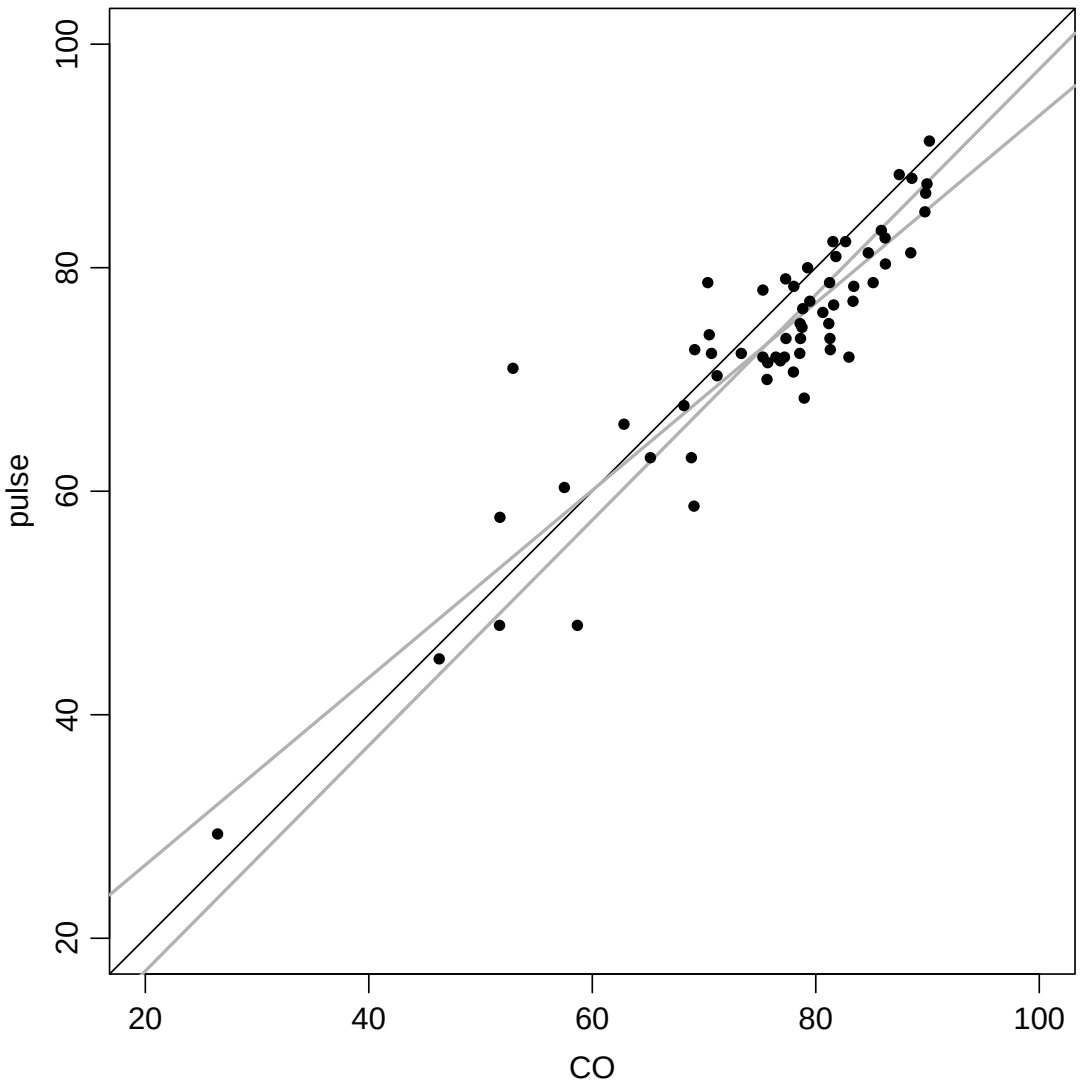
MEB, KI, Stockholm, 29 November 2007

Outline

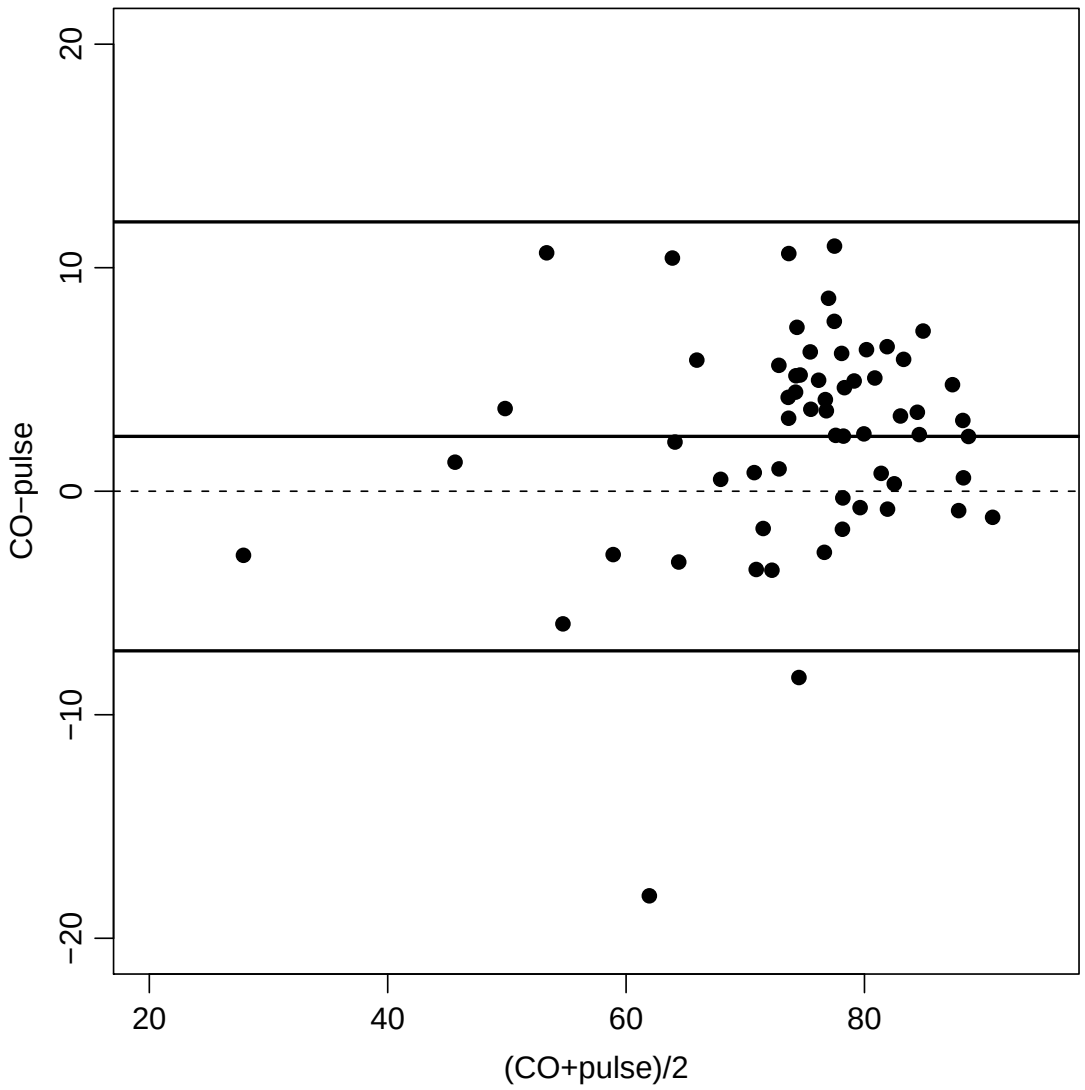
Comparing measurement methods

- Are results systematically different?
- Can one method safely be replaced by another?
- What is the size of measurement errors?
- Different centres use different methods of measurement:
  How do we convert from one method to another?
- How precise is the conversion?

Two methods for oxygen saturation:



Limits of agreement:



Plot differences (D_i) versus averages (A_i).

Limits of agreement:

How large is the difference between a measurement with method 1 and one with method 2 on a (randomly chosen) person?

$$D_i = y_{1i} - y_{2i}, \qquad \bar{D}, \qquad \text{s.d.}(D)$$

"Limits of agreement:"

$$\bar{D} \pm 2 \times \text{s.d.}(D)$$

95% prediction interval for the difference between a pair of future measurements by methods 1 and 2. [**?**, **?**]

Terminology

- **Methods**: $m = 1, \ldots, M$
  Machines, specimens, combinations of these.
- **Items**: $i = 1, \ldots, I$
  Individuals, patients, samples.
- **Replicates**: $r = 1, \ldots, R_{mi}$
  Not necessarily the same number of replicates throughout.

A dataset

```
> subset(ox,item<3)
     meth item repl    y
1      CO    1    1 78.0
2      CO    1    2 76.4
3      CO    1    3 77.2
4      CO    2    1 68.7
5      CO    2    2 67.6
6      CO    2    3 68.3
184 pulse    1    1 71.0
185 pulse    1    2 72.0
186 pulse    1    3 73.0
187 pulse    2    1 68.0
188 pulse    2    2 67.0
189 pulse    2    3 68.0
```

```
> subset(to.wide(ox), item<3)
  item repl  id   CO pulse
1    1    1 1.1 78.0    71
2    1    2 1.2 76.4    72
3    1    3 1.3 77.2    73
4    2    1 2.1 68.7    68
5    2    2 2.2 67.6    67
6    2    3 2.3 68.3    68
```

Limits of agreement: Model

Methods $m = 1, \ldots, M$, applied to $i = 1, \ldots, I$ individuals:

$$\begin{aligned} y_{mi} &= \alpha_m + \mu_i + e_{mi} \\ & \quad \text{s.d.}(e_{mi}) = \sigma_m \quad \text{— measurement error} \end{aligned}$$

