Adjust for baseline — or not in studies of change

Bendix Carstensen Steno Diabetes Center, Gentofte, Denmark
http://BendixCarstensen.com

EDEG, Kildare House, Ireland

May 2016

http://BendixCarstensen/SDC/LEAD

Basic set-up

Measurement at two time points

- Randomized study:
  - Effect of randomization
  - 1st point special (**pre**-intervention)
- Observational study
  - Differences between groups
  - — and changes in difference
  - 1st point **not** special

Observational scheme

- Observations / measurements for each individual, i:
  - Baseline: y_{0i}
  - Follow-up: y_{1i}
  - Treatment group
  - Covariates
- Topic of interest:
  - how much is the change from from baseline to follow-up
  - how much does this depend on treatment (and other covariates)
  - in **observational** studies covariate effects at baseline may be of interest

Observation characteristics

- Baseline is subject to random error
- If the random error at baseline is large **positive**:
  - baseline measurements are "artificially" **large**
  - $\Rightarrow$ **change** from baseline to follow-up is **smaller**
- If the random error at baseline is large **negative**:
  - baseline measuremnts are "artificially" **small**
  - $\Rightarrow$ **change** from baseline to follow-up is **larger**
- $\Rightarrow$ **change** depends on the baseline measurement
- . . . regression to the mean

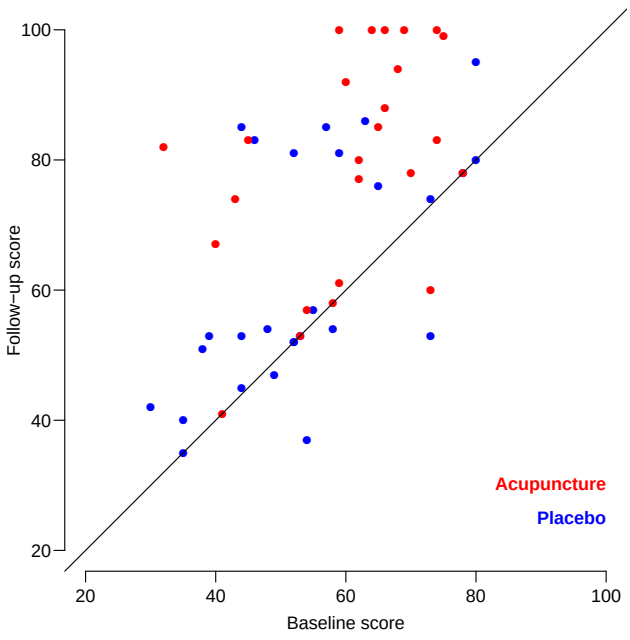
Example from Vickers et al.[?]

```
> library( Epi )
> library( foreign )
> acp <- read.dta( "../data/sportsmen.dta" )[,-4]
> names( acp ) <- c("bl","fu","gr")
> acp$gr <- factor( acp$gr, labels=c("Placebo","Acupuncture") )
> str( acp )
'data.frame': 54 obs. of  3 variables:
 $ bl: num  59 53 46 38 52 63 30 73 44 48 ...
 $ fu: num  81 53 83 51 81 86 42 74 45 54 ...
 $ gr: Factor w/ 2 levels "Placebo","Acupuncture": 1 1 1 1 1 1 1 1 1 1 ...
> head( acp )
  bl fu      gr
1 59 81 Placebo
2 53 53 Placebo
3 46 83 Placebo
4 38 51 Placebo
5 52 81 Placebo
6 63 86 Placebo
```


Randomized to acupuncture / placebo

Outcome: Shoulder pain rating (scale from 0 to 100)

Change is the vertical distance from the identity line to the point

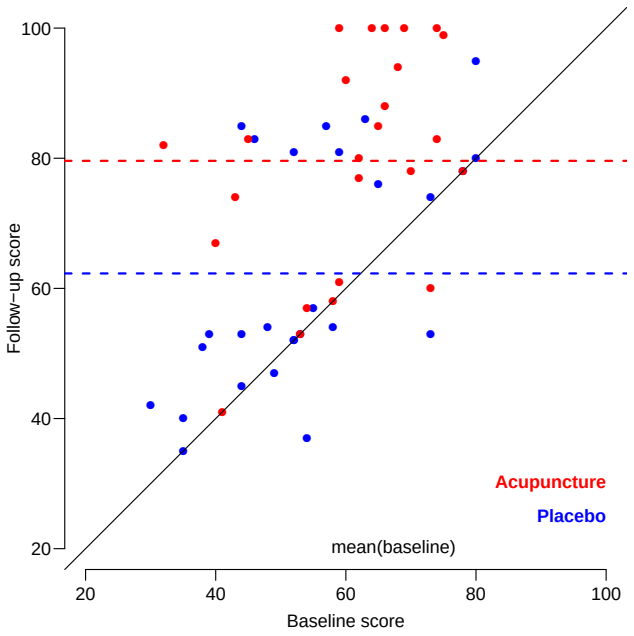


Follow-up analysis

$$y_{fi} \ \mu_g$$

Randomized study: Analysis of the follow-up measurements is in principle unbiased

because the baseline-distribution is the same in the two groups.



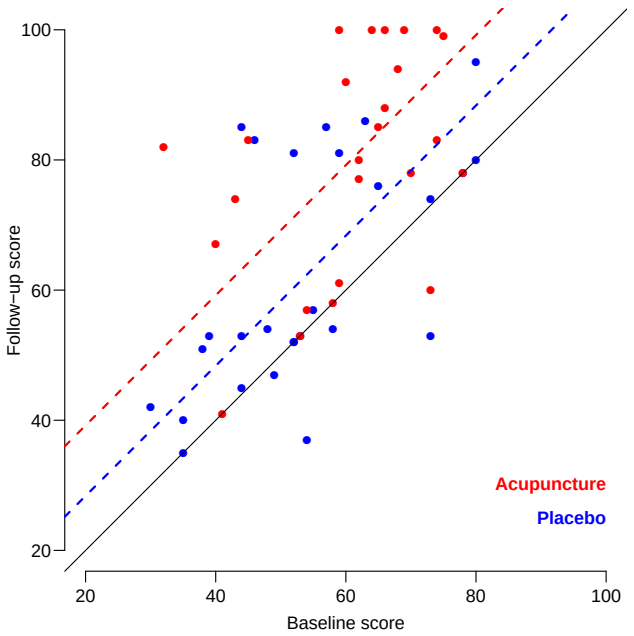
Analysis of follow-up values

```
> # Follow-up
> fm <- with( acp, tapply( fu, gr, mean ) )
> c( fm, diff(fm) )
    Placebo Acupuncture Acupuncture
    62.2963     79.6000     17.3037
> mf <- lm( fu ~ gr, data = acp )
> round( ci.lin( mf ), 4 )
              Estimate StdErr       z     P    2.5%   97.5%
(Intercept)    62.2963 3.3783 18.4400 0e+00 55.6749 68.9177
grAcupuncture  17.3037 4.8723  3.5515 4e-04  7.7542 26.8532
```


Analysis of change-scores

$$y_{fi} - y_{bi} \ \mu_g$$

The change score result (treatment effect) is the vertical difference between the lines.



Analysis of change-scores

```
> # Follow-up
> cm <- with( acp, tapply( fu-bl, gr, mean ) )
> c( cm, diff(cm) )
    Placebo Acupuncture Acupuncture
    8.37037    19.20000    10.82963
> mc <- lm( fu-bl ~ gr, data = acp )
> round( ci.lin( mc ), 4 )
              Estimate StdErr      z      P   2.5%   97.5%
(Intercept)     8.3704 2.9480 2.8394 0.0045 2.5924 14.1483
grAcupuncture  10.8296 4.2516 2.5472 0.0109 2.4966 19.1627
```

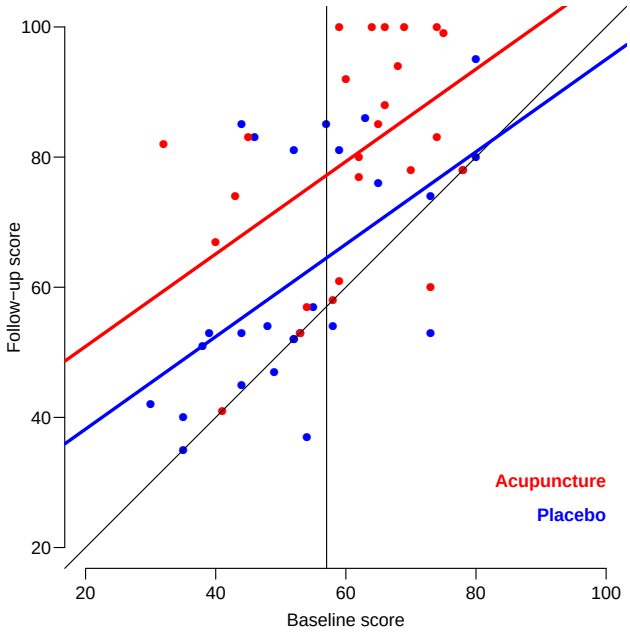

Conditioning on baseline

$$y_{1i} | y_{0i} \ \mu_g$$

Accounts for possible imbalances in baseline distribution

Controlling for confounding by baseline value

Effect is vertical distance between lines



Conditioning on baseline

```
> ml <- lm( fu ~ bl + gr, data = acp )
> round( ci.lin( ml ), 4 )
              Estimate StdErr      z      P   2.5%   97.5%
(Intercept)    23.9973 9.1092 2.6344 0.0084 6.1435 41.8511
bl              0.7102 0.1602 4.4323 0.0000 0.3962  1.0243
grAcupuncture  12.7057 4.2857 2.9647 0.0030 4.3059 21.1056
```

- $y_{1i} = M + By_{0i} + D_g$
- Treatment effect is 12.7 points:
  - change on placebo:
    $M + (B-1)y_{0i} + D_{\text{pl}} = 23.997 + 0.290y_{01} + 0$
  - change on treatment:
    $M + (B-1)y_{0i} + D_{\text{tr}} = 23.997 + 0.290y_{01} + 12.706$

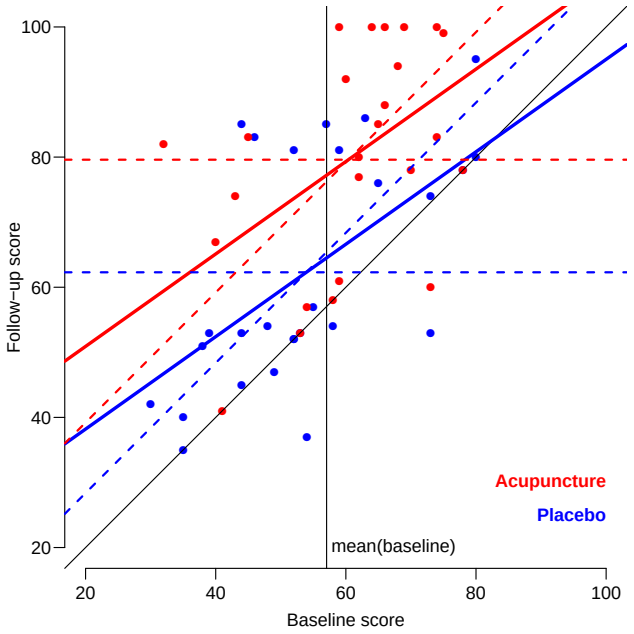
Conditioning on baseline

- $y_{1i} = M + By_{0i} + D_g$
- Treatment effect is 12.7 points:
  - change on placebo:
    $M + (B-1)y_{0i} + D_{\text{pl}} = 23.997 + 0.290y_{01} + 0$
  - change on treatment:
    $M + (B-1)y_{0i} + D_{\text{tr}} = 23.997 + 0.290y_{01} + 12.706$
- Change from baseline depends on baseline value
- **Difference** in change between does **not**
- . . . but that is a model **assumption**.

Comparing the three approaches

Effect is vertical distance between lines

Three sets of lines — three different estimates.



It all depends on the **baseline imbalance**

Random effects model

- We have repeated measures on each person
- . . . so why not use a random effects model?
- Greater flexibility:
  - accommodate more than two measurements
  - not necessarily the same no. measurements per person
  - accommodate **actual** measurement times
- For two points it is close to the ANCOVA approach, but not the same

Random effects model

Random effects model: $\Rightarrow$ data in the long format:

```
> lg <- reshape( acp, varying=1:2, v.names="score", direction="long" )
> head( lg )
         gr time score id
1.1 Placebo    1    59  1
2.1 Placebo    1    53  2
3.1 Placebo    1    46  3
4.1 Placebo    1    38  4
5.1 Placebo    1    52  5
6.1 Placebo    1    63  6
> str( lg )
'data.frame': 108 obs. of  4 variables:
 $ gr   : Factor w/ 2 levels "Placebo","Acupuncture": 1 1 1 1 1 1 1 1 1 ...
 $ time : int  1 1 1 1 1 1 1 1 1 ...
 $ score: num  59 53 46 38 52 63 30 73 44 48 ...
 $ id   : int  1 2 3 4 5 6 7 8 9 10 ...
 - attr(*, "reshapeLong")=List of 4
  ..$ varying:List of 1
  .. ..$ score: chr  "bl" "fu"
  .. ..- attr(*, "v.names")= chr "score"
      - attr(*, "times")= int  1 2
```


Random effects model

```
> library( lme4 )
> mr <- lmer( score ~  gr + gr:factor(time) + (1|id), data=lg )
> round( ci.lin( mr ), 2 )
                           Estimate StdErr     z    P  2.5% 97.5%
(Intercept)                   53.93   2.99 18.03 0.00 48.06 59.79
grAcupuncture                  6.47   4.31  1.50 0.13 -1.98 14.93
grPlacebo:factor(time)2        8.37   2.95  2.84 0.00  2.59 14.15
grAcupuncture:factor(time)2   19.20   3.06  6.27 0.00 13.20 25.20
```

- baseline mean in Placebo is 53.93
- baseline difference is 6.47
- change in the placebo group is 8.37
- change in the acupuncture group it is 19.20
- difference thus 10.83

But the difference is not the difference between the **conditional** means. . .

Random effects model

Formally the model is:

$$\begin{aligned} y_{it} =& \mu + \delta_g + \beta_t + \gamma_{gt} + \eta + a_i + e_{it} \\ & i = 1, \ldots, I, \quad t = 0, 1, \quad g = \text{pl, int} \\ & a_i \sim \mathcal{N}(0, \tau^2), \quad e_{it} \sim \mathcal{N}(0, \sigma^2) \end{aligned}$$

. . . this is a 2-dimensional normal distribution, and in this

$$y_1 | y_0 \sim \mathcal{N}\left(\mu_1 + \frac{\rho\sigma_1}{\sigma_0}(y_0 - \mu_0), \sigma_1^2(1-\rho^2) \right)$$

. . . don't worry

The random effects model shows how to compute the **conditional** distribution (well, mean) of:

follow-up measurements given baseline:

$$\mathrm{E}(y_1 | y_0) = \mu_1 + \frac{\rho\sigma_1}{\sigma_0}(y_0 - \mu_0),$$

μ_1, μ_0 are follow-up and baseline means,
σ_1, σ_0 are baseline variances, ρ the correlation
— all functions of the parameters specified in the random effects model.

Conditional mean

```
> mR <- lmer( score ~  gr*factor(time) + (1|id), data=lg )
> round( ci.lin( mR ), 2 )
                          Estimate StdErr     z    P  2.5% 97.5%
(Intercept)                  53.93   2.99 18.03 0.00 48.06 59.79
grAcupuncture                 6.47   4.31  1.50 0.13 -1.98 14.93
factor(time)2                 8.37   2.95  2.84 0.00  2.59 14.15
grAcupuncture:factor(time)2  10.83   4.25  2.55 0.01  2.50 19.16
> cf    <- fixef( mR )                  # regression coef
> tausq <- as.numeric( VarCorr( mR )$id ) # tau-squared
> sigsq <- attr( VarCorr( mR ), "sc" )^2  # sigma-squared
> rho   <- tausq/(tausq+sigsq)          # rho - correlation
```

Hence what we need to compute is:

```
> ( 1- rho ) * cf[2] + cf[4]
grAcupuncture
     13.97486
```

— close to the intervention effect 12.7 in the conditional model.

Treatment effects from different models

Treatment effect from model:	Estimate	s.e.	Cond.diff
Conditional (ANCOVA)	12.71	4.29	12.71
Random effects:			
identical baseline	13.94	3.72	13.94
different baseline	10.83	4.25	13.97
Change score difference	10.83	4.25	
Follow-up difference	17.30	4.87	

What goes on?

- Fitting a random effects model is just fitting a 2-dimensional normal distribution to (y_1, y_2)
- . . . subject to some mild restrictions
- In a 2-dimensional normal distribution, the conditional mean of y_2 given y_1 is just the regression of y_2 on y_1
- The random effects model puts a few restrictions on mean and variance of the 2-dimensional normal.
- . . . but the ANCOVA approach does **not**
- Treatment difference as evaluated by conditional means are almost the same.
- And will be in all sane examples.

Recommendations

- Always control for the obvious confounder: **baseline value**
- ANCOVA: Uses only a linearity assumption for the effects
- Random effects model (with or without baseline difference) also assume that variances are the same
- Including the baseline difference in the random effects model requires extra calculations.
- Omitting it does not, and gives the conditional difference as an explicit parameter.

Conclusions

- Not much difference between ANCOVA and random effects model
- But beware when using random effects models
  — must use conditional mean given baseline
  — use the model without baseline difference
  — otherwise you are effectively analyzing change-scores
- Use a random effects model:
  - actual dates of measurement
  - several measurements
  - **first** measurement has the status of baseline
- Since your data are not getting smaller and simpler:
- . . . you might as well get used to it sooner than later.