Method Comparison Studies in Practise

Bendix Carstensen

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

28–30 November 2007
Dept. of Medical Epidemiology and Biostatistics,
Karolinska Institutet, Stockholm

Introduction to computing

Wednesday 28 November 2007, afternoon

Bendix Carstensen

Method Comparison Studies in Practise
28–30 November 2007
Dept. of Medical Epidemiology and Biostatistics,
Karolinska Institutet, Stockholm

Course structure

The course is both theoretical and practical, i.e. the aim is to convey a basic understanding of the problems in method comparison studies, but also to convey practical skills in handling the statistical analysis.

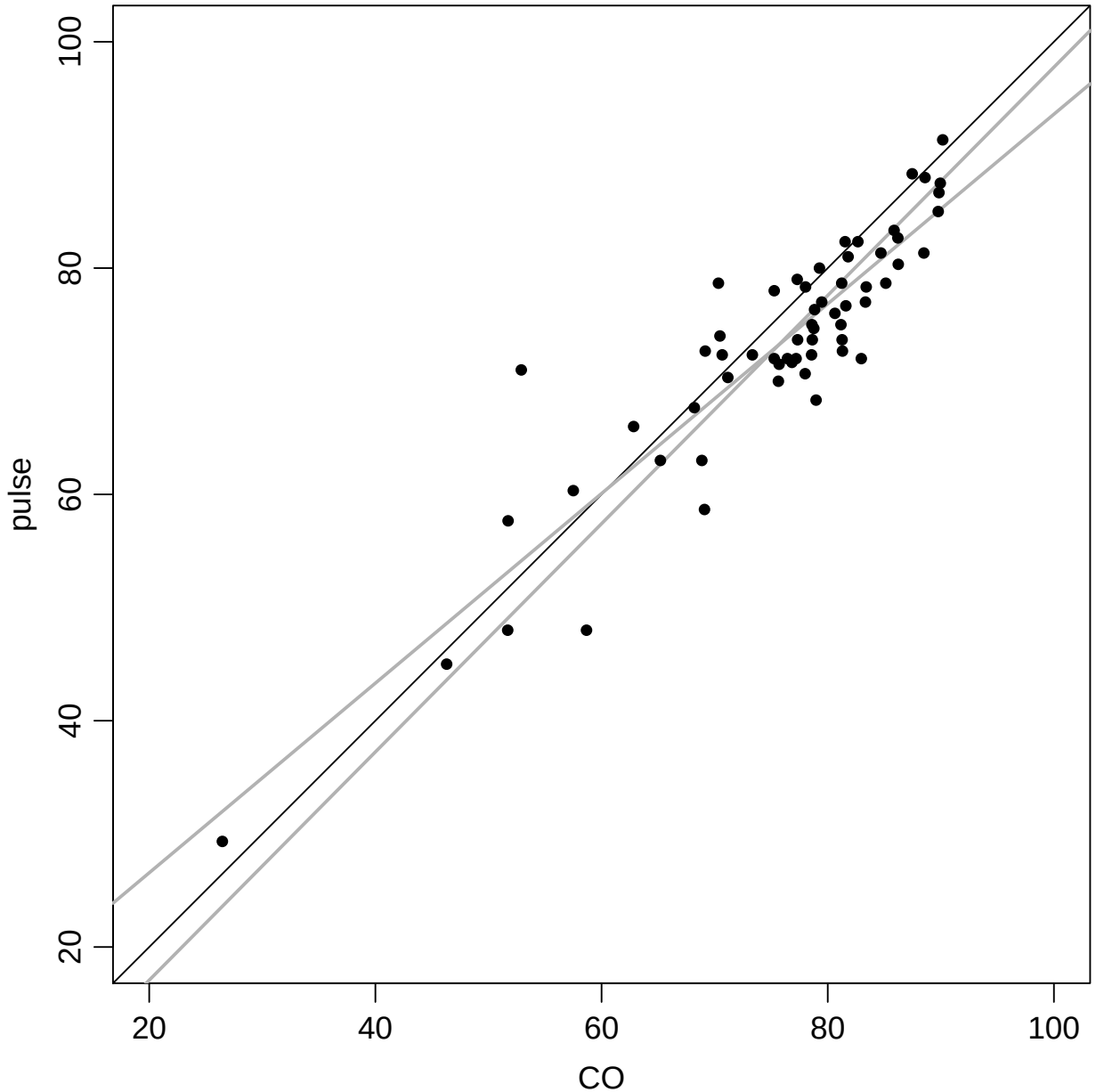
- **R** for data manipulation ad graphics.
- `WinBUGS` for estimation in non-linear variance component models.

Software considerations

- **R**, SAS and Stata all have interfaces to WinBUGS.
- But **R** have more flexible graphical facilities.
- The `MethComp` is written for **R**.

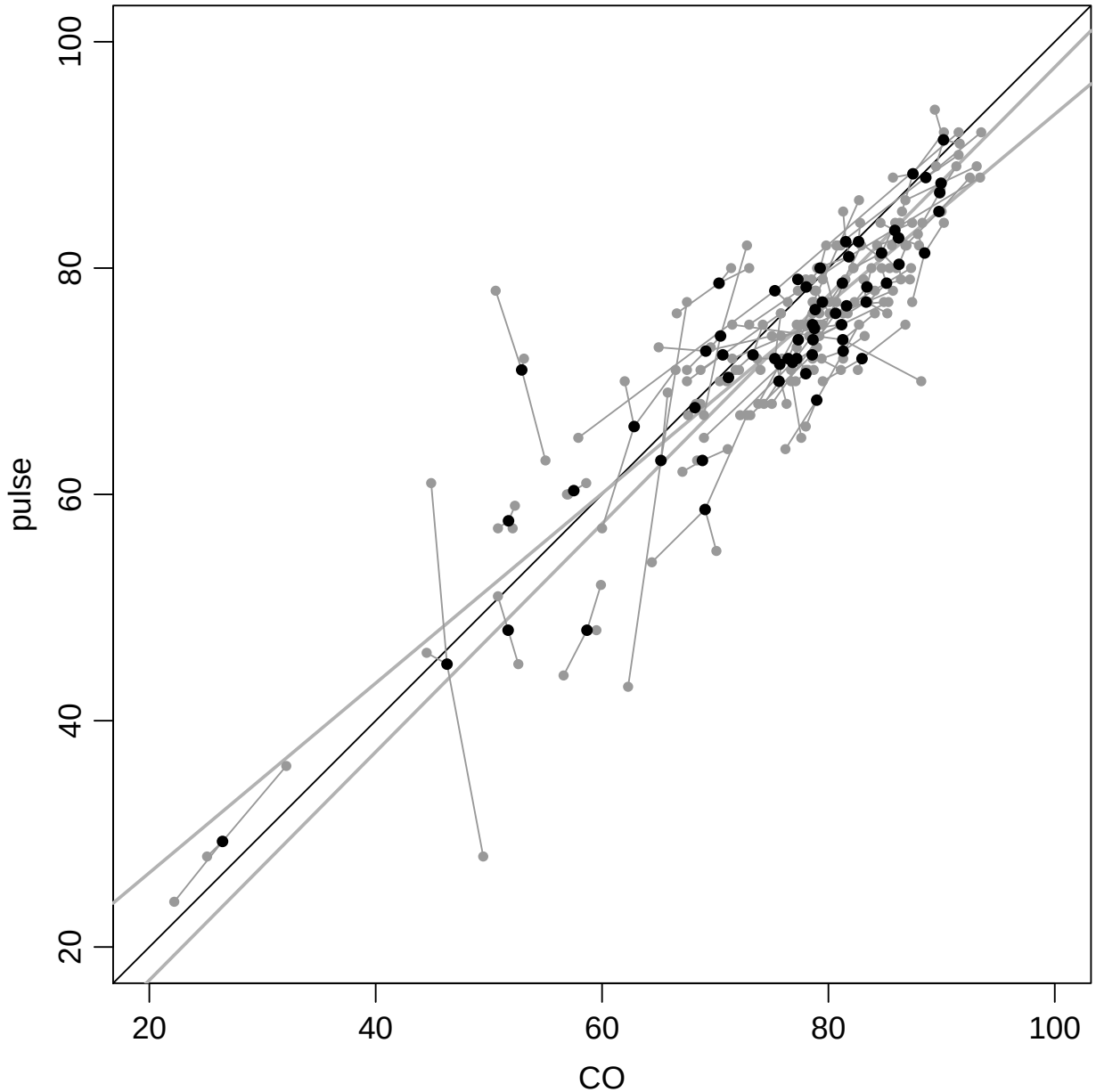
Therefore we use **R** in this course.