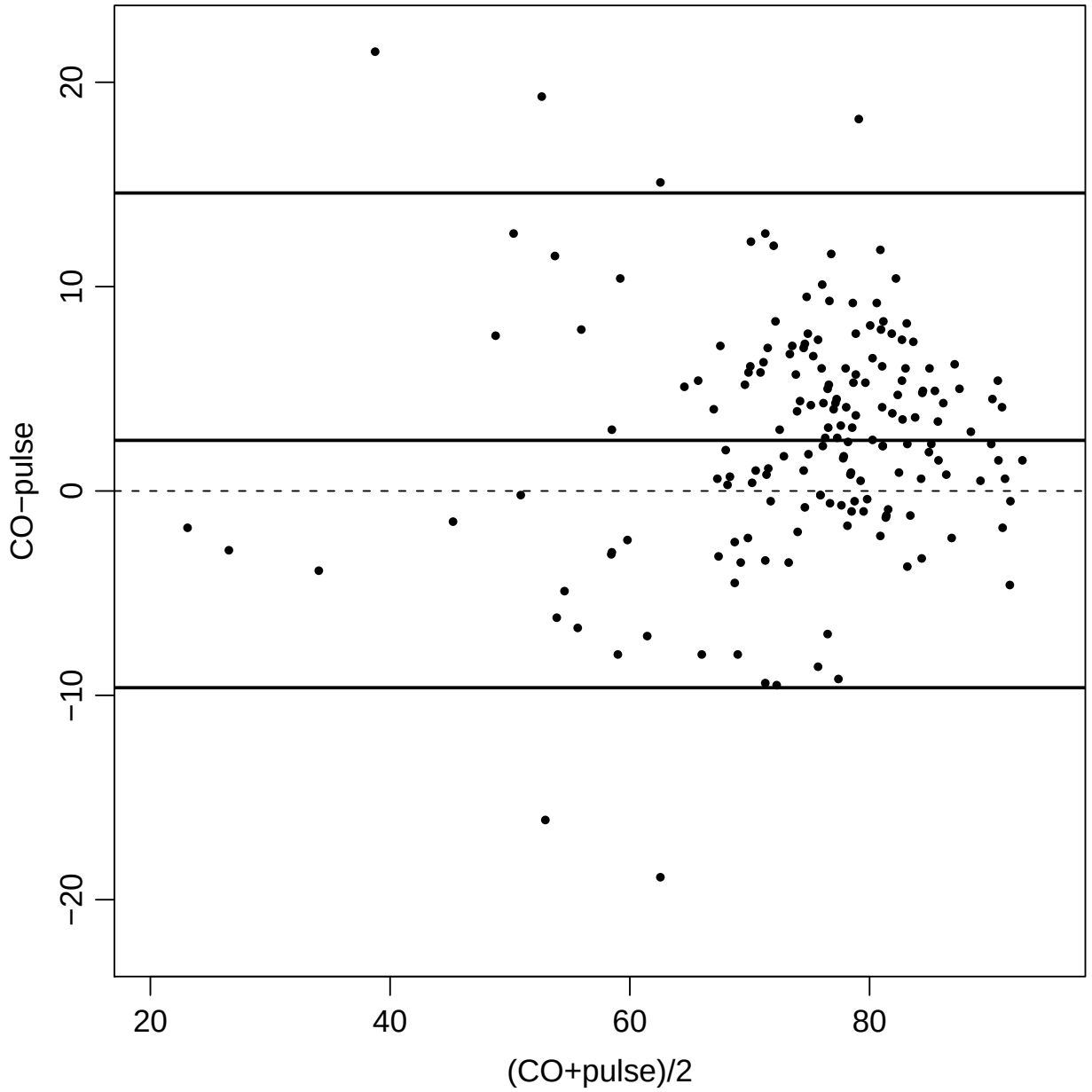
- Two-way analysis of variance model, with unequal variances in columns.
- Different variances are not identifiable without replicate measurements for $M = 2$.
- Unequal variances induce correlation between D_i and A_i: $\rho = \frac{1-\sigma_1^2/\sigma_2^2}{2(1+\sigma_1^2/\sigma_2^2)}$

Extension of the model: replicate measurements

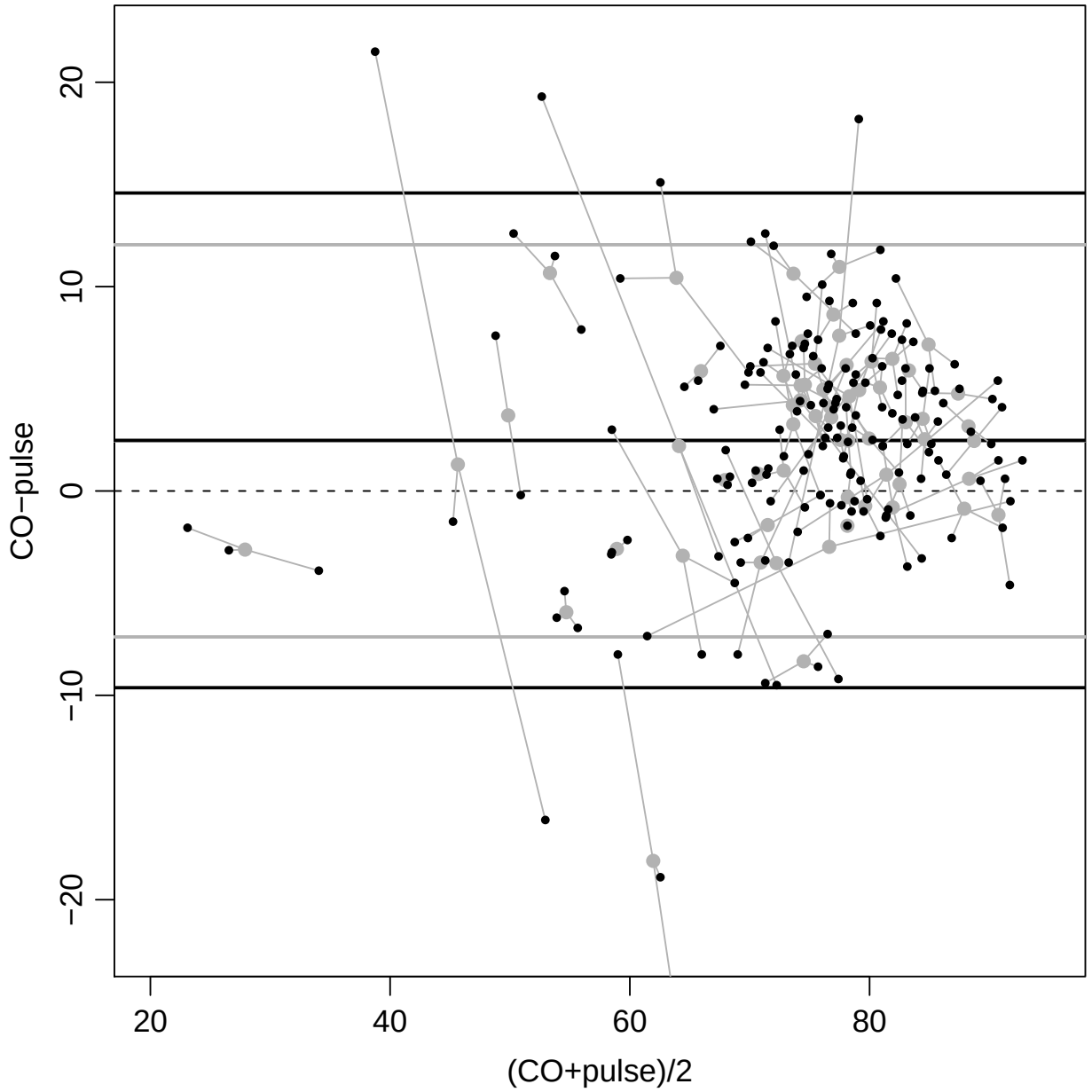
$$
\begin{aligned}
y_{mir} &= \alpha_m + \mu_i + c_{mi} + e_{mir} \\
&\quad \text{s.d.}(c_{mi}) = \tau_m \quad \text{— "matrix"-effect} \\
&\quad \text{s.d.}(e_{mir}) = \sigma_m \quad \text{— measurement error}
\end{aligned}
$$

- Replicates within (m, i) is needed to separate τ and σ.
- Even with replicates, the τs are only estimable if $M > 2$.
- Still assumes that the difference between methods is constant.

