Repeated measures - LEAD

SDC
2014
http://bendixcarstensen.com/SDC/LEAD
1.0

Compiled Monday 18th April, 2016, 10:54
from: /home/bendix/teach/Lead/background.tex

Bendix Carstensen Steno Diabetes Center, Gentofte, Denmark
& Department of Biostatistics, University of Copenhagen
bxc@steno.dk
http://BendixCarstensen.com

Contents

Chapter 1

Repeated measures at two points

1.1 Full model

We consider outcome data for individuals i, measured at times 1 and 2, where time 1 is pre-randomization and 2 is post-treatment, so a full model would be:

$$y_{it} = \mu + \delta_g + \beta_t + \gamma_{gt} + \eta + a_i + e_{it}, \quad i = 1, \ldots, I, \quad t = 1, 2, \quad g = \text{pl, int} \tag{1.1}$$

$$\begin{aligned} a_i &\sim \mathcal{N}(0, \tau^2), \\ e_{it} &\sim \mathcal{N}(0, \sigma^2) \end{aligned} \tag{1.2}$$

where η represents the effect of possible confounders to be included in the model. For convenience we assume $\beta_1 = \gamma_{g1} = \delta_{\text{pl}} = 0$, where $g = 0$ is what would normally be taken as the placebo group.

Note that in the model (1.2), we allow different baseline means between randomization groups $g = g(i)$, in the parameter δ_g.

Thus the model with δ_g has 4 parameters to describe the baselines and follow-up measurements in the two groups, so the estimated means under this model is identical to the empirical means (at least if there are no other covariates in the model). In the parametrization chosen, the mean difference at follow-up is:

$$(\mu + \delta_{\text{int}} + \beta_2 + \gamma_{\text{int},2}) - (\mu + \delta_{\text{pl}} + \beta_2 + \gamma_{\text{pl},2}) = \delta_{\text{int}} + \gamma_{\text{int},2} - \gamma_{\text{pl},2}$$

and the difference in mean change scores is:

$$\begin{gathered} (\mu + \delta_{\text{int}} + \beta_2 + \gamma_{\text{int},2}) - (\mu + \delta_{\text{int}} + \beta_1 + \gamma_{\text{int},1}) - \\ (\mu + \delta_{\text{pl}} + \beta_2 + \gamma_{\text{pl},2}) + (\mu + \delta_{\text{pl}} + \beta_1 + \gamma_{\text{pl},1}) = \\ \gamma_{\text{int},2} - \gamma_{\text{int},1} - \gamma_{\text{pl},2} + \gamma_{\text{pl},1} = \\ \gamma_{\text{int},2} - \gamma_{\text{pl},2} \end{gathered}$$

In a randomized study one would expect that $\delta_g = 0$, hence a model without δ_g should be considered, in which case the two would be the same.

In the latter model the difference of the fitted mean at follow-up is identical to the difference between the mean expected change scores, that is a model that considers follow-up and change-scores on equal footing.

A corollary of this is that in the evauation of the difference between the randomization groups, the baseline mean is ancillary, and by first principles inference should be made in the conditional distribution given the sufficient statistic for the (common) baseline mean. Conditioning on the *single* individual means is a further narrowing of the sample space.

Also note that the formal interpretation of the model is that confounder effects are the *same* at baseline ($t = 1$) and follow-up ($t = 2$), which implies that change from baseline to follow-up is the same regardless of covariate values. Another way to state this is that there is no confounder by time interaction.

1.2 Conditioning on baseline

The usual approach to the analysis of repeated measures with a baseline and one follow-up measurement is to use the follow-up measurement as response in analysis with the baseline measurement as covariate [4, 1, 2, 3].

This can be viewed as using the basic statistical principle that inference should be made in the conditional distribution given the sufficient statistics for the ancillary parameters, which in this case is the individual-specific values of the random effects (a_i). The baseline measurement y_{i1} is not formally the sufficient statistic for this in model 1.2, but it is close and easier to handle.