Oximetry data



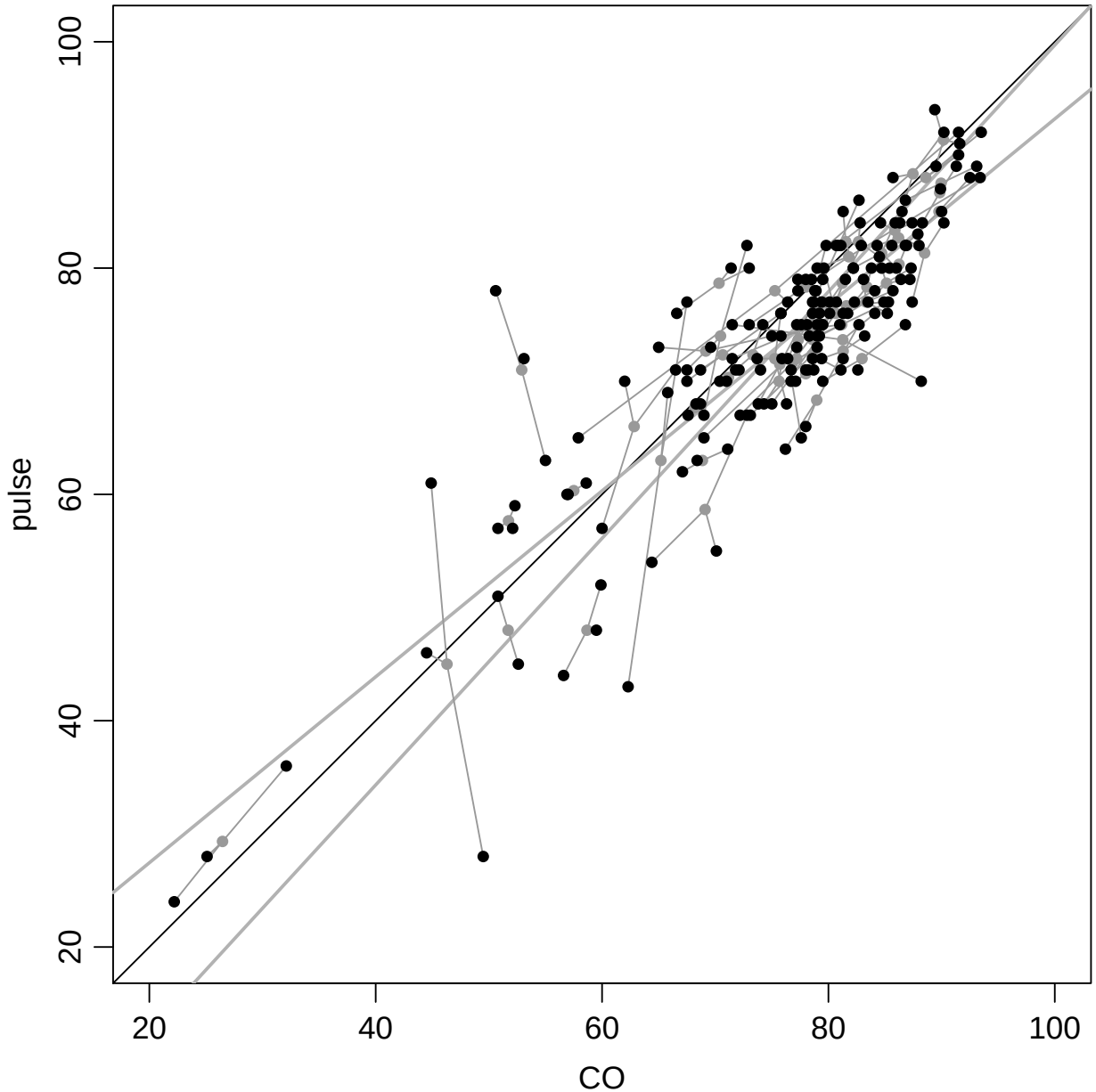
Means over replicates.

Oximetry data



Linked replicates.

Oximetry data



Linked replicates.

How it works

Example data sets are included in the `MethComp` package. Contains the following variables.

- `meth` — method
- `item` — item, person, individual, sample
- `repl` — replicate (if present)
- `y` — the actual measurement

— or rather *should* in order for the functions in `MethComp` to work.

How it looks

```
> 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)
Note:
 Replicate measurements are ta
  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
```

Analyses/plots in this course

- Scatter plots.
- Bland-Altman plots ($y - x$ vs. $(x + y)/2$)
- Limits of agreement.
- Models with constant bias.
- Models with linear bias.
- Conversion formulae between methods (single replicates)
- Plots of converison equations.
- Graphical reporting of variance components.

Requirements

- **R** for data manipulation and graphics:
- `Tinn-R` convenience editior with syntax highlighting for **R**.
- `nlme`-package for variance component models — constant bias.
- `WinBUGS` for fitting models with linear bias (non-linear variance component models, over-parametrized).

All of it works from within **R**.

Functions in the `MethComp` package

4 broad categories of functions in `MethComp`:

- Graphical — just exploring data.
- Data manipulation — reshaping and changing. Simulation.
- Analysis function — fitting models to data.
- Reporting functions — displaying the results from analyses.

Graphical functions

- `BA.plot` Makes a Bland-Altman plot of two methods from a data frame with method comparison data, and computes limits of agreement. The plotting etc is really done by a call to
- `BlandAltman` Draws a Bland-Altman plot and computes limits of agreement.
- `plot.meth` Plots all methods against all other, both as a scatter plot and as a Bland-Altman plot.
- `bothlines` Adds regression lines of y on x and vice versa to a scatter plot.

Data manipulating functions

- `make.repl` Generates a `repl` column in a data frame with columns `meth`, `item` and `y`.
- `perm.repl` Randomly permutes replicates within (method,item) and assigns new replicate numbers.
- `to.wide` Transforms a data frame in the long form to the wide form.
- `to.long` Reverses the result of `to.wide`.
- `tab.repl` Tabulates replicates by methods and items.
- `sim.meth` Simulates a dataset from a method comparison experiment for given parameters for bias, exchangeability and variances.

Analysis functions

- `Deming` Performs Deming regression, i.e. regression with errors in both variables.
- `BA.est` Estimates in the variance components models underlying the concept of limits of agreement, and returns the bias and the variance components. Assumes constant bias between methods.
- `MethComp` Estimates via `BUGS` in the general model with non-constant bias (and in the future) possibly non-constant standard deviations of the variance components. Produces a `MethComp` object.

Reporting functions

These functions all take a `MethComp` object as input.

- `print.MethComp` Prints a table of conversion equation between methods analyzed, with prediction standard deviations. Also gives summaries of the posteriors for the parameters that constitute the conversion algorithms.
- `plot.MethComp` Plots the conversion lines between methods with prediction limits.
- `plot.VarComp` Plots smoothed posterior densities for the variance component estimates.

Does it work?

You should get something reasonable out of this:

```
library(MethComp)
data(ox)
plot.meth(ox)
plot.meth(perm.repl(ox))
BA.plot(ox)
BA.est(ox)
BA.est(perm.repl(ox))
MethComp(ox,code.only=TRUE)
m1 <- MethComp(ox)
print(m1)
plot(m1)
plot.VarComp(m1)
```

— if it works we are ready for tomorrow!

Any practical examples?