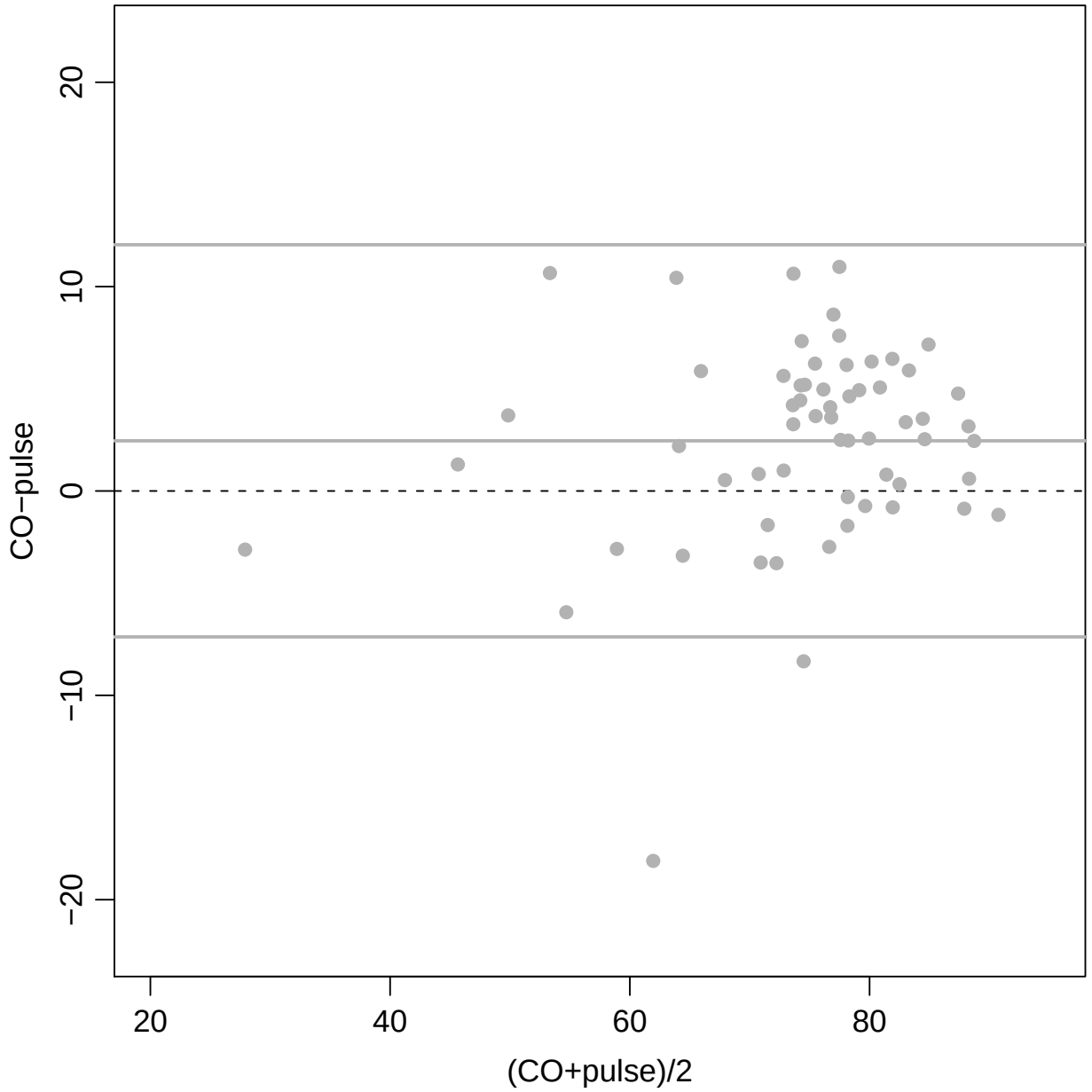
Oximetry data



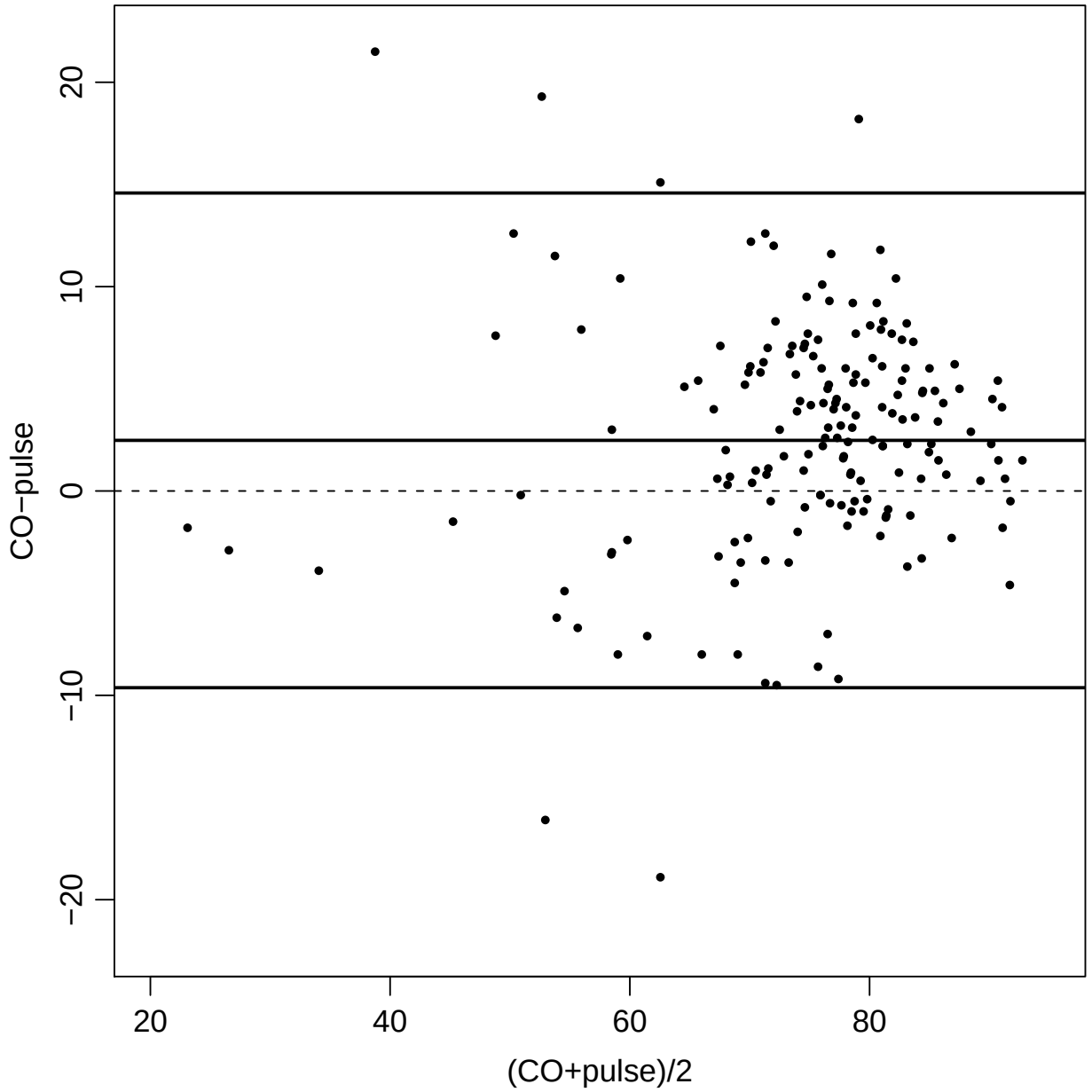
Oximetry data



Oximetry data



Oximetry data



Replicate measurements

- The limits of agreement should still be for difference between future **single** measurements.
- Analysis based on the **means** of replicates is therefore **wrong**.
- Model:

$$y_{mir} = \alpha_m + \mu_i + c_{mi} + e_{mir}$$

- $\text{var}(y_{1jx} - y_{2jx}) = 2\tau^2 + \sigma_1^2 + \sigma_2^2$

How to get the variance components: R

```
> library( nlme )
> lme( y ~ factor( item ) + factor( meth ),
+      random = ~1 | MI,
+      weights = varIdent( form = ~1 | meth ),
+      method ="REML",
+      data = cbind( ox,
+                    MI=interaction(ox$meth,ox$item) ) )
```

How to get the variances: R

```
Random effects:
 Formula: ~1 | MI
        (Intercept) Residual
StdDev:    2.190678 4.069055

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | meth
 Parameter estimates:
      CO    pulse
1.000000 1.288972
Number of Observations: 354
Number of Groups: 122
```

$\tau = 2.191$ $\sigma_1 = 4.069$ $\sigma_2 = 4.069 \times 1.289 = 5.245$

How to get the variance components: R

A convenience wrapper for this, `BA.est`, is in the `MethComp` package:

```
> BA.est(ox,exch=TRUE)
$bias
       CO     pulse
 0.000000 -2.475899

$sd.s
     MxI.CO   MxI.pulse    resid.CO resid.pulse
   2.190678    2.190677    4.069056    5.244897
```

How to get the variances: Stata

```
gen meth2 = meth - 1
gen obsid = _n
gen MI = item + 100*(meth==1)
xi:xtmixed y meth i.item ||MI: || obsid: meth2, nocons
```

```
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.
-----------------------------+------------------------------------------------
MI: Identity       sd(_cons) |   2.190685    .532584
-----------------------------+------------------------------------------------
obsid: Identity    sd(meth2) |    3.30932   .6333033
-----------------------------+------------------------------------------------
                   sd(Residual) |   4.069061   .2661295
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(2) =    14.07
```

$\tau = 2.191 \qquad \sigma_1 = 4.069 \qquad \sigma_2 = \sqrt{4.069^2 + 3.309^2} = 5.245$

How to get the variances: SAS

```
proc mixed  data = ox ;
  class item meth ;
  model y = meth item ;
  random meth * item ;
  repeated item / group = meth ;
run ;
```

```
Covariance Parameter Estimates
Cov Parm        Group      Estimate
meth*item                    4.7991
item            meth 1      16.5572
item            meth 2      27.5089
```

$$\tau = \sqrt{4.7991} = 2.191, \qquad \sigma_1 = \sqrt{16.5572} = 4.069$$

$$\sigma_2 = \sqrt{27.5089} = 5.245$$

Extension of the model: non-constant differences

$$y_{mi} = \alpha_m + \beta_m \mu_i + e_{mi}$$

μ_i : "true" individual level

e_{mi} : measurement error, σ_m

- Measurements linearly related to a "true" value, μ_i.
- Extension of the main-effects model with a parametric interaction term, $\beta_m \mu_i$.
- Non-linear model because of the product $\beta_m \mu_i$.

Extension of the model: non-constant differences

$$y_{mi} = \alpha_m + \beta_m \mu_i + e_{mi}$$

μ_i : "true" individual level

e_{mi} : measurement error, σ_m

- Not all (α_m, β_m) can be identified.
- The μs are only unique up to a linear transformation.
- $2M + I$ mean value parameters, dimension only $2M + I - 2$.

Relationship between methods

Translation formula from method 1 to method 2 is (for the mean):

$$\begin{aligned} y_2 = \alpha_2 + \beta_2 \mu &= \alpha_2 + \beta_2 (y_1 - \alpha_1)/\beta_1 \\ &= \left(\alpha_2 - \alpha_1 \beta_2/\beta_1\right) + (\beta_2/\beta_1) y_1 \end{aligned}$$

Intercept and slope going from method 1 to 2:

$$\begin{aligned} \alpha_{2\cdot 1} &= \alpha_2 - \alpha_1 \beta_2/\beta_1 \\ \beta_{2\cdot 1} &= \beta_2/\beta_1 \end{aligned}$$

Invariant under linear transformation of the μs.

Extension with variance components

Three-way layout:
Method, Individual, Replicate.
Three two-way interactions:

$$y_{mir} = \alpha_m + \beta_m(\mu_i + a_{ir}) + c_{mi} + d_{mr} + e_{mir}$$

Exchangeability of replicates within methods and individuals determine which interactions are relevant.

Variance components

Method, Item, Replicate.

$$\begin{aligned} y_{mir} &= \alpha_m + \beta_m(\mu_i + a_{ir}) + c_{mi} + d_{mr} + e_{mir} \\ & \quad \text{s.d.}(c_{mi}) = \tau_m \end{aligned}$$

Matrix-effect: Each item reacts differently to each method.

If only two methods compared:
τ_1 and τ_2 cannot be separated:

$$\begin{aligned} y_{mir} &= \alpha_m + \beta_m(\mu_i + a_{ir} + c_{mi}) + d_{mr} + e_{mir} \\ & \quad \text{s.d.}(c_{mi}) = \tau \end{aligned}$$

Variance components

Method, Item, Replicate.

$$\begin{aligned} y_{mir} &= \alpha_m + \beta_m(\mu_i + a_{ir}) + c_{mi} + d_{mr} + e_{mir} \\ & \quad \text{s.d.}(d_{mr}) = \nu_m \end{aligned}$$

Number of methods and replicates are normally small.
More likely to be included as a fixed effect, for example as specific effects of analysis day for each method.