The formal analysis of this is as follows: The random effects model 1.2 induces a 2-dimensional normal distribution of the measurements y_1 and y_2; in general terms:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right]$$

From standard statistical theory we know that under this model, the conditional distribution of y_2 given y_1 is:

$$y_2|y_1 \sim \mathcal{N}\left(\mu_2 + \frac{\rho\sigma_2}{\sigma_1}(y_1 - \mu_1), \sigma_2^2(1-\rho^2)\right)$$

Now in the model (1.2) we have the following values for the parameters μ_1, μ_2, σ_1^2, σ_2^2 and ρ in the 2-dimensional normal model outlined above:

$$\begin{aligned} \mu_1 &= \mu + \delta_g + \eta \\ \mu_2 &= \mu + \delta_g + \eta + \beta_2 + \gamma_{g2} \\ \sigma_1^2 = \sigma_2^2 &= \tau^2 + \sigma^2 \\ \rho &= \frac{\tau^2}{\sigma^2 + \tau^2} \end{aligned}$$

As opposed to the ANCOVA-approach that can be formally derived from a completely unspecified 2-dimensional normal distribution, the 2-dimensional normal distribution induced by the random effects model has the *same* variance at baseline and follow-up, namely $\sigma^2 + \tau^2$. But there are no restrictions on the covariance and it is a model with seperate means for the two groups, namely $(\mu + \eta, \mu + \eta + \beta_2)$ in the placebo group, and $(\mu + \eta + \delta_{\text{int}}, \mu + \eta + \beta_2 + \delta_{\text{int}} + \gamma_{\text{int},2})$ in the intervention group.

Thus the only difference between the ANCOVA approach and the random effects model is that the random effects model assumes that the marginal variances are the same at baseline and follow-up.

Using the relationships above, the formulae for the conditional distribution gives the conditional distribution of y_2 given y_1 in terms of the model parameters from (1.2) (well, we maintain ρ):

$$\begin{aligned} y_2|y_1 &\sim \mathcal{N}\left(\mu + \delta_g + \beta_2 + \gamma_{g2} + \eta + \rho\big(y_1 - (\mu + \delta_g + \eta)\big), (\sigma^2 + \tau^2)(1 - \rho^2)\right) \\ &= \mathcal{N}\left(\big((1-\rho)\mu + \beta_2\big) + \big((1-\rho)\delta_g + \gamma_{g2}\big) + \rho y_1 + (1-\rho)\eta, (\sigma^2 + \tau^2)(1 - \rho^2)\right) \end{aligned}$$

So if data were generated by model (1.2), and we fitted the regression of y_2 on y_1, we get the results in the form of a regression:

$$y_{i2} = M + By_{i1} + D_g + Z + e_i, \quad e_i \sim \mathcal{N}(0, \omega^2) \tag{1.3}$$

where Z is the effect of possible confounders.

We would then expect to see the following relationships between the parameters from the regression and the parameters from the model generating data:

- the term $(1-\rho)\mu + \beta_2$ should show up as the intercept M,
- the term $(1-\rho)\delta_g + \gamma_{g2}$ as the coefficient to the treatment indicator D_g,
- ρ as the coefficient to the baseline measurement y_1, B
- the coefficients to the confounders (in Z) should appear as the coefficients in η scaled by $1 - \rho$, and
- the residual standard deviation, ω should be $\sqrt{(\sigma^2 + \tau^2)(1 - \rho^2)}$.

In any practical circumstances, when fitting the two different models (the random effects model and the conditional model) we should find these relationships quite accurately if the random effects model fit well, because the relationships are derived under the assumption that the random effects model is the correct model.

Under this assumption it seems that when conditioning on the first measurement y_{i1}, we are implicitly assuming that $\delta_g = 0$ if we interpret the coefficient to the treatment indicator as the treatment effect.

However, when fitting a regression of follow-up on baseline, we are formally not making any assumptions about the *marginal* distribution of y_1, the baseline, only that the *conditional* of y_2 given y_1 is normal (and has a structure as it would have been if the marginal of y_1 were normal).