Comparing two methods with one measurment on each

Thursday 29 November 2007, morning

Bendix Carstensen

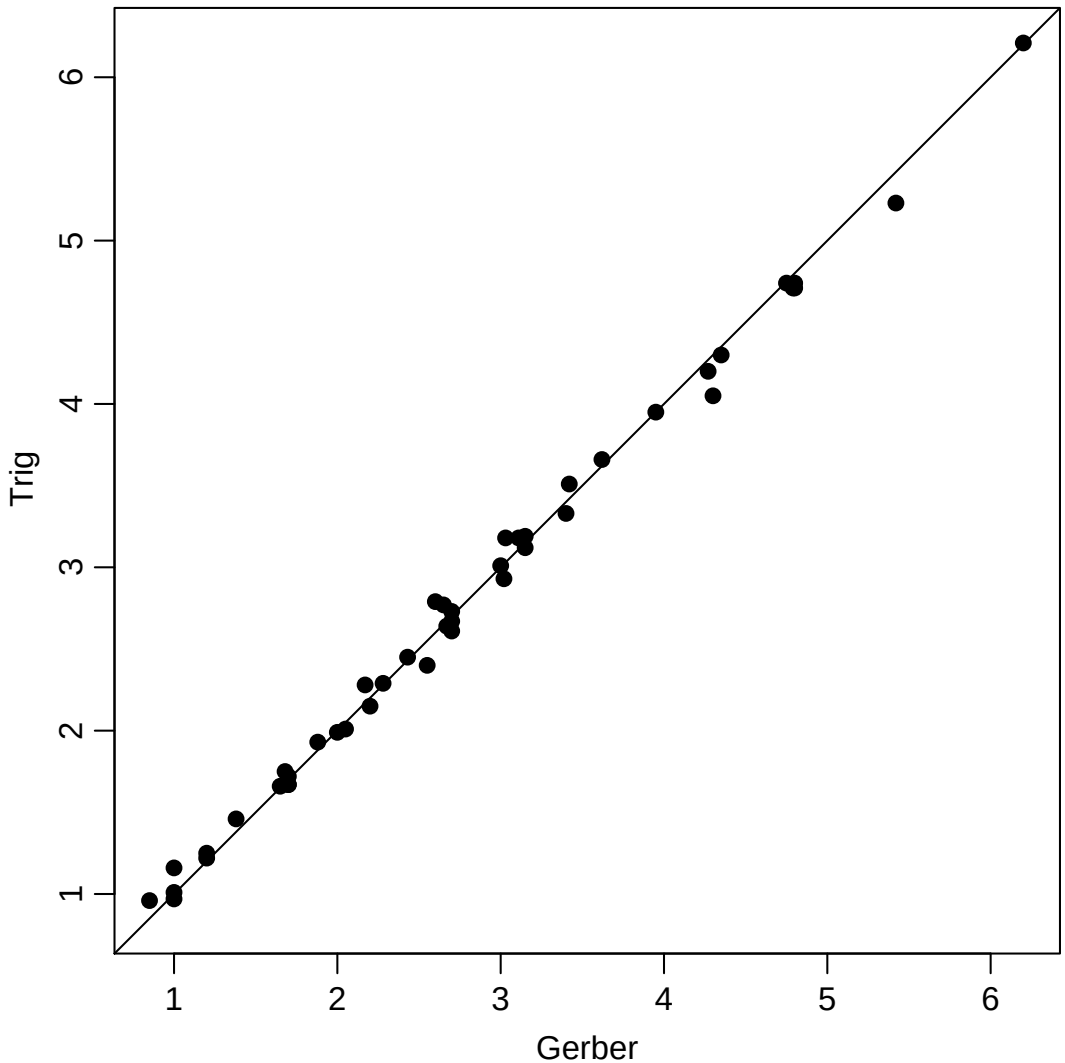
Method Comparison Studies in Practise
28–30 November 2007
Dept. of Medical Epidemiology and Biostatistics,
Karolinska Institutet, Stockholm

Comparing measurement methods

General questions:

- 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 can we convert from one method to another?

Two methods for measuring fat content in human milk:



The relationship looks like:

$$y_1 = a + by_2$$

Two methods — one measurement by each

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 measurement by method 1 and one by method 2.
[**?**, **?**]

Limits of agreement: Interpretation

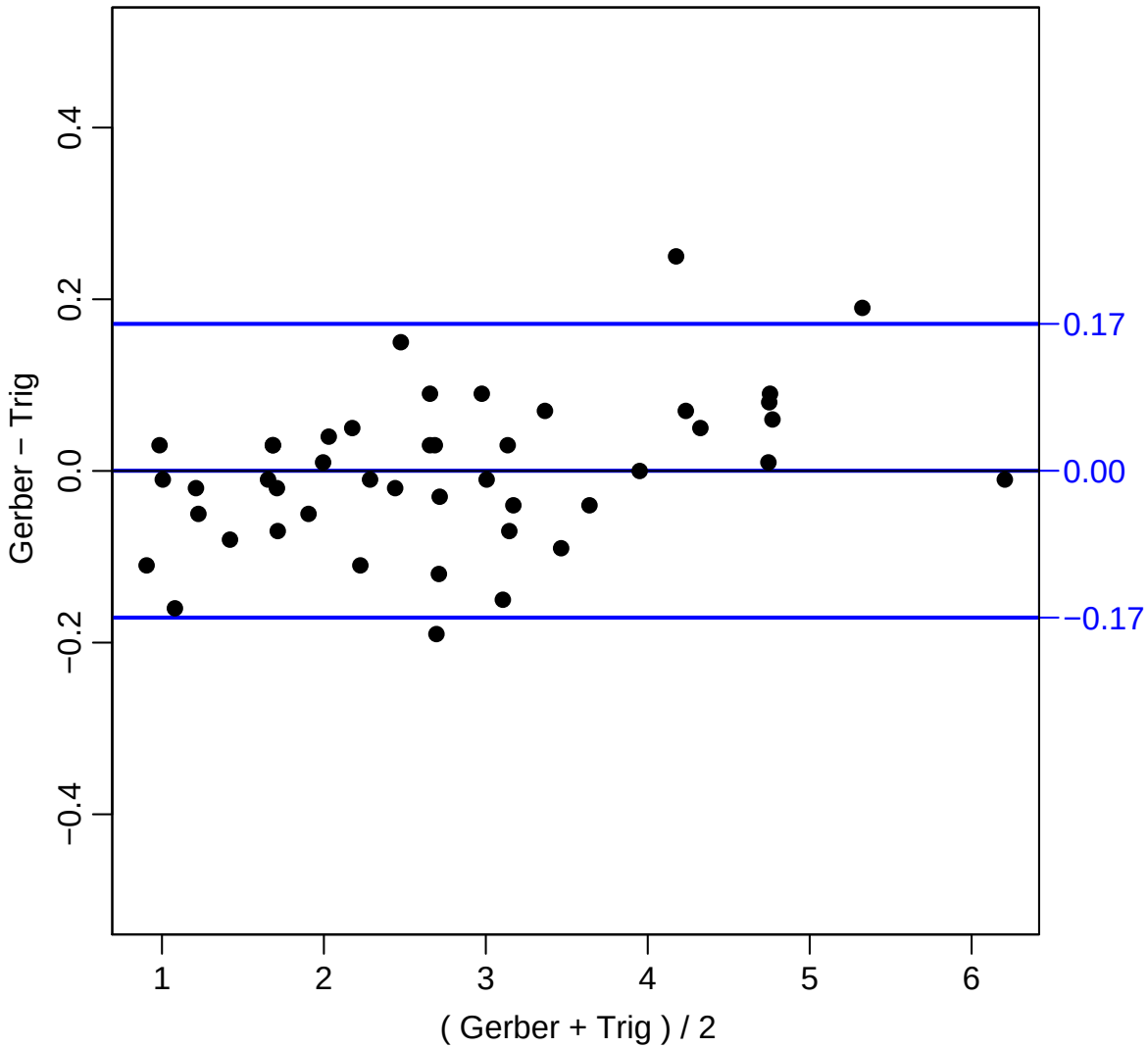
- If a new patient is measured **once** with each of the two methods, the difference between the two values will with 95% probability be within the limits of agreement.
- This is a **prediction** interval for a (future) difference.
- Requires a **clinical** input:
  Are the limits of agreement sufficiently narrow to make the use of either of the methods clinically acceptable?
- Is it relevant to test if the mean is 0?