Variance components

Method, Item, Replicate.

$$\begin{aligned} y_{mir} &= \alpha_m + \beta_m(\mu_i + a_{ir}) + c_{mi} + d_{mr} + e_{mir} \\ & \quad \text{s.d.}(a_{ir}) = \omega \end{aligned}$$

Common across methods — must be scaled relative to the methods.
Included if replicates are linked across methods, e.g. if there is a sequence in the replicates.
The relevant quantities to reports are $\beta_m\omega$ — the s.d. on the scale of the mth method.

Predicting method 2 from method 1

$$
\begin{aligned}
y_{10r} &= \alpha_1 + \beta_1(\mu_0 + a_{0r}) + c_{10} + e_{10r} \\
y_{20r} &= \alpha_2 + \beta_2(\mu_0 + a_{0r}) + c_{20} + e_{20r} \\
&\Downarrow \\
y_{20r} &= \alpha_2 + \frac{\beta_2}{\beta_1}(y_{10r} - \alpha_1 - c_{10} - e_{10r}) \\
&\qquad + c_{20} + e_{20r}
\end{aligned}
$$

The random effects have expectation 0, so:

$$
\mathrm{E}(y_{20r}|y_{10r}) = \hat{y}_{20r} = \alpha_2 + \frac{\beta_2}{\beta_1}(y_{k0r} - \alpha_1)
$$

$$
\begin{aligned}
y_{20r} &= \alpha_2 + \frac{\beta_2}{\beta_1}(y_{10r} - \alpha_1 - c_{10} - e_{10r}) \\
&\quad + c_{20} + e_{20r} \\
\text{var}(\hat{y}_{20r}|y_{10r}) &= \left(\frac{\beta_2}{\beta_1}\right)^2(\tau_1^2 + \sigma_1^2) + (\tau_2^2 + \sigma_2^2)
\end{aligned}
$$

The slope of the prediction line from method 1 to method 2 is β_2/β_1.
The width of the prediction interval is:

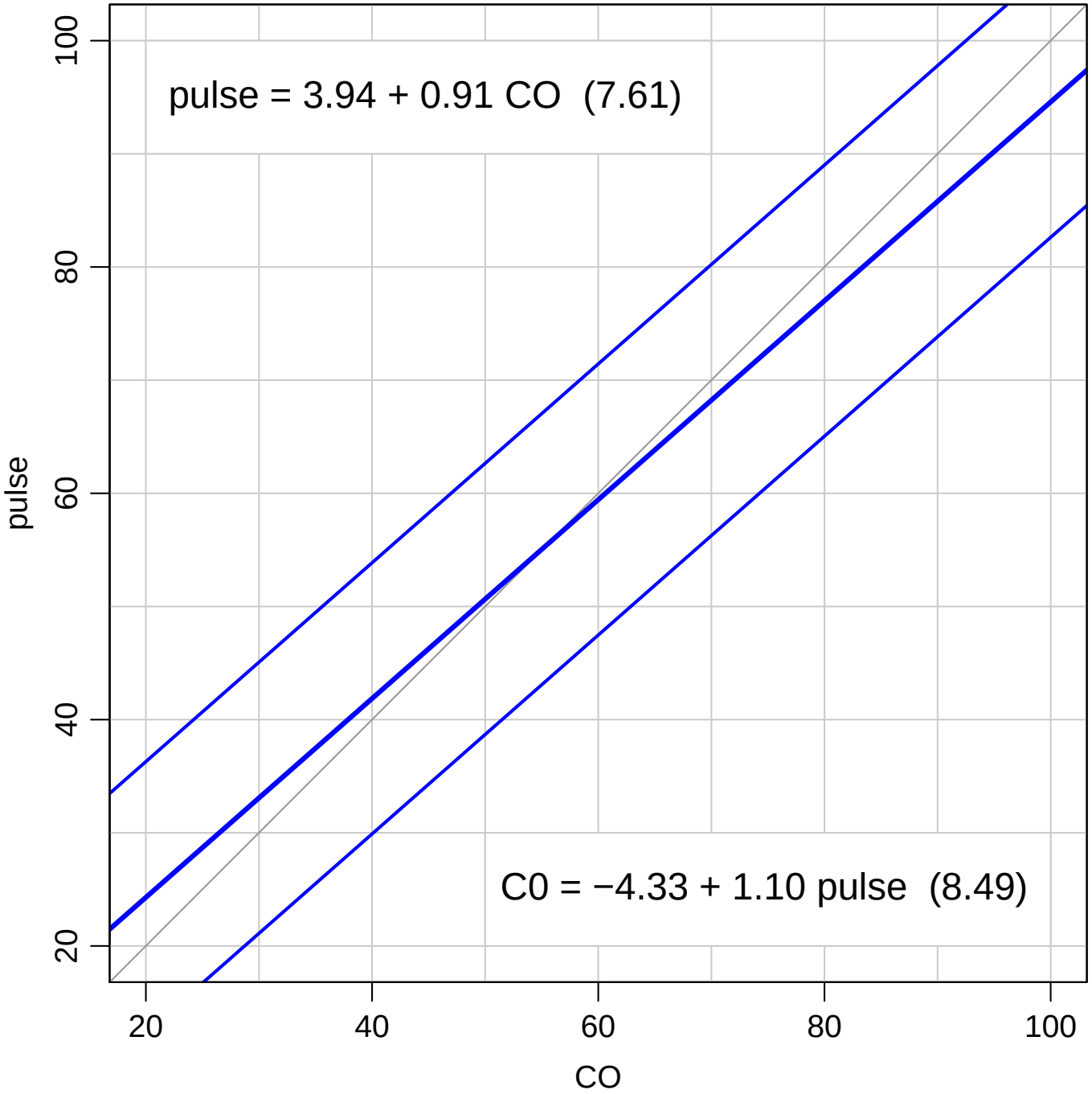
$$
2 \times 1.96 \times \sqrt{\left(\frac{\beta_2}{\beta_1}\right)^2(\tau_1^2 + \sigma_1^2) + (\tau_2^2 + \sigma_2^2)}
$$

If we do the prediction the other way round $(y_1|y_2)$ we get the same relationship i.e. a line with the inverse slope, β_1/β_2.
The width of the prediction interval in this direction is:

$$2 \times 1.96 \times \sqrt{(\tau_1^2 + \sigma_1^2) + \left(\frac{\beta_1}{\beta_2}\right)^2 (\tau_2^2 + \sigma_2^2)}$$

$$= 2 \times 1.96 \times \frac{\beta_1}{\beta_2} \sqrt{\left(\frac{\beta_2}{\beta_1}\right)^2 (\tau_1^2 + \sigma_1^2) + (\tau_2^2 + \sigma_2^2)}$$

i.e. if we draw the prediction limits as traight lines they can be used both ways.



Alternating random effects models

Carstensen [**?**] proposed a ridiculously complicated approach to fit the model

$$y_{mir} = \alpha_m + \beta_m \mu_i + c_{mi} + e_{mir}$$

- For fixed μ it's just a linear mixed model.
- For fixed (α, β) it's just a regression through 0.

— plus a bit of fidgeting with the BLUPs.

Implementation in BUGS

$$y_{mir} = \alpha_m + \beta_m(\mu_i + a_{ir} + c_{mi}) + e_{mir}$$

Non-linear hierarchical model:
Implement in BUGS.

- The model is *symmetrical* in methods.
- Mean is overparametrized.
- Choose a prior (and hence posterior!) for the μs with finite support.
- Keeps the chains nicely in place.

Results from fitting the model

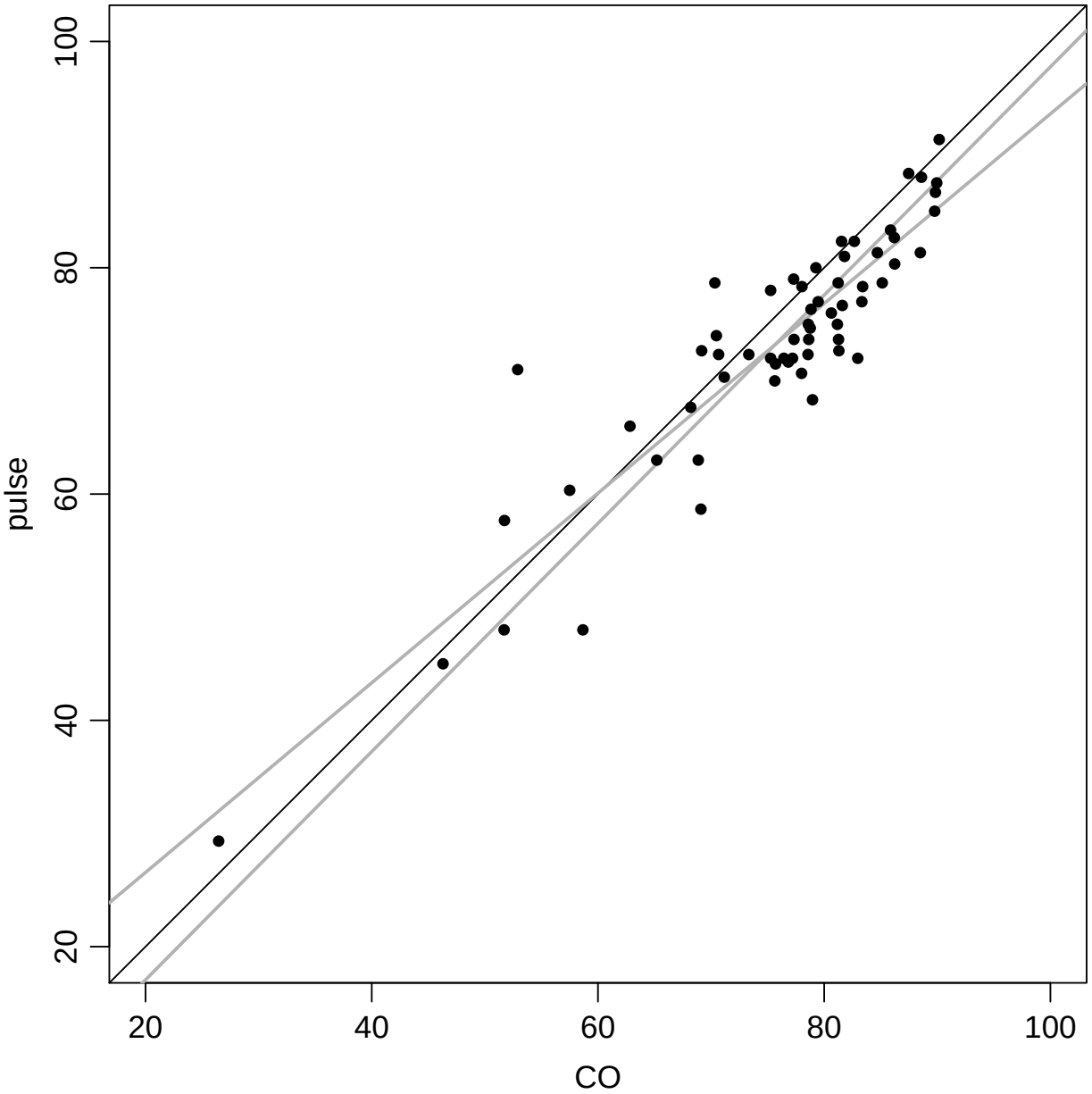
The posterior dist'n of $(\alpha_m, \beta_m, \mu_i)$ is singular.