If we want to allow for baseline imbalance in the random effects model, we must fit the random effects model with δ_g. In this model, the mean change is γ_{g2} which is estimated as the mean of the changes — but the mean change *given* baseline equal to y_1 is $(1-\rho)\delta_g + \gamma_{g2} = \big(1 - \tau^2/(\tau^2 + \sigma^2)\big)\delta_g + \gamma_{g2}$. Thus if we want the conditional mean as calculated under the random effects model we must compute it from the parameters as above. But there is no easy way to get a standard error for this quantity.

1.3 Reporting effects

Very often researchers in addition to the treatment effect also want to report the change in each randomization group separately, and that is usually done by just computing the mean change in each group with the corresponding empirical standard deviation.

But when using the conditional model, it is not sensible to ask for the mean change in each group; conditioning on baseline is actually expressing a firm disinterest in this.

Insisting on mean changes in each group requires a re-introduction of the *marginal* distribution of y_1, and hence a reconstruction the entire *joint* distribution of (y_1, y_2). Which of course need not be the bivariate normal as induced by the random-effects model.

However, since the expected (mean) change in each group depends on the mean baseline in the group, it would be sensible to compute the mean change in both groups at some fixed baseline value. For example the overall mean, but even if these changes have a difference equal to the estimated treatment effect, we could compute a number of other sets fo changes with the same property, by just conditioning on some other baseline value.

1.3.1 Conditional model

Usually the treatment effect is reported as the coefficient D_g to the treatment indicator from an analysis with y_1 as covariate.

In the conditional model, that is modelling the follow-up with the baseline measurement as covariate as in model 1.3, we have:

$$y_{i2} - y_{i1} = M + (B-1)y_{i1} + D_g + Z + e_i$$

so the expected difference from baseline to follow-up depends both on the covariates in the model and on the baseline value.

Still, the dependence is the *same* in the two treatment groups, and hence the difference between these differences (D_g) can be taken as the treatment effect — how much larger is the change in the treatment group than in the placebo group at `any` baseline value.

1.3.2 Full random effects model

It could be argued that if we want to report within-group changes, it would be more reasonable to fit the random effects model (with or without δ_g), and report the quantities $\beta_2 + \gamma_{g2}$. Because of the obvious redundancy, these group-specific changes have a difference corresponding to the claimed treatment effect in the full model 1.2, which would seem an obvious advantage[1].

Thus the random effects model gives the possibility to model baseline imbalance and sensibly report changes observed within groups; the model conditioning on the baseline does not — if we want to report the *expected* change in each of the treatment groups, it must of course be an expectation with respect to some assumed distribution of the variables upon which the change depends, in this case baseline values and confounders.

The random effects model makes assumptions about the distribution of the baseline-values, and hence we can derive the expected change to follow-up, essentially using the assumed normal distribution of the baseline measurements.

[1] Unless of course you subscribe to the notion that the amount of information obtained is proportional to the number of different models fitted to a given dataset.

Chapter 2

Two examples

This chapter illustrates the above piece of theory by two examples; the classical acupuncture example used in BMJ by Vickers & Altman [4], and the other a simulation example.

2.1 Acupuncture example

Here we read the data from acupuncture example in the BMJ article by Vickers and Altman [4] — data has kindly been put at my disposal by DGA.

```
> 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
```

2.1.1 Naive analyses

The simplest analyses would be to compute either the difference in follow-up score or the difference in change-scores:

```
> round( ci.lin( lm( fu    ~ gr, data=acp ) ), 2 )
             Estimate StdErr     z P  2.5% 97.5%
(Intercept)      62.3   3.38 18.44 0 55.67 68.92
grAcupuncture    17.3   4.87  3.55 0  7.75 26.85
> round( ci.lin( lm( fu-bl ~ gr, data=acp ) ), 2 )
             Estimate StdErr    z    P 2.5% 97.5%
(Intercept)      8.37   2.95 2.84 0.00 2.59 14.15
grAcupuncture   10.83   4.25 2.55 0.01 2.50 19.16
```