Limits of agreement: Test?

Testing whether the difference is 0 is a bad idea:

- If the study is sufficiently small this will be accepted even if the difference is important.
- If the study is sufficiently large this will be rejected even if the difference is clinically irrelevant.
- It is an **eqivalence** problem:
  Clinical input is required!

Limits of agreement:



Plot differences (D_i) versus averages (A_i).

Model in "Limits of agreement"

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

$$\begin{aligned} y_{mi} &= \alpha_m + \mu_i + e_{mi} \\ e_{mi} &\sim \mathcal{N}(0, \sigma_m^2) \quad \text{measurement error} \end{aligned}$$

- Two-way analysis of variance model, with unequal variances in columns.
- Different variances are not identifiable without replicate measurements for $M = 2$ because the variances cannot be separated.

Limits of agreement:

Unequal variances induce correlation between D_i and A_i:

$$\text{cov}(D_i, A_i) = \frac{1}{2}(\sigma_x^2 - \sigma_y^2) \neq 0 \quad \text{if } \sigma_x \neq \sigma_y$$

In correlation terms:

$$\rho(D, A) = \frac{1}{2}\frac{\sigma_x^2 - \sigma_y^2}{\sigma_x^2 + \sigma_y^2}$$

i.e. the correlation depends on whether the difference between the variances is large relative to the sizes of the two.

Limits of agreement:

Usually interpreted as the likely differnence between two future measurements, one with each method:

$$\widehat{y_2 - y_1} = \hat{D} = \alpha_2 - \alpha_1 \pm 1.96\,\mathrm{s.d.}(D)$$

But it can of course also be converted to a prediction interval for y_2 given y_1:

$$\hat{y}_2|y_1 = \alpha_2 - \alpha_1 + y_1 \pm 1.96\,\mathrm{s.d.}(D)$$

Repeatability and reproducibility

Thursday 29 November 2007, morning

Bendix Carstensen

Method Comparison Studies in Practise
28–30 November 2007
Dept. of Medical Epidemiology and Biostatistics,
Karolinska Institutet, Stockholm

Accuracy of a measurement method

- Repeatability:
  The accuracy of the method under exactly similar circumstances; i.e. the same lab, the same technician, and the same day.
  (**Repeata**bility conditions)
- Reproducibility:
  The accuracy of the method under comparable circumstances, i.e. the same machinery, the same kit, but possibly different days or laboratories or technicians.
  (**Reproduci**bility conditions)

Quantification of accuracy

- Upper limit of a 95% confidence interval for the difference between two measurments.
- Suppose the variance of the measurement is σ^2:

$$\mathrm{var}(y_{mi1} - y_{mi2}) = 2\sigma^2$$

i.e the standard error is $\sqrt{2}\sigma$, and a confidnece interval for the difference:

$$0 \pm 1.96 \times \sqrt{2}\sigma = 0 \pm 2.772\sigma \approx 2.8\sigma$$

- This is called the reproducibility coefficient or simply the reproducibility. (The number 2.8 is used as a convenient approximation).

Quantification of accuracy

- Where do we get the σ?
- Repeat measurements on the same item (or even better) several items.
- The conditions under which the repeat (replicate) measurements are taken determines whether we are estimating repeatability or reproducibility.
- In larger experiments we must consider the **exchangeability** of the replicates — i.e. which replicates are done under (exactly) similar conditions and which are not.

Comparing two methods with replicate measurements

Thursday 29 November 2007, morning

Bendix Carstensen

Method Comparison Studies in Practise
28–30 November 2007
Dept. of Medical Epidemiology and Biostatistics,
Karolinska Institutet, Stockholm

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.
- Assumes *exchangeability* of replicates.

Extension of the model: replicate measurements

$$
\begin{aligned}
y_{mir} &= \alpha_m + \mu_i + a_{ir} + c_{mi} + e_{mir} \\
&\quad \text{s.d.}(a_{ri}) = \omega \quad \text{— between replicates} \\
&\quad \text{s.d.}(c_{mi}) = \tau_m \quad \text{— “matrix”-effect} \\
&\quad \text{s.d.}(e_{mir}) = \sigma_m \quad \text{— measurement error}
\end{aligned}
$$

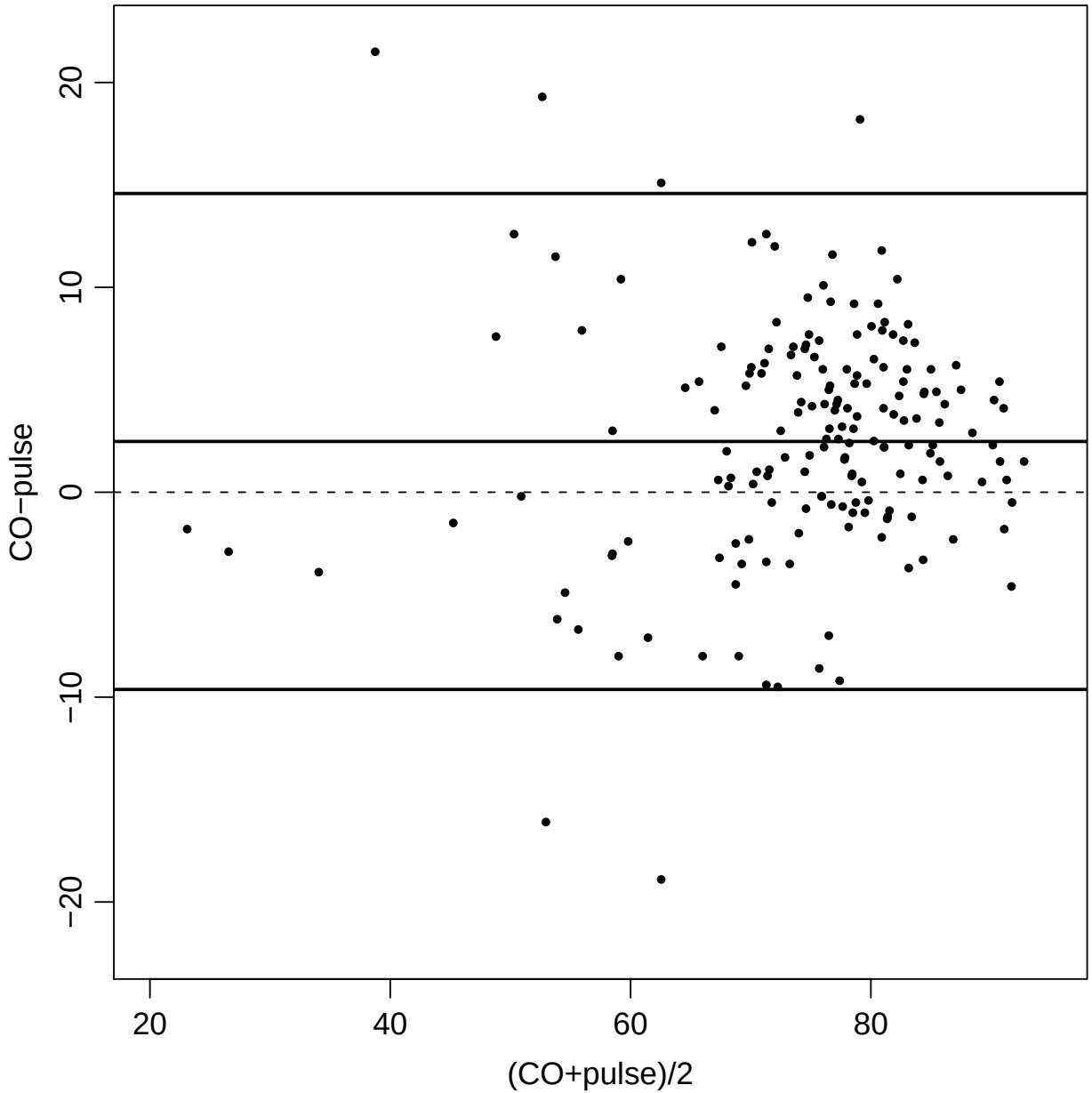
- Still assumes that the difference between methods is constant.
- Replicates are *linked* between methods: a_{ir} is common across methods, i.e. the first replicate on a person is made under similar conditions for all methods (i.e. at a specific day or the like).