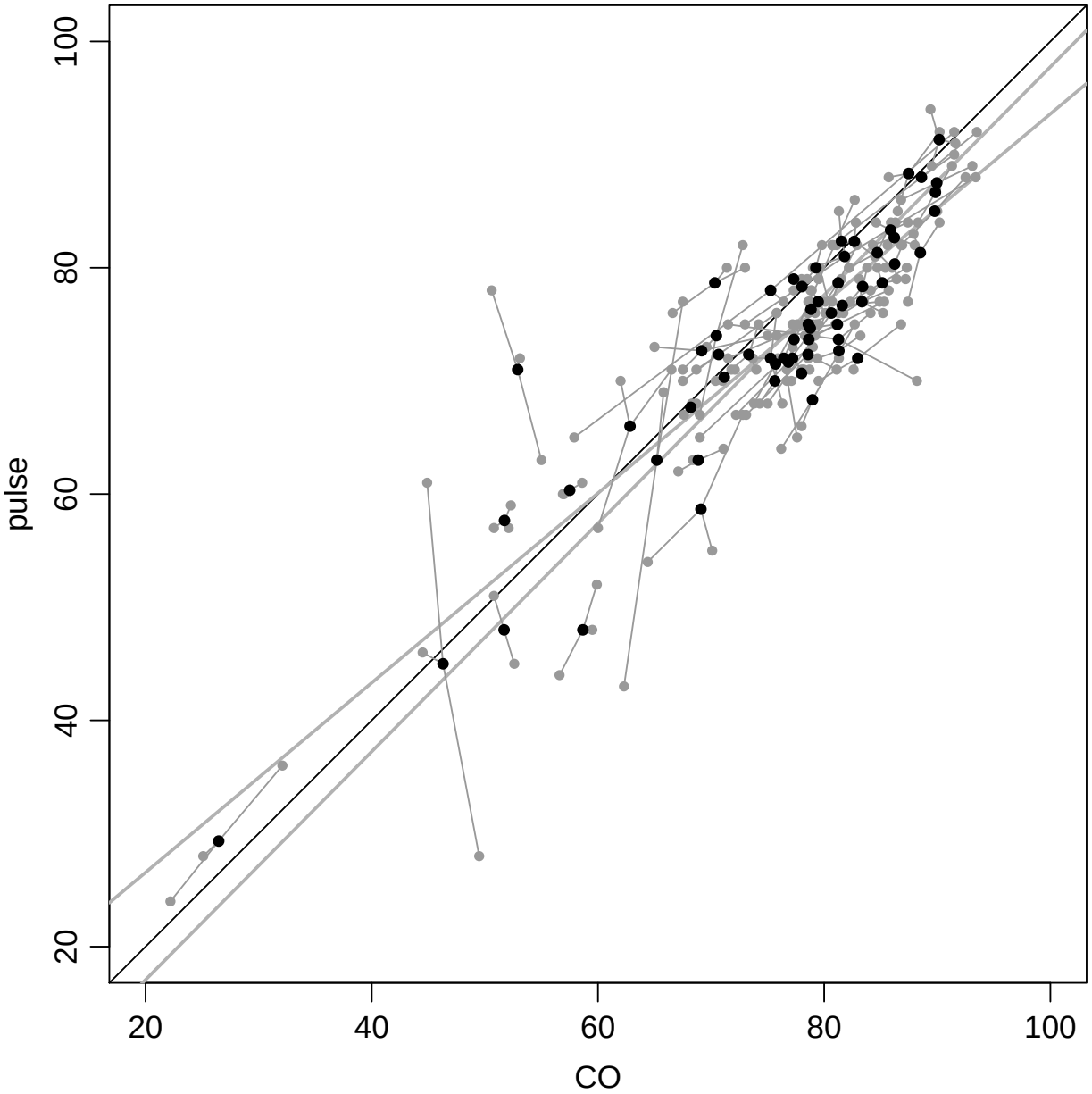
But the relevant translation quantities are identifiable:

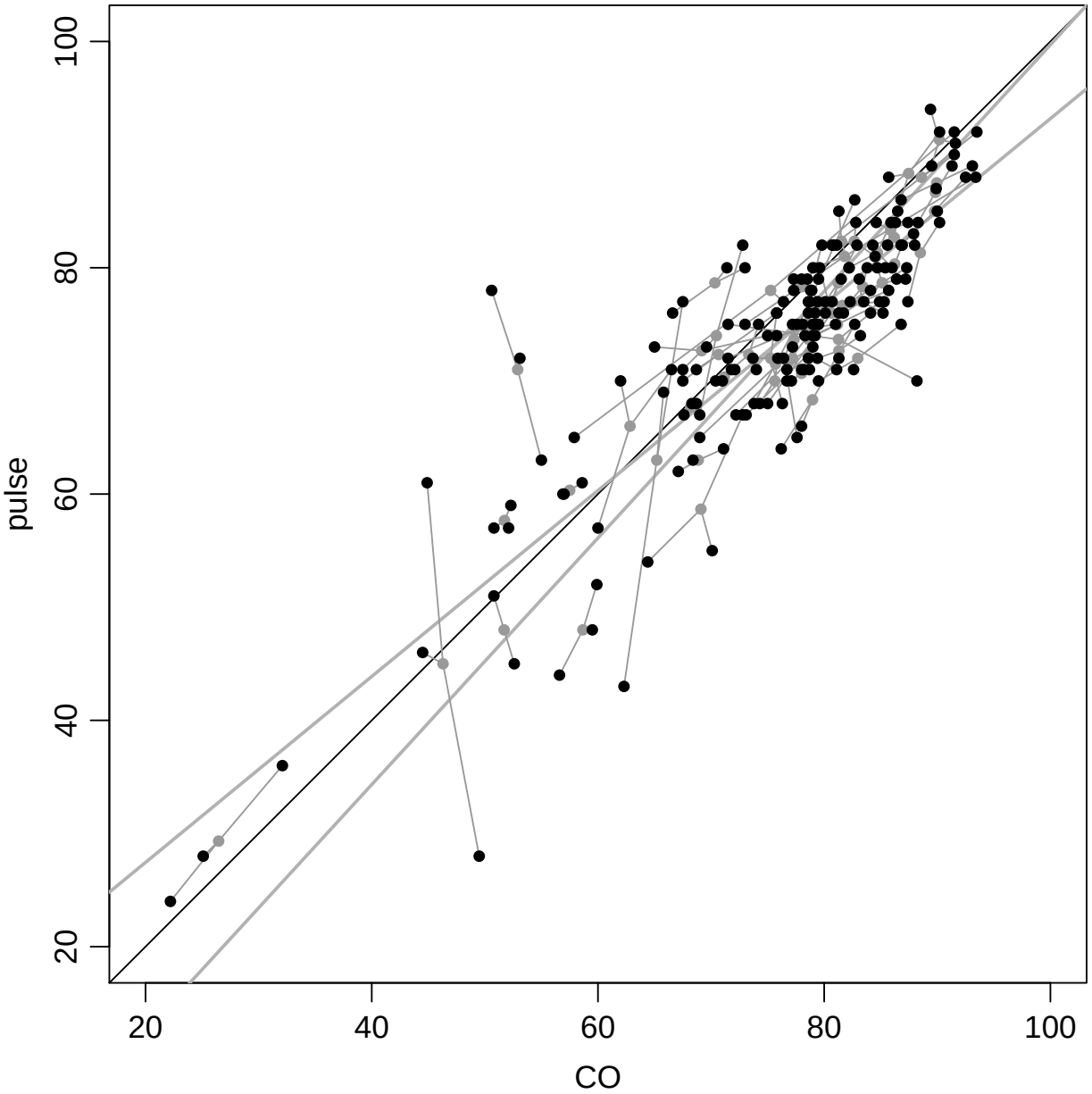
$$\begin{aligned} \alpha_{2\cdot1} &= \alpha_2 - \alpha_1\beta_2/\beta_1 \\ \beta_{2\cdot1} &= \beta_2/\beta_1 \end{aligned}$$