2.1.2 Conditional model

The model is fitted very simply:

```
> m0 <- lm( fu ~ bl + gr, data=acp )
> summary( m0 )
Call:
lm(formula = fu ~ bl + gr, data = acp)

Residuals:
    Min      1Q  Median      3Q     Max
-28.549  -9.258  -1.104  13.059  29.753

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   23.9973     9.1092   2.634  0.01125
bl             0.7102     0.1602   4.432 5.25e-05
grAcupuncture 12.7057     4.2857   2.965  0.00467

Residual standard error: 14.98 on 49 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:   0.43,        Adjusted R-squared:  0.4067
F-statistic: 18.48 on 2 and 49 DF,  p-value: 1.046e-06
```

From the model conditioning on baseline we see that the treatment effect is 12.7 points, that is, for *any* given baseline value the (mean) follow-up score will be 12.7 larger in the intervention group.

From the formulae above we have (since we assume no confounders present) that the change in the placebo group for a person with baseline y_1 is:
$M + (B-1)y_1 = 24.00 - 0.29 \times y_1$ and in the acupuncture group
$M + (B-1)y_1 + D_g = 24.00 - 0.29 \times y_1 + 12.71$. In order to report any of these two sensibly, we need some value for y_i; for example we could stick in the mean of the baseline measurements:

```
> ( mb <- mean( acp$bl ) )
[1] 57.04259
```

so we get:

```
> ( cf <- coef( m0 ) )
  (Intercept)            bl grAcupuncture
   23.9973054    0.7102148    12.7057205
> (cf-c(0,1,1)) %*% cbind( c(1,mb,0),  c(1,mb,1) )
         [,1]     [,2]
[1,] 7.467206 19.17293
```

The particular choice of y_1 is essentially arbitrary, if we use the joint mean we report a change which is certainly not generalizable to any other context, so it could be argued that an arbitrary reference value should be chosen.

2.1.3 Graphical illustration

We can illustrate this in figure **??**, where the thin vertical line is drawn at the mean baseline (for *all* persons), and the *mean* (expected) change for a person with baseline equal to the overall baseline mean is the distance from the intersect with the identity to either the red or blue line depending on the treatment group.

We could also consider the expected changes for persons with baseline score of say 40 and 80:

```
> (cf-c(0,1,1)) %*% cbind( c(1,40,0),  c(1,40,1) )
        [,1]     [,2]
[1,] 12.4059 24.11162
> (cf-c(0,1,1)) %*% cbind( c(1,80,0),  c(1,80,1) )
          [,1]     [,2]
[1,] 0.8144897 12.52021
```

as we see dramatically different changes, but with *differences* between changes equal to 12.71 in both cases.

When making the graph we need the regression coefficients and the mean baseline and the mean follow-up:

```
> cf <- coef( m0 )
> df <- with( acp, tapply( fu-bl, gr, mean ) )
> fu <- with( acp, tapply( fu   , gr, mean ) )


> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, bty="n", las=1 )
> with( acp, plot( bl, fu, pch=16, col=c("blue","red")[gr],
+                  xlim=c(20,100), ylim=c(20,100),
+                  xlab="Baseline score", ylab="Follow-up score" ) )
> abline( 0, 1 )
> abline( v=mb )
> text( mb, 20, " mean(baseline)", adj=0 )
> #abline( v=c(40,80), lty="26" )
> abline( cf[1]      , cf[2], lwd=3, col="blue" )
> abline( cf[1]+cf[3], cf[2], lwd=3, col="red" )
> abline( h=fu[1], lwd=2, lty=2, col="blue" )
> abline( h=fu[2], lwd=2, lty=2, col="red" )
> abline( df[1], 1, lwd=2, lty=2, col="blue" )
> abline( df[2], 1, lwd=2, lty=2, col="red" )
> text( rep(100,2), c(25,30), levels(acp$g), font=2, col=c("blue","red"), adj=1 )
```