Replicate measurements

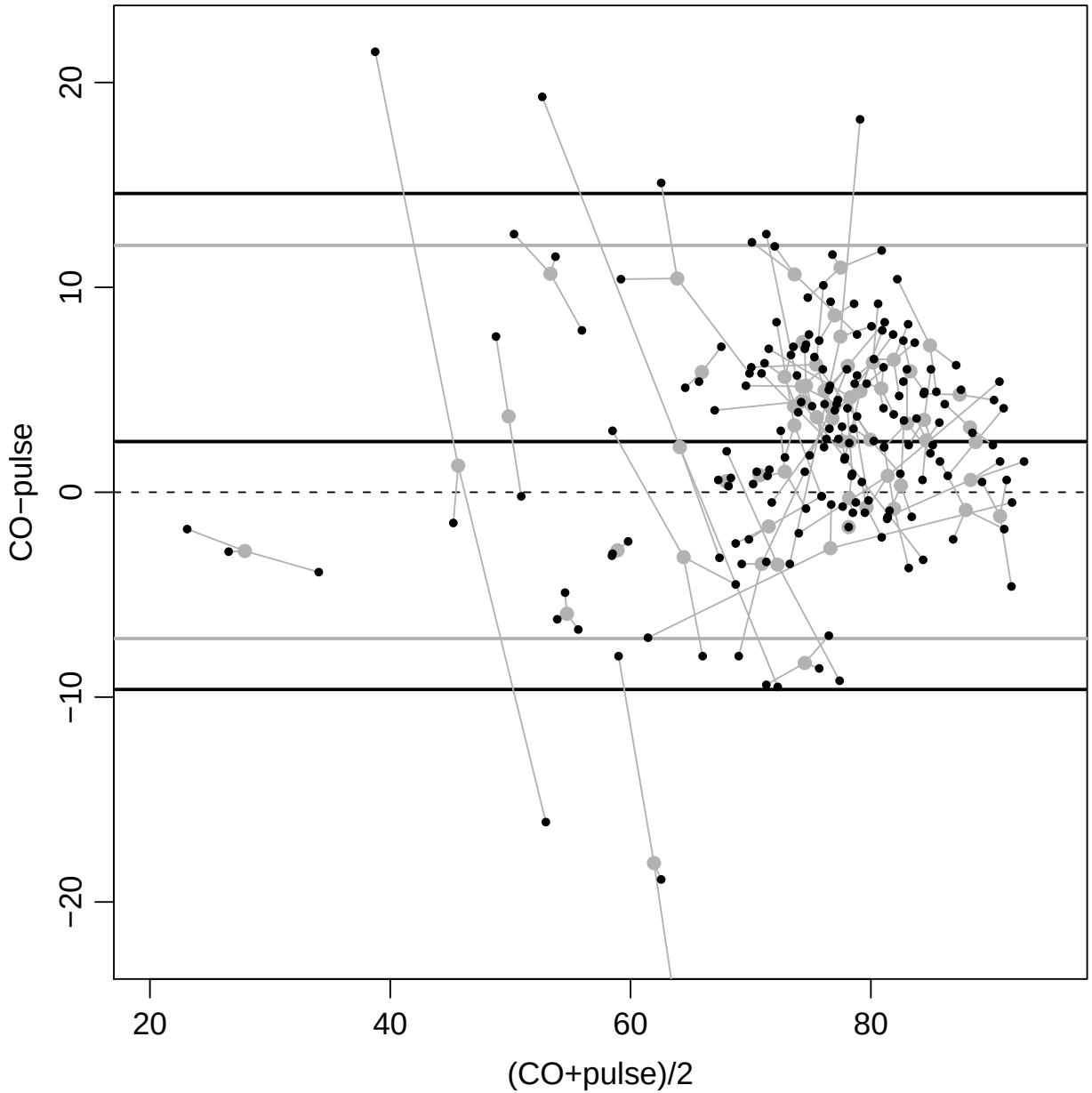
Two approaches to limits of agreement with replicate measurements:

1. Take means over replicates within each method by item stratum.
2. Replicates within item are taken as items.

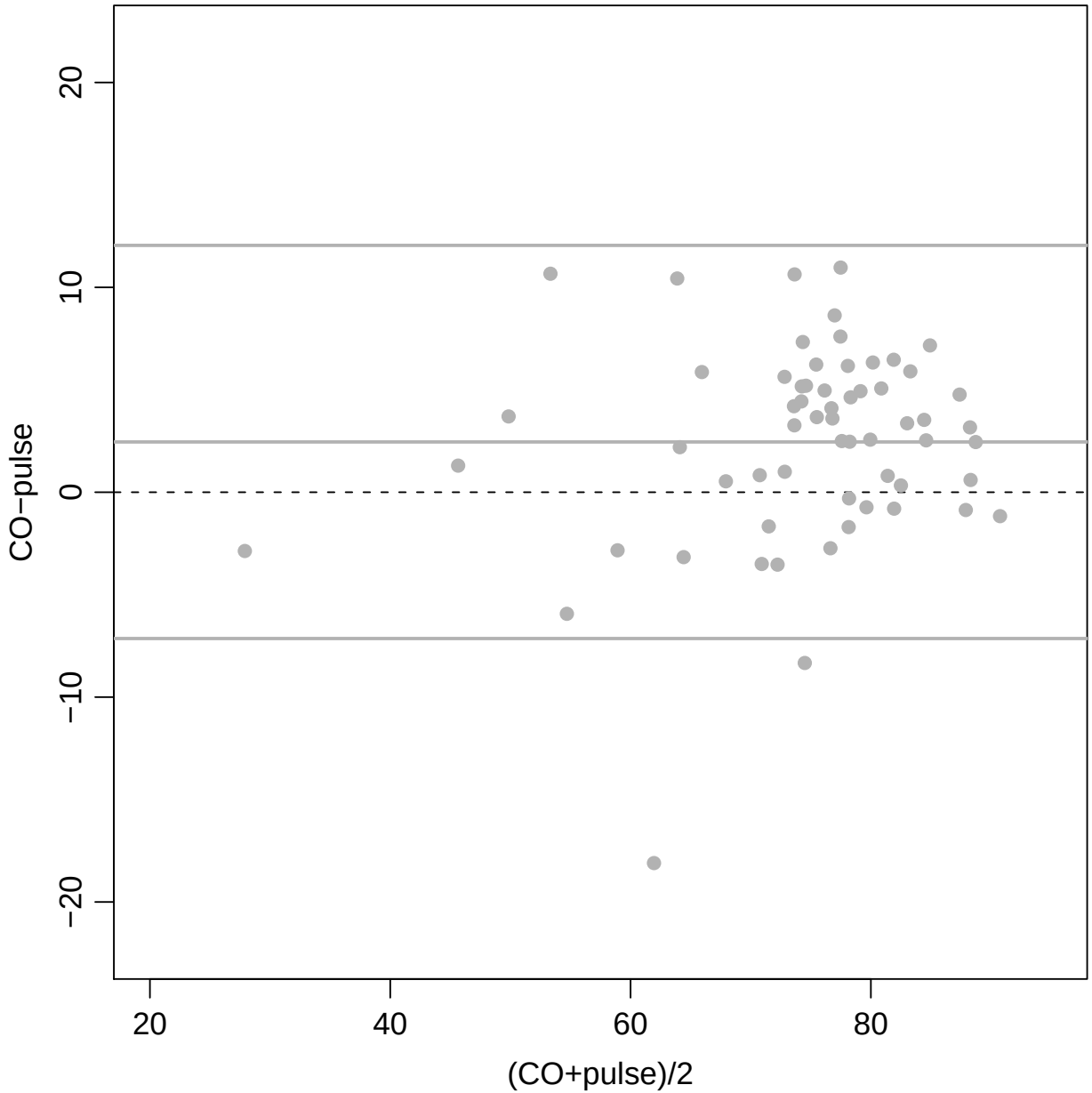
Oximetry data



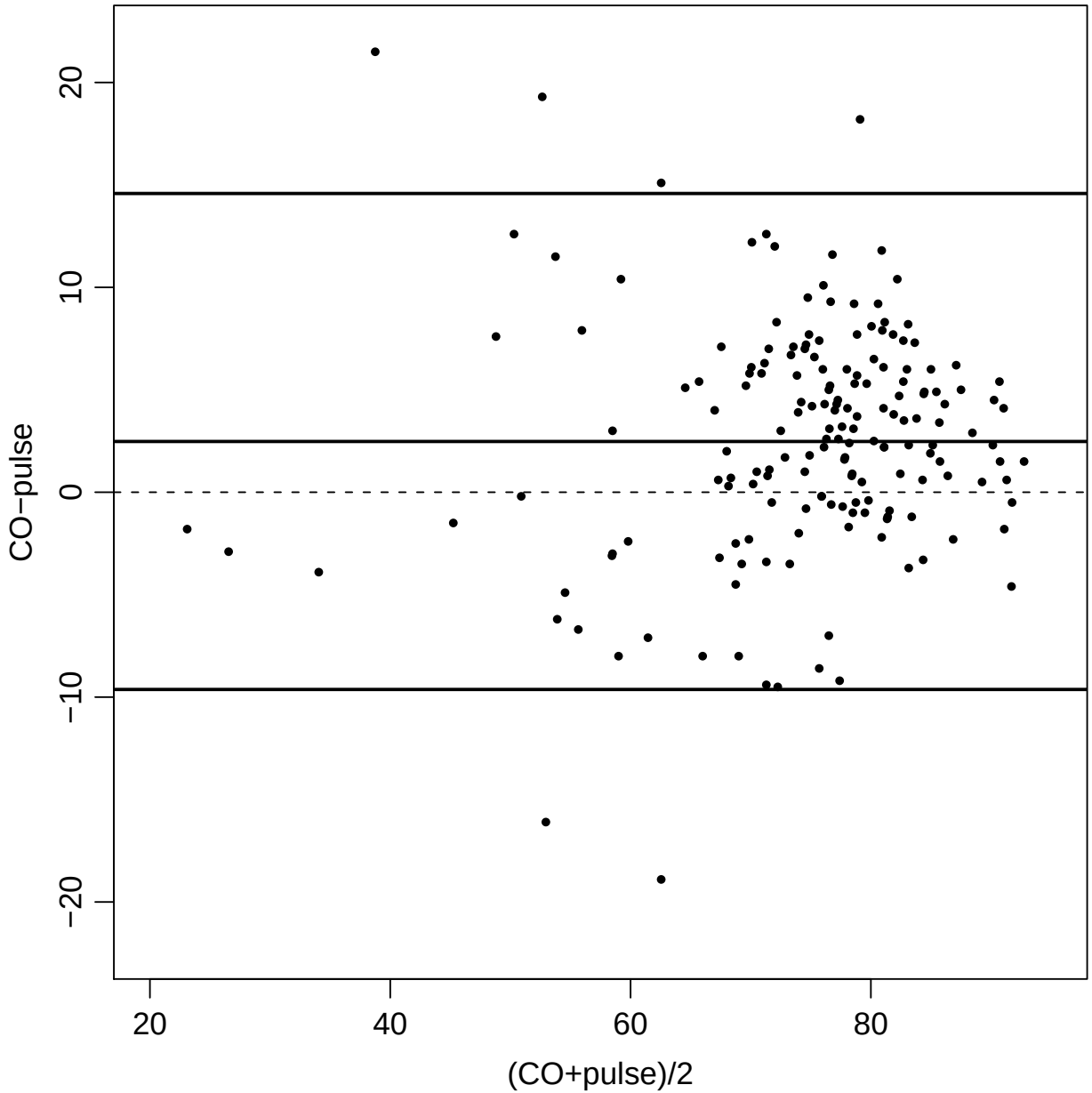
Oximetry data



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:

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

- $\mathrm{var}(y_{1jr} - y_{2jr}) = \tau_1^2 + \tau_2^2 + \sigma_1^2 + \sigma_2^2$
  — note that the term $a_{ir} - a_{ir}$ cancels because we are referring to the *same* replicate.

Wrong or almost right

In the model the correct limits of agreement would be:

$$\alpha_1 - \alpha_2 \pm 1.96\sqrt{\tau_1^2 + \tau_2^2 + \sigma_1^2 + \sigma_2^2}$$

If we are using means of replicates to form the differences we have:

$$\begin{aligned} \bar{d}_i = \bar{y}_{1i\cdot} - \bar{y}_{2i\cdot} &= \alpha_1 - \alpha_2 + \frac{\sum_r a_{ir}}{R_{1i}} - \frac{\sum_r a_{ir}}{R_{2i}} \\ &\quad + c_{1i} - c_{2i} + \frac{\sum_r e_{1ir}}{R_{1i}} - \frac{\sum_r e_{2ir}}{R_{2i}} \end{aligned}$$

The terms with a_{ir} are only relevant for linked replicates in which case $R_{1i} = R_{2i}$ and therefore the term vanishes. Thus:

$$\mathrm{var}(\bar{d}_i) = \tau_1^2 + \tau_2^2 + \sigma_1^2/R_{1i} + \sigma_2^2/R_{2i} < \tau_1^2 + \tau_2^2 + \sigma_1^2 + \sigma_2^2$$

so the limits of agreement calculated based on the means are much too narrow as prediction limits for differences between future *single* measurements.

(Linked) replicates as items

If replicates are taken as items, then the calculated differences are:

$$d_{ir} = y_{1ir} - y_{2ir} = \alpha_1 - \alpha_2 + c_{1i} - c_{2i} + e_{1ir} - e_{2ir}$$

which has variance $\tau_1^2 + \tau_2^2 + \sigma_1^2 + \sigma_2^2$, and so gives the correct limits of agreement. However, the differences are not independent:

$$\mathrm{cov}(d_{ir}, d_{is}) = \tau_1^2 + \tau_2^2$$

Negligible if the residual variances are very large compared to the interaction, variance likely to be only slightly downwards biased.

Exchangeable replicates as items?

If replicates are exchangeable it is not clear how to produce the differences using replicates as items.

If replicates are paired at random (se the function `perm.repl`), the variance will still be correct using the model without the $i \times r$ interaction term (a_{ir}):

$$\text{var}(y_{1ir} - y_{2is}) = \tau_1^2 + \sigma_1^2 + \tau_2^2 + \sigma_2^2$$

Differences will be positively correlated within item:

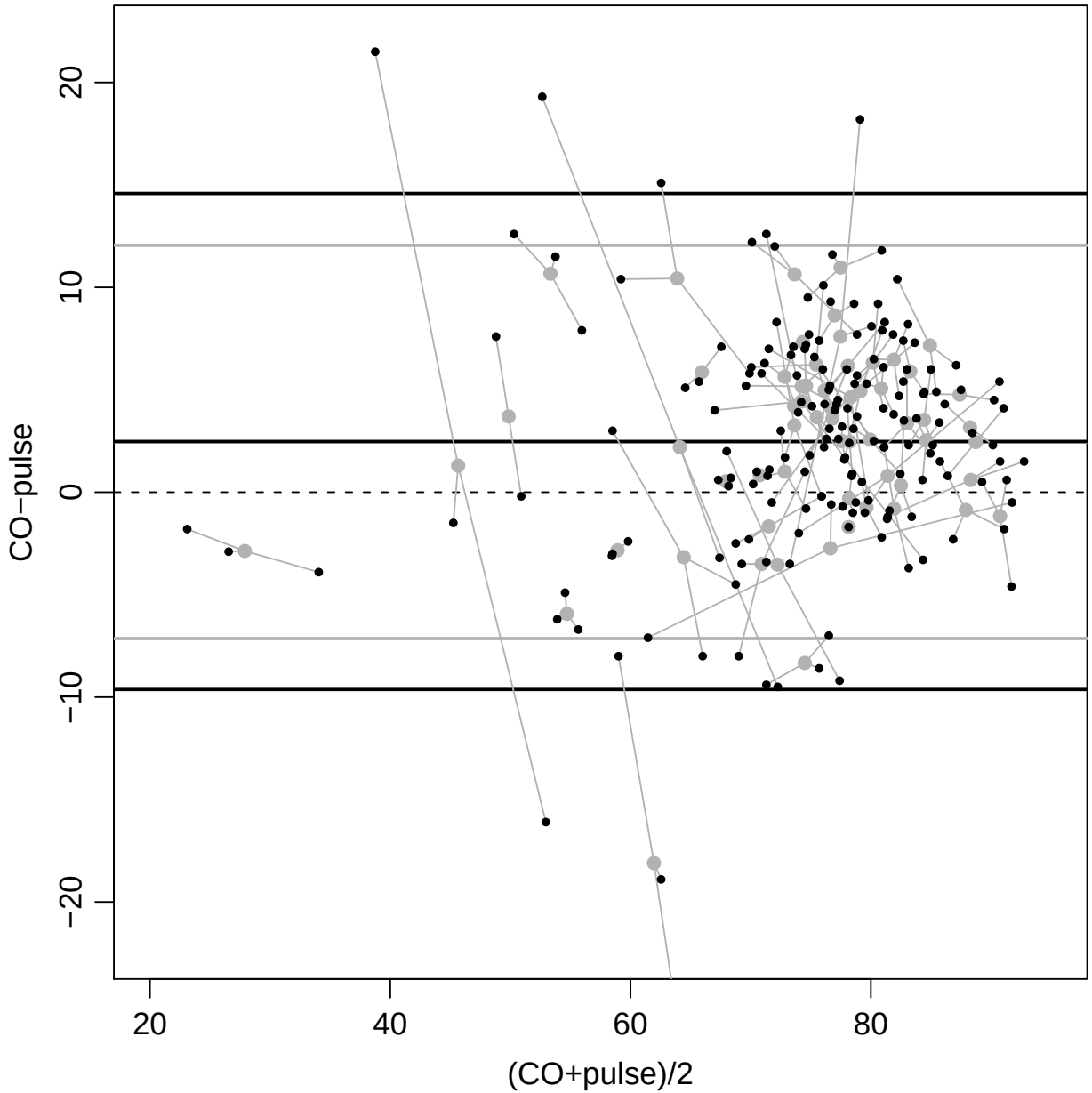
$$\text{cov}(y_{1ir} - y_{2is}, y_{1it} - y_{2iu}) = \tau_1^2 + \tau_2^2$$

— slight underestimate of the true variance.

Recommandattion

- Fit the correct model, and get the estimates fron that, e.g. by using `BA.est`.
- If you must:
  - Use linked replicates as item.
  - If replicates are not linked; make a random linking. Note: If this give a substantially different picture than using the original replicate numbering as linking key, there might be something fishy about the data.

Oximetry data



A general model

Thursday 29 November 2007, morning

Bendix Carstensen

Method Comparison Studies in Practise
28–30 November 2007
Dept. of Medical Epidemiology and Biostatistics,
Karolinska Institutet, Stockholm

Extension of the model:

$$
\begin{aligned}
y_{mir} &= \alpha_m + \mu_i + a_{ir} + c_{mi} + d_{mr} + e_{mir} \\
&\quad \text{s.d.}(a_{ir}) = \omega \quad \text{— between replicates} \\
&\quad \text{s.d.}(c_{mi}) = \tau_m \quad \text{— "matrix"-effect} \\
&\quad \text{s.d.}(d_{mr}) = \nu_m \quad \text{— } m \times r \\
&\quad \text{s.d.}(e_{mir}) = \sigma_m \quad \text{— measurement error}
\end{aligned}
$$

Method, Item, Replicate

- 1 3-way interaction
- 3 2-way interactions

What part of the interactions should be systematic (fixed) and what part should be random?

(m, r) - between replicates within method

This effect has $M \times R$ levels, usually a rather small number.

This effect will therefore normally be modelled as a fixed effect, but not necessarily with $M \times R$ parameters, presumably fewer.

If replicates are times of sampling or analysis, we may consider different time trends for each method, e.g.

$$d_{mr} = \gamma_m t_r$$

A random $m \times r$-effect would be hard to interpret.

(i, r) - between replicates within individual

Observations with same (i, r) — but different method — will be correlated.

Use if all methods are applied to each item at

- different times
- at different locations
- at different conditions

This means there is a minimal structure to replicates — they are linked.

There might be further structure, e.g. a systematic effect of a time.

(m, i) - between methods within individual

This is what is often called a "matrix" effect.

Matrix in the chemical sense: The surrounding matter ("matrix") in which the stuff of interest is dissolved.

Represents random effects of items reacting differently on each measurement method.

Logical to require that the variance of these methods was allowed to differ between methods.

Variance component model!

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

Note we do not consider the method by replicate interaction any more.

The model is a (standard) variance component model, where two of the variance components depend on method.

Fitting the variance component model

Complicated and counter-intuitive in **R**:

```
> library( nlme )
> lme( y ~ meth + item,
       random = list( item = pdIdent(~meth - 1),
                      repl = ~1),
       weights = varIdent(form = ~1 | meth),
       data = ox)
```

```
Random effects:
 Formula: ~meth - 1 | item
 Structure: Multiple of an Identity
          methCO methpulse
StdDev: 2.928042  2.928042

 Formula: ~1 | repl %in% item
        (Intercept) Residual
StdDev:    3.415692 2.224868

Variance function:
 Structure: Different standard deviations per stratum
 Formula: ~1 | meth
 Parameter estimates:
      CO    pulse
1.000000 1.795365
Number of Observations: 354
Number of Groups:
          item repl %in% item
            61             177
```

Tease out variances for later use?

Even worse.

Therefore it has been packaged in a function that calls `lme` and then tease out the relevant parameters.

```
> BA.est(ox)
$bias
        CO     pulse
 0.000000 -2.470446

$sd.s
   MxI.CO  MxI.pulse       IxR   resid.CO resid.pulse
 2.928042   2.928042  3.415692   2.224868    3.994451

Warning message:
In pt(q, df, lower.tail, log.p) : NaNs produced
```

Unequal bias

Thursday 29 November 2007, afternoon

Bendix Carstensen

Method Comparison Studies in Practise
28–30 November 2007
Dept. of Medical Epidemiology and Biostatistics,
Karolinska Institutet, Stockholm

Extension with non-constant bias

$$y_{mir} = \alpha_m + \beta_m \mu_i + \text{random effects}$$

There is now a *scaling* between the methods.

Methods do not measure on the same scale — the relative scaling is *estimated*, between method 1 and 2 the scale is β_2/β_1.

Consequence: Multiplication of all measurements on one method by a fixed number does not change results of analysis:

The corresponding β is multiplied by the same factor as is the variance components for this method.

Variance components

All two-way interactions:

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

The random effects c_{mi}, d_{mr} and e_{mir} have variances specific for each method.

But a_{ir} does not depend on m — must be scaled to each of the metods by the corresponding β.

Implies that $\omega = \mathrm{s.d.}(a_{ir})$ is irrelevant — the scale is arbitrary. The relevant quantities are $\beta_m\omega$ — the between replicate variation within item *as measured on the mth scale*.

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.

Conversion between methods

Friday 30 November 2007, morning

Bendix Carstensen

Method Comparison Studies in Practise
28–30 November 2007
Dept. of Medical Epidemiology and Biostatistics,
Karolinska Institutet, Stockholm

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:

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

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

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

The width of the prediction interval is:

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

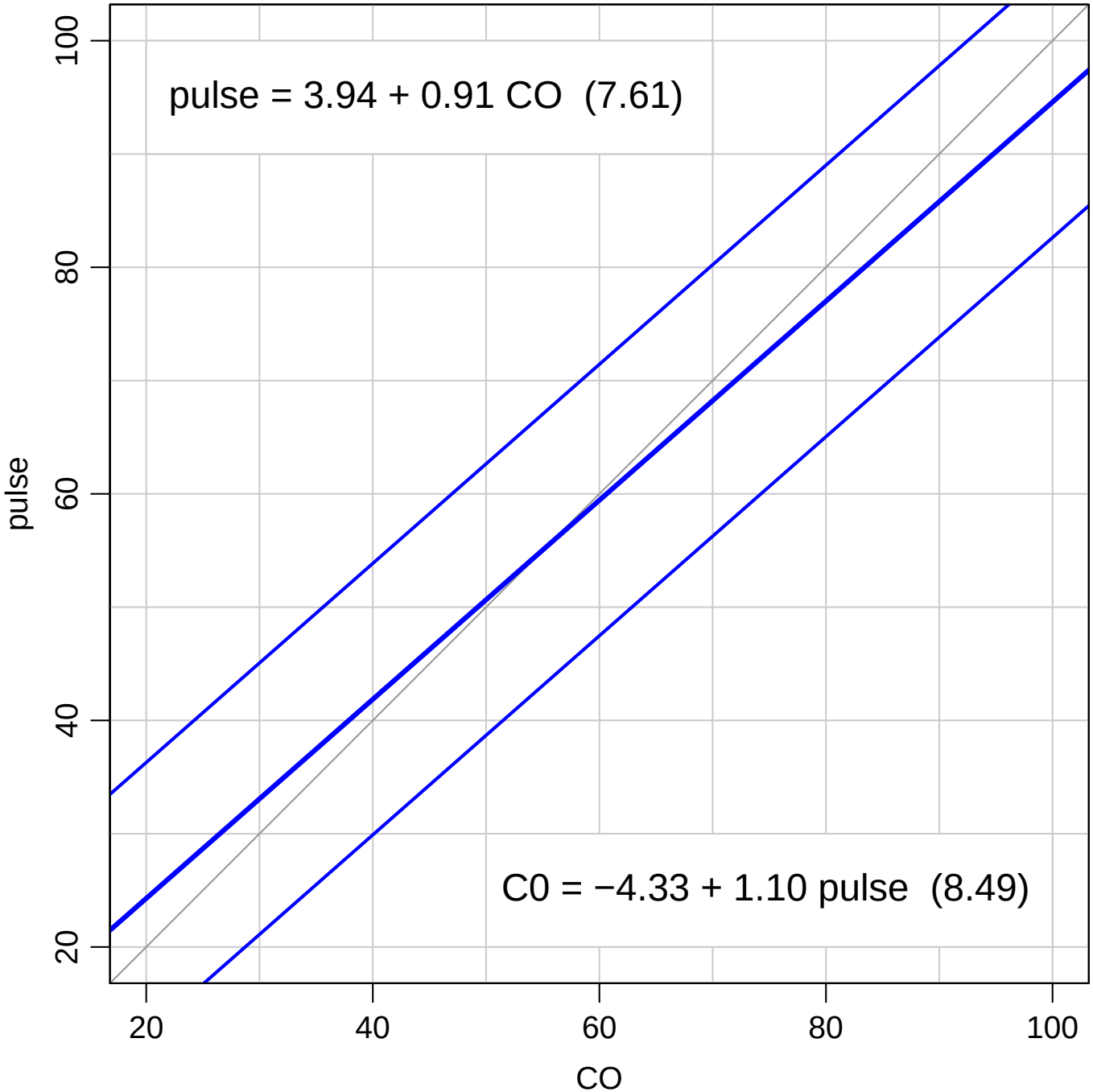
If we do the prediction the other way round $(y_1|y_2)$ we get the same relationship i.e. a line with the inverse slope, β_1/β_2.

The width of the prediction interval in this direction is:

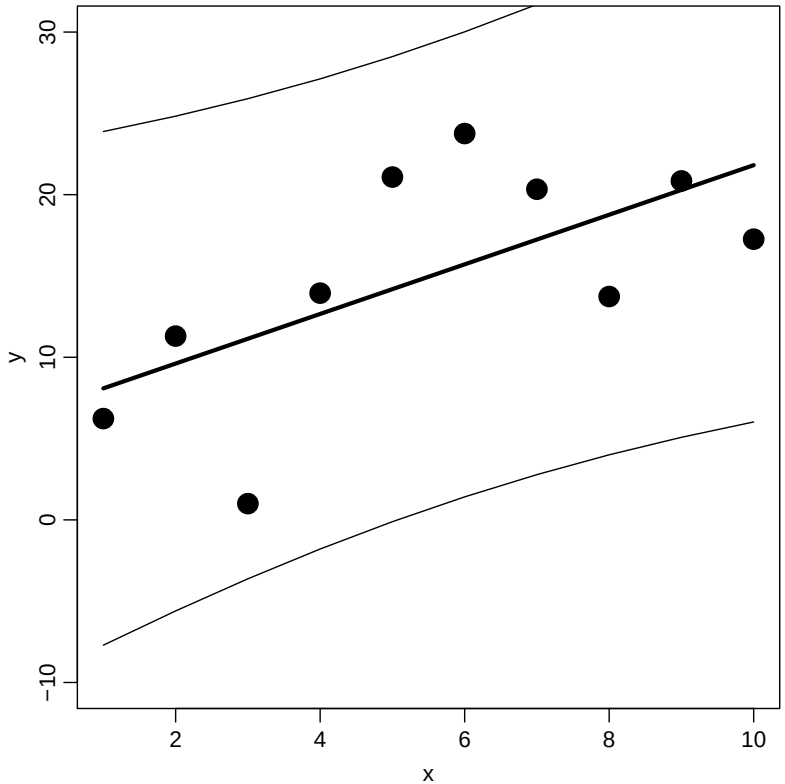
$$2 \times 1.96 \times \sqrt{(\tau_1^2 + \sigma_1^2) + \left(\frac{\beta_1}{\beta_2}\right)^2 (\tau_2^2 + \sigma_2^2)}$$

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

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


pulse = 3.94 + 0.91 CO (7.61)
C0 = −4.33 + 1.10 pulse (8.49)
pulse
CO
20
40
60
80
100

What happened to the curvature?



Usually the prediction limits are curved:

$$\hat{y}|x \pm 1.96 \times \hat{\sigma}\sqrt{1 + x'x}$$

In our prediction we have ignored the last term $(x'x)$, i.e. effectively assuming that there is no estimation error on $\alpha_{2.1}$ and $\beta_{2.1}$.

Variance components

Friday 30 November 2007, morning

Bendix Carstensen

Method Comparison Studies in Practise
28–30 November 2007
Dept. of Medical Epidemiology and Biostatistics,
Karolinska Institutet, Stockholm

Variance components

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

3 variance components / random effects:

- a_{ir}: between replicates within item, ω^2
  $\beta_m\omega$ is the relevant quantity.
- c_{mi}: matrix effect τ_m^2
  τ_m is the relevant quantity.
- e_{mir}: measurement error, residual variation σ_m^2
  σ_m is the relevant quantity.

Variance components

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

The total variance of a measurement is:

$$\sqrt{\beta_m^2\omega^2 + \tau_m^2 + \sigma_m^2}$$

These are the variance components reported by `print.MethComp` and shown by `plot.VarComp`.

Repeatabiliy and reproducibility

Repeatability is based on the difference between measurements made under comparable, though not exactly identical conditions.

Reproducibility is based on the difference between measurements made under comparable, though not exactly identical conditions.

This is a different setting from the one underlying the modelling of data from a comparison experiment.

The exchangeability has no meaning, we are discussing future measurements in different circumstances.

Repeatabiliy and reproducibility

Repeatability: $2.8\sigma_m$:
same individual, same replicate, but not considering the variation that constitute differences between replicates *in the experiment*.

Hence *reproducibility* is not estimable from a classical experiment, unless an extra layer of replication is introduced — i.e. different laboratories.