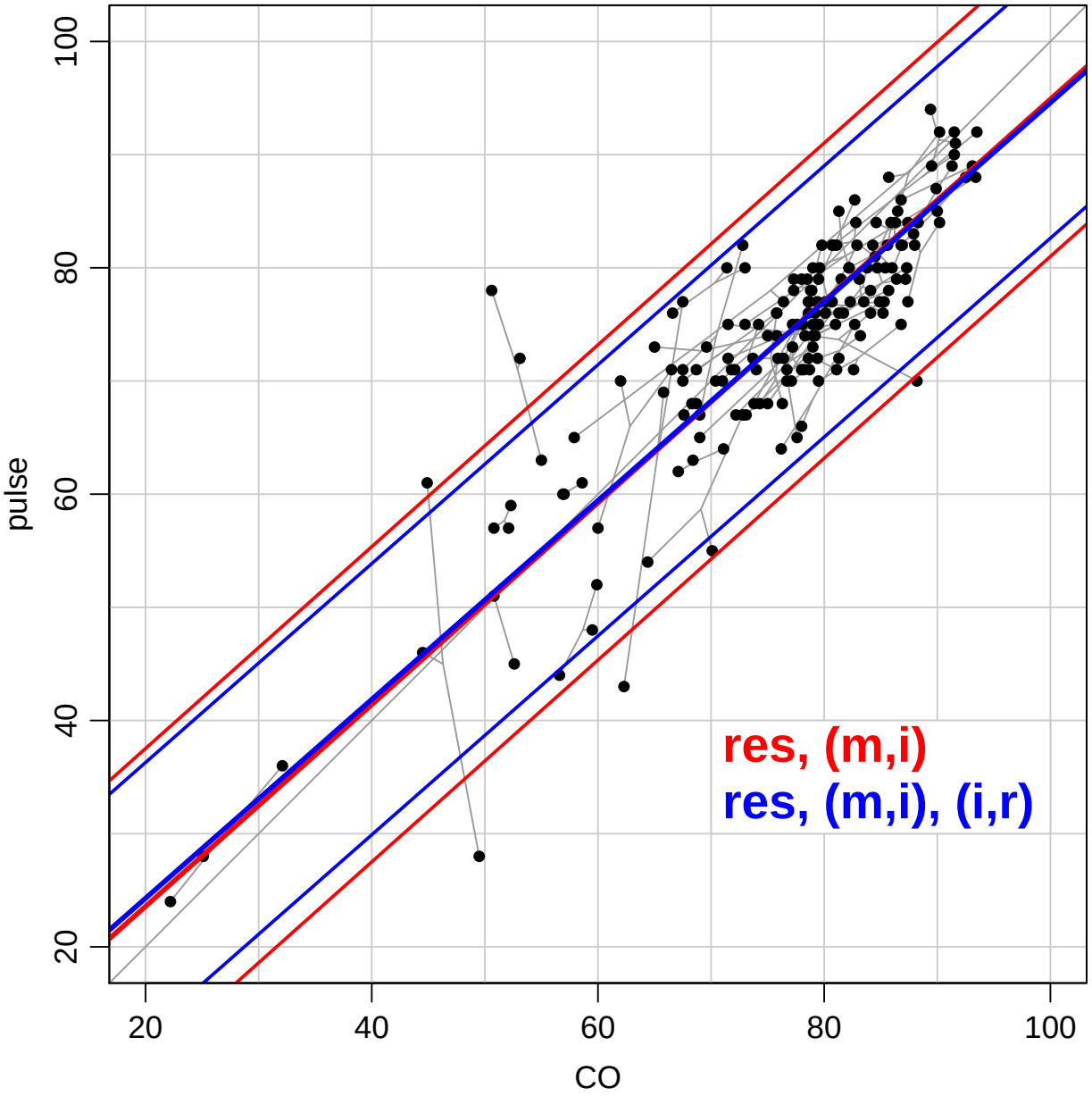
So are the variance components.

Posterior medians used to devise prediction equations with limits.

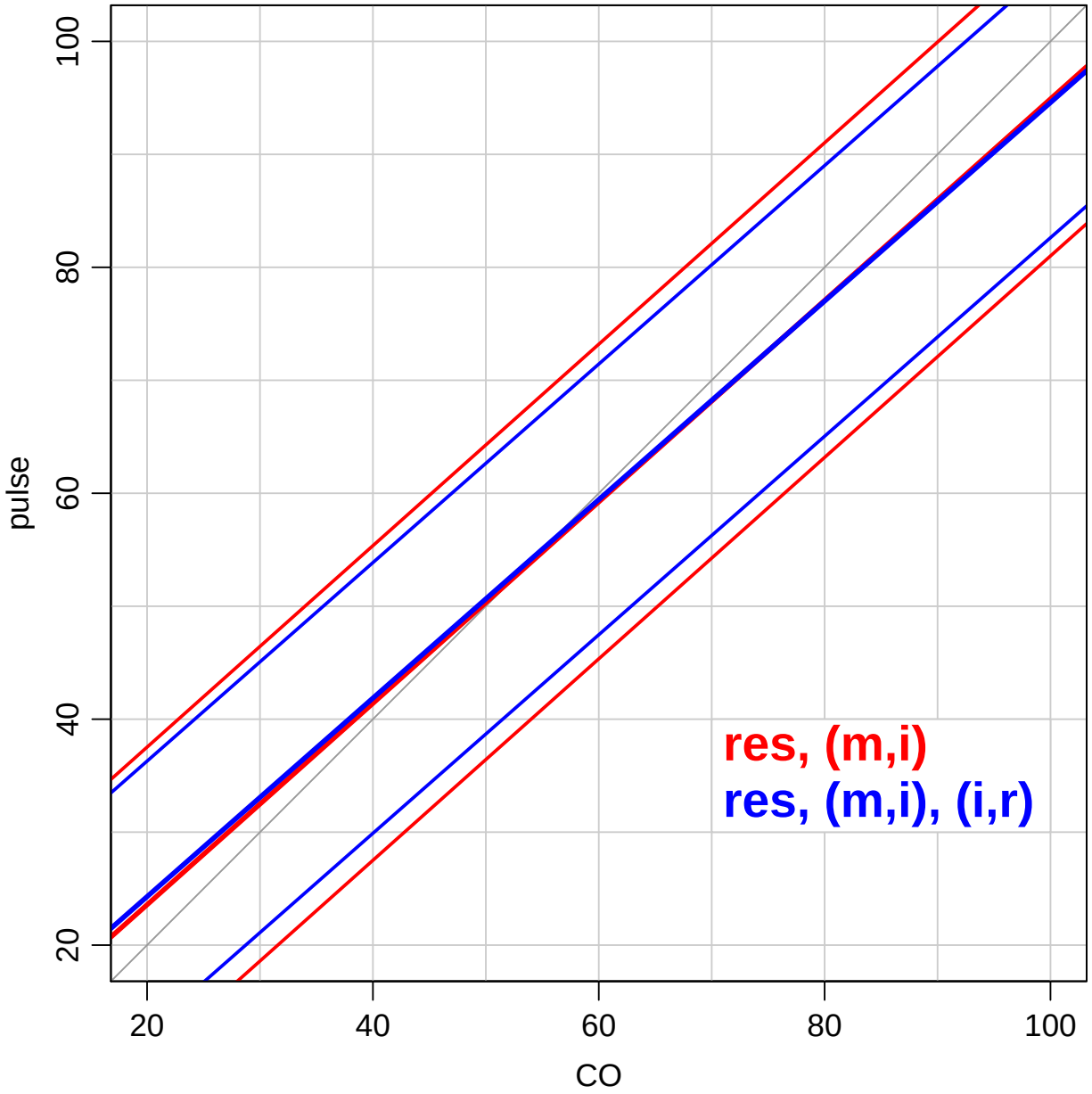

pulse
CO
20
40
60
80
100
20
40
60
80
100


pulse
CO
20
40
60
80
100

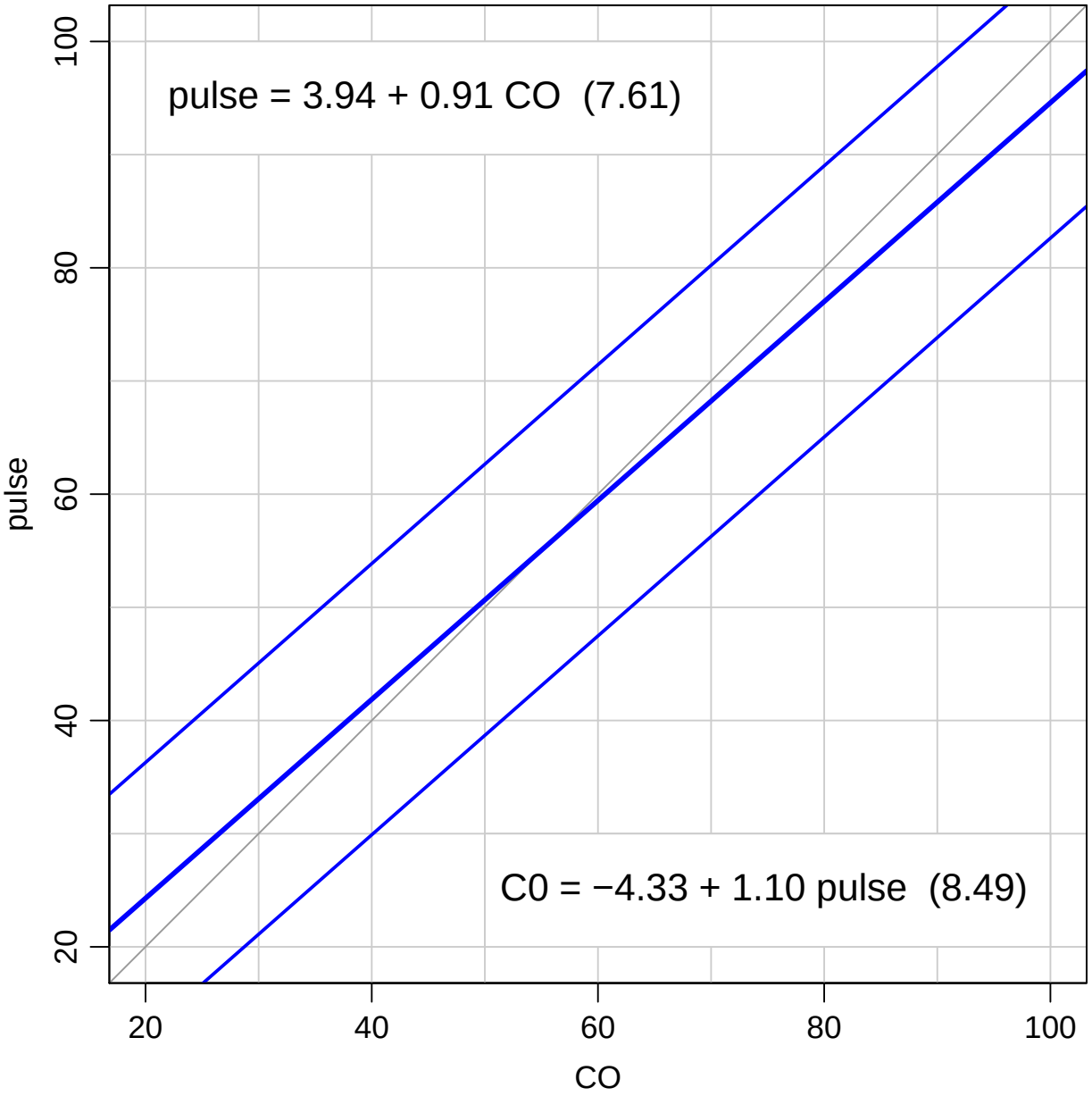

pulse
CO
20
40
60
80
100
20
40
60
80
100



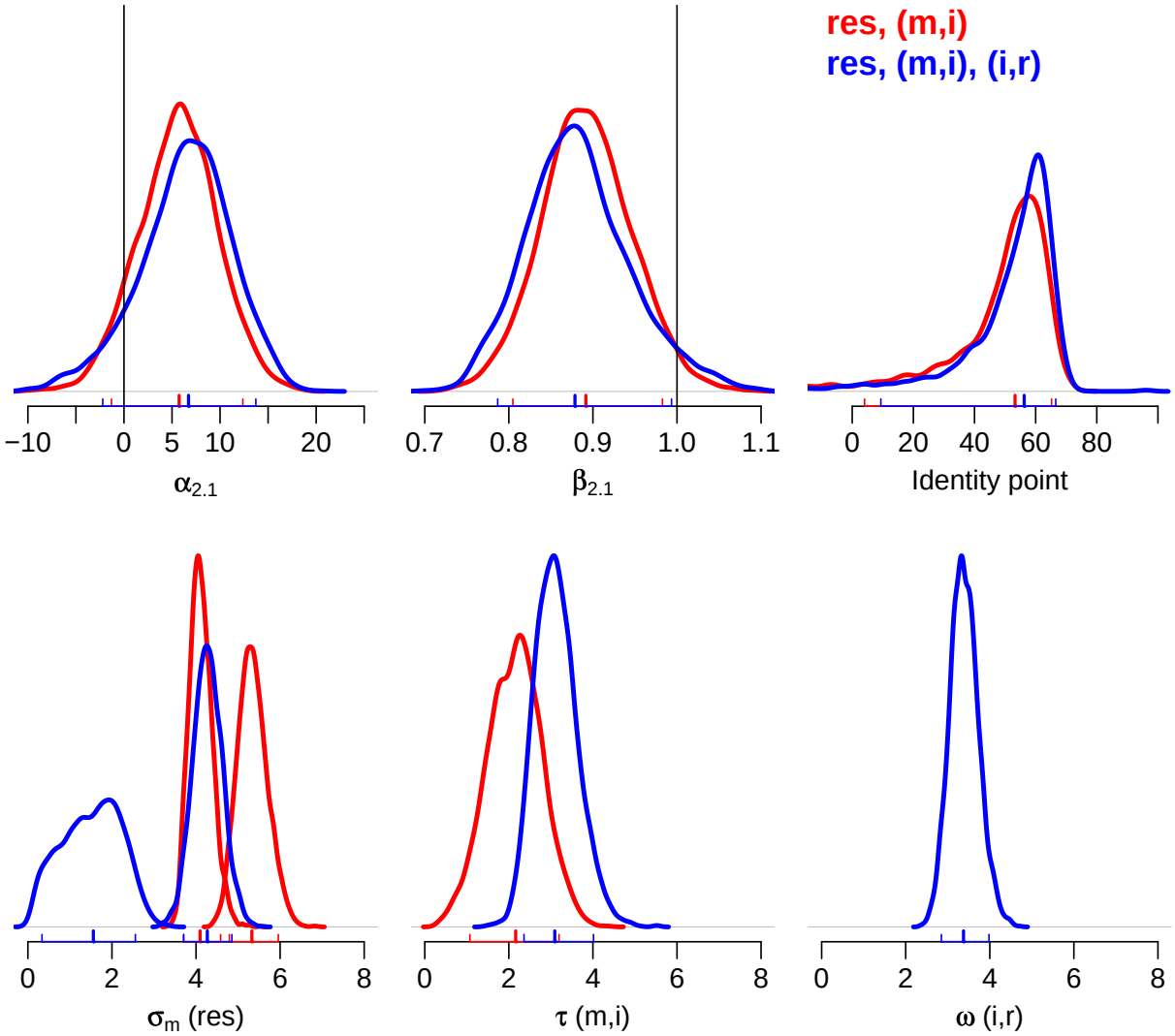
$$y_{mir} = \alpha_m + \beta_m(\mu_i + a_{ir} + c_{mi}) + e_{mir}$$



$$y_{mir} = \alpha_m + \beta_m(\mu_i + a_{ir} + c_{mi}) + e_{mir}$$



$$y_{mir} = \alpha_m + \beta_m(\mu_i + a_{ir} + c_{mi}) + e_{mir}$$



$$y_{mir} = \alpha_m + \beta_m(\mu_i + a_{ir} + c_{mi}) + e_{mir}$$

Morale

- Use a proper model for your problem.
- Get the exchangeability right.
- Report the model in a useful way.

The MethComp package for R

Implemented model:

$$y_{mir} = \alpha_m + \beta_m(\mu_i + a_{ir} + c_{mi}) + e_{mir}$$

- Replicates required.
- R2WinBUGS is required.
- Dataframe with variables meth, item, repl and y.
- The function MethComp writes a BUGS-program, initial values and data to files.
- Runs WinBUGS and sucks results back in to **R**, and gives a nice overview of the conversion equations.

Example output: Oximetry

```
> ox.mi.ir <- MethComp( ox, n.iter=5000 )
> ox.mi.ri

Comparison of 2 methods, using 354 measurements
on 61 individuals, with up to 3 replicate measurements.
( 2 * 61 * 3 = 366 ):

No. individuals with measurements on each method:
       # replicates
Method  1 2  3 Sum
  CO    1 4 56  61
  pulse 1 4 56  61
```

Example output: Oximetry

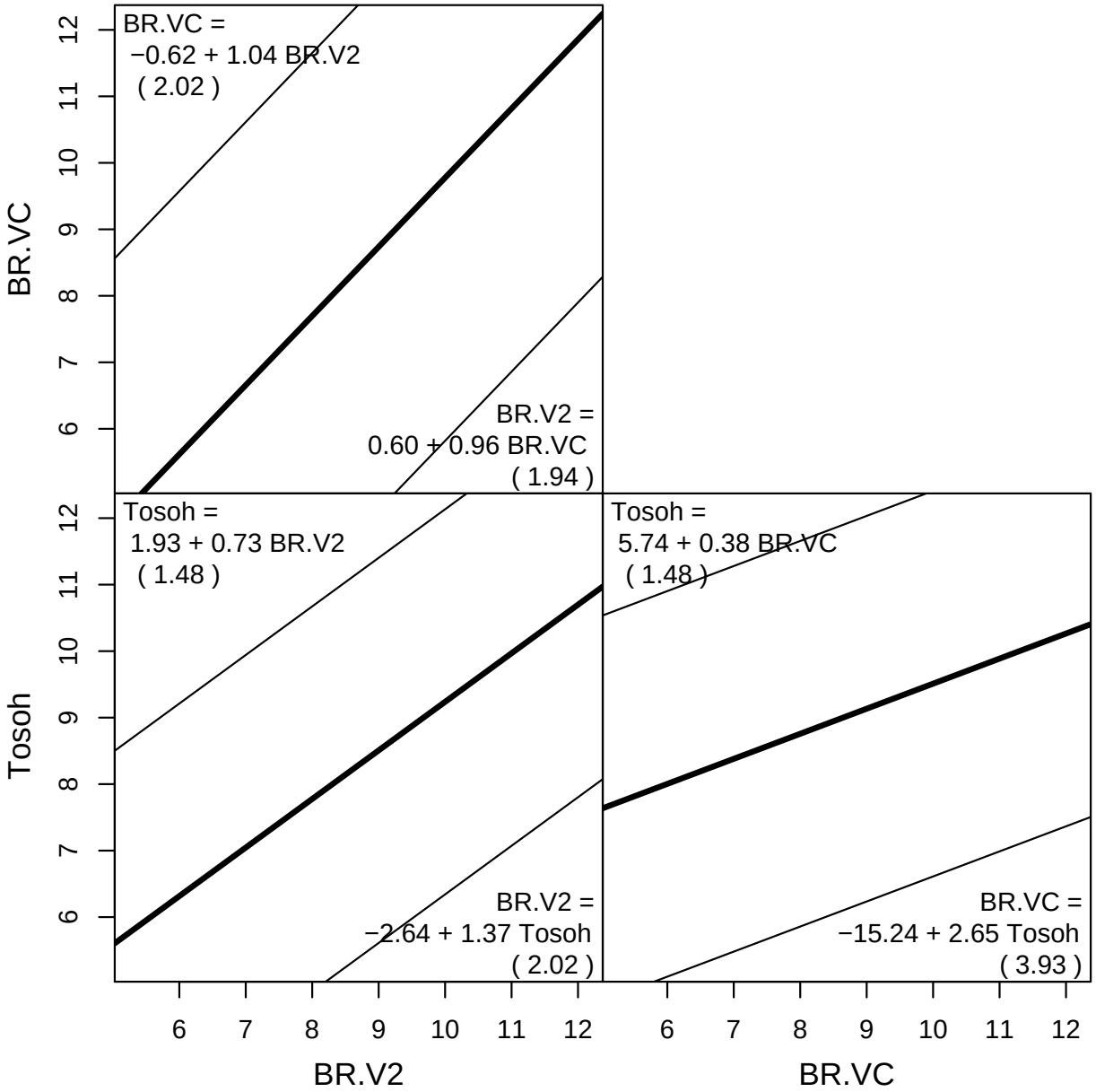
```
Conversion formulae (y_to = alpha + beta*y_from +/- 2*sd.pred):
    From:     CO                          pulse
           alpha   beta sd.pred      alpha   beta sd.pred
To:
CO         0.000  1.000   4.266     -4.328  1.098   8.487
pulse      3.939  0.911   7.606      0.000  1.000   5.534

Variance components (standard deviations):
                      50%   2.5%   97.5%     0%     100%
sigma.mir[CO]      1.6285 0.2092  2.8274 0.0724  3.4330
sigma.mir[pulse] 4.2580 3.5390  4.9725 3.0670  5.9800
sigma.mi[CO]       4.8043 2.7504 13.3685 2.2597 17.6134
sigma.mi[pulse]  4.3123 2.4981 11.5859 1.9248 13.2186
sigma.ir[CO]       3.9213 3.1452  4.7038 2.7289  5.3129
sigma.ir[pulse]  3.5433 2.7542  4.3516 2.2610  4.8723
```

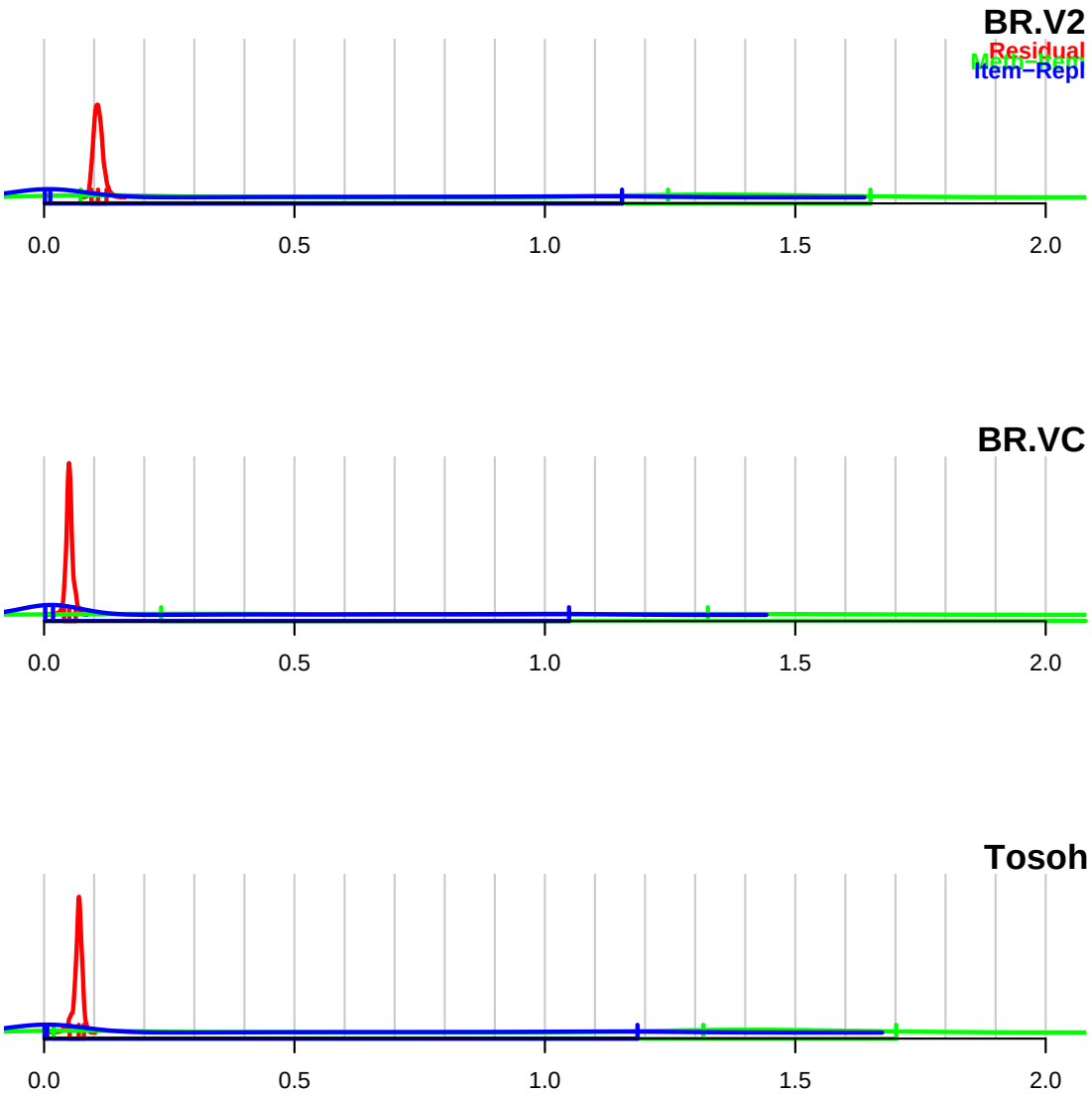
HbA_{1c} - 3 different instruments

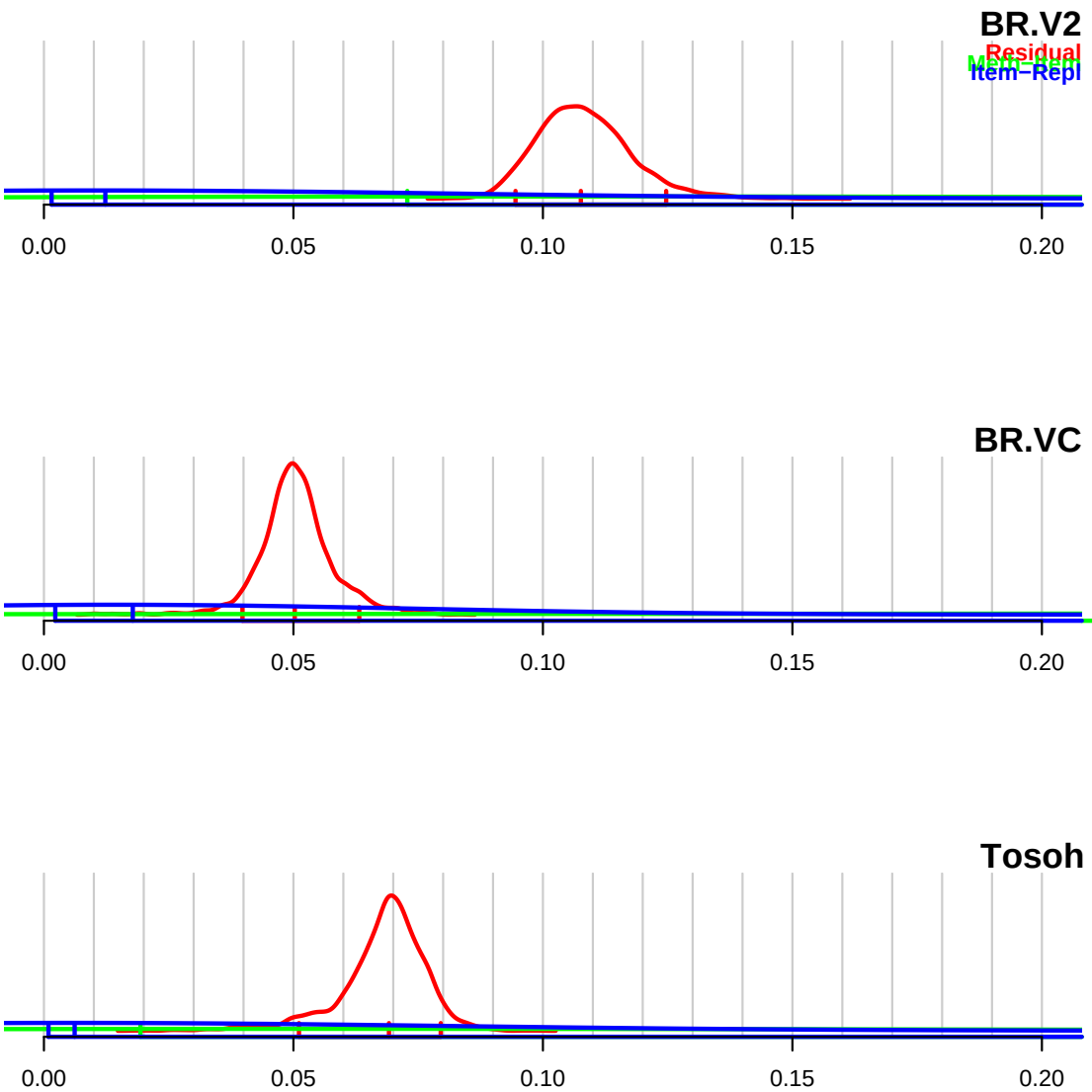
```
> hbv.mi.ir <- MethComp( hbv, n.iter=5000 )
> print( hbv.mi.ir, across=FALSE )

Conversion formula:
 y_to = alpha + beta * y_from +/- 2*sd.pred:
            From:  BR.V2  BR.VC  Tosoh
To:
BR.V2 alpha        0.000 -1.627  1.413
      beta         1.000  1.154  0.946
      sd.pred      0.254  2.079  2.099
BR.VC alpha        1.417  0.000  2.412
      beta         0.867  1.000  0.819
      sd.pred      1.800  0.164  1.927
Tosoh alpha       -1.591 -3.144  0.000
      beta         1.057  1.220  1.000
      sd.pred      2.145  2.249  0.156
```

HbA_{1c} - 3 different instruments

```
Variance components (standard deviations):
                      50%   2.5%  97.5%     0%   100%
sigma.mir[BR.V2] 0.2089 0.1816 0.2401 0.1614 0.2692
sigma.mir[BR.VC] 0.1074 0.0813 0.1286 0.0642 0.1467
sigma.mir[Tosoh] 0.0345 0.0006 0.0824 0.0004 0.0984
sigma.mi[BR.V2]  1.3495 1.0780 1.7742 0.9194 2.1615
sigma.mi[BR.VC]  1.3088 1.0498 1.6979 0.8615 2.1350
sigma.mi[Tosoh]  1.4416 1.0782 5.3653 0.9250 6.3534
sigma.ir[BR.V2]  0.1418 0.1037 0.1882 0.0855 0.2319
sigma.ir[BR.VC]  0.1239 0.0928 0.1572 0.0797 0.1827
sigma.ir[Tosoh]  0.1496 0.1231 0.1815 0.0950 0.2002
```


BR.VC =
−0.62 + 1.04 BR.V2
(2.02)
BR.V2 =
0.60 + 0.96 BR.VC
(1.94)
Tosoh =
1.93 + 0.73 BR.V2
(1.48)
Tosoh =
5.74 + 0.38 BR.VC
(1.48)
BR.V2 =
−2.64 + 1.37 Tosoh
(2.02)
BR.VC =
−15.24 + 2.65 Tosoh
(3.93)
BR.VC
Tosoh
BR.V2
BR.VC
6
7
8
9
10
11
12


BR.V2
Residual
Meth−Repl
Item−Repl
0.0
0.5
1.0
1.5
2.0
BR.VC
0.0
0.5
1.0
1.5
2.0
Tosoh
0.0
0.5
1.0
1.5
2.0


BR.V2
Residual
Meth−Meth
Item−Repl
0.00
0.05
0.10
0.15
0.20
BR.VC
0.00
0.05
0.10
0.15
0.20
Tosoh
0.00
0.05
0.10
0.15
0.20

The MethComp package

Also (currently) contains:

- `BA.plot` — make a Bland-Altman plot and compute limits of agreement.
- `BA.est` — estimates in the variance component model for the constant bias situation.
- `Deming` — regression with errors in both variables.
  A `.pdf` with a detailed derivation of the formulae (by Anders C Jensen) is included in the package too.
- A number of example data sets, amongst them all examples from [**?**].

References

MethComp (0.1.19) is available at:
http://www.biostat.ku.dk/~bxc/MethComp
— but not on CRAN yet.