2.1.4 Random effects model

In order to fit the random effects model we must have the 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 1 ...
 $ time : int  1 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
  ..$ v.names: chr "score"
  ..$ idvar  : chr "id"
  ..$ timevar: chr "time"
```

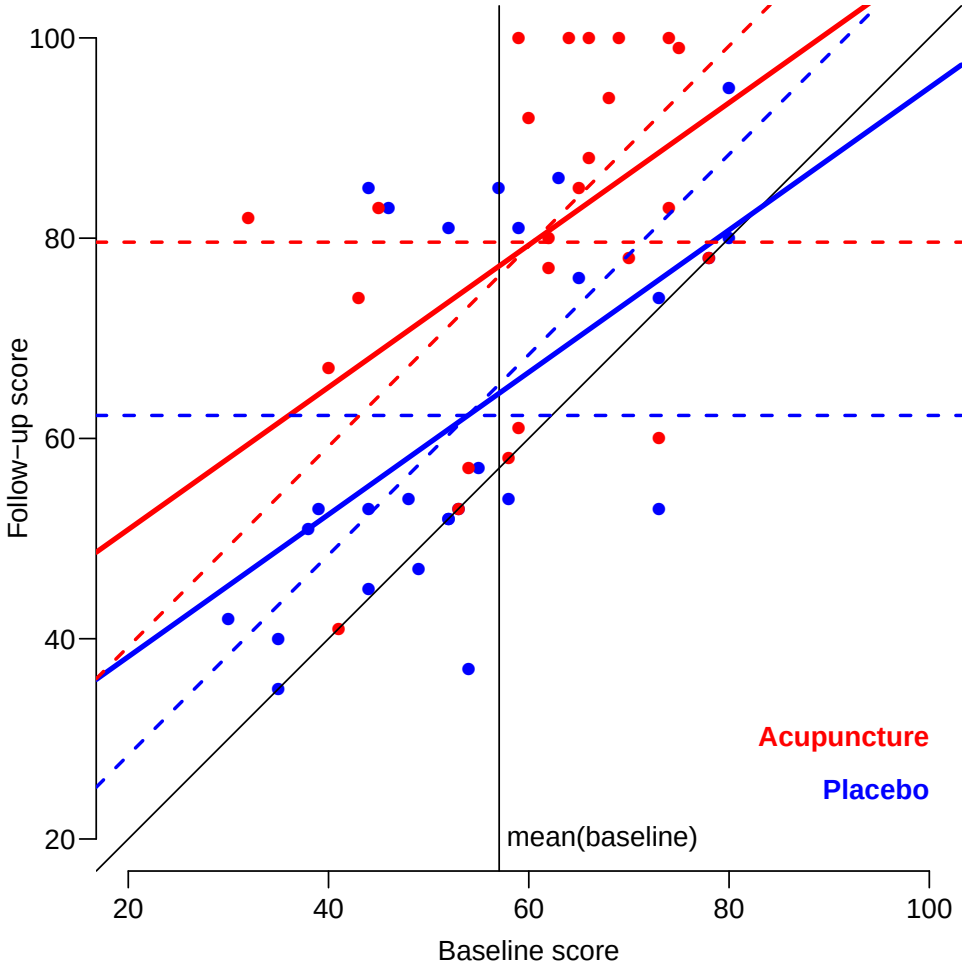

Figure 2.1: *Follow-up versus baseline score for acupuncture data. Regression lines are from the ANCOVA model, the horizontal dashed lines are the means of the follow-up data, and the* 45° *dashed lines correspond to the analysis of the change scores — the change for each person is the* vertical *distance to the identity line.*
Note that the three lines of each color necessarily all pass through the point (mean(`bl`), mean(`fu`)).

2.1.4.1 Unequal means at baseline

Fitting a model for the long-form data is easily done using `lmer` — we first fit the model with baseline imbalance:

```
> 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
```

Thus the acupuncture group has a mean at baseline which is 6.47 larger then the placebo group; the change in the placebo group is 8.37, in the acupuncture group it is 19.20, the difference thus 10.83, not far from the difference we saw in the conditional model.

If we were to compare to the parameter estimated in the conditional model it should be $(1-\rho)\delta_g + \gamma_{g2}$. This formula refers to a slightly different parametrization:

```
> 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
```

The latter parametrization is the classical interaction parametrization; as above the change in the placebo (reference) group is 8.37, but we how have the interaction parameter as the *difference* in changes between groups, $10.83 = 19.20 - 8.37$, so we have $\gamma_{g2} = 10.83$ and $\delta_g = 6.47$. The ρ is derived from the variance components (labeled `Random effects`) in the model as $\rho = \tau^2/(\tau^2 + \sigma^2)$, τ^2 is the between-person variance, labeled `(Intercept)` and σ^2 is the residual variance, labeled `Residual`.

```
> summary( mR )
Linear mixed model fit by REML ['lmerMod']
Formula: score ~ gr * factor(time) + (1 | id)
   Data: lg

REML criterion at convergence: 830.1

Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.80685 -0.56741  0.01961  0.58225  1.69548

Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept) 124.2    11.14
 Residual             117.3    10.83
Number of obs: 104, groups:  id, 52

Fixed effects:
                           Estimate Std. Error t value
(Intercept)                  53.926      2.991  18.031
grAcupuncture                 6.474      4.313   1.501
factor(time)2                 8.370      2.948   2.839
grAcupuncture:factor(time)2  10.830      4.252   2.547

Correlation of Fixed Effects:
            (Intr) grAcpn fct()2
grAcupunctr -0.693
factor(tm)2 -0.493  0.342
grAcpnc:()2  0.342 -0.493 -0.693
```

But we can actually fish them out of the `mR` object, a `lmerMod` object. But it is very clumsy.

```
> VarCorr( mR )
 Groups   Name        Std.Dev.
 id       (Intercept) 11.143
 Residual             10.832
> VarCorr( mR)
 Groups   Name        Std.Dev.
 id       (Intercept) 11.143
 Residual             10.832
> ( tausq <- as.numeric( VarCorr( mR )$id ) )
[1] 124.1719
> ( sigsq <- attr( VarCorr( mR ), "sc" )^2 )
```

```
[1] 117.323
> ( rho <- tausq/(tausq+sigsq) )
[1] 0.5141802
```

Hence what we need to compute is:

```
> 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
> round( cf <- fixef( mR ), 2 )
                (Intercept)                grAcupuncture                factor(time)2
                      53.93                         6.47                         8.37
grAcupuncture:factor(time)2
                      10.83
> ( 1- rho ) * cf[2] + cf[4]
grAcupuncture
     13.97486
```

— also a little bit from the 12.7 in the conditional model.

2.1.4.2 Identical means at baseline

Now if we fit a random effects model where we assume equal levels at baseline, that is the model with $\delta_g = 0$, we have a non-standard model. This is a model with no main effect of g, but with a $g \times t$ interaction. In order to get this we must hand-code the interaction; here the `Relevel` function (from the `Epi` package is used to merge the two baseline levels of the $g \times t$ interaction):

```
> lg <- transform( lg, G2 = Relevel( interaction( gr, time ),
+                                    list( B=1:2 ) ) )
> with( lg, ftable( gr, time, G2 ) )
                 G2  B Placebo.2 Acupuncture.2
gr          time
Placebo     1       27         0             0
            2        0        27             0
Acupuncture 1       25         0             0
            2        0         0            25
> ms <- lmer( score ~ G2 + (1|id), data=lg )
> round( ci.lin( ms ), 3 )
                Estimate StdErr      z     P   2.5%  97.5%
(Intercept)       57.038  2.172 26.262 0.000 52.782 61.295
G2Placebo.2        6.873  2.780  2.472 0.013  1.423 12.322
G2Acupuncture.2   20.817  2.874  7.242 0.000 15.184 26.451
```

We would want not only the change in each group, but also the difference between them — the intervention effect, so we use `ci.lin` for this:

```
> CM <- rbind( diag(3), c(0,-1,1) )
> rownames( CM ) <- c( names( fixef(ms) ), "Acp-eff" )
> round( ci.lin( ms, ctr.mat=CM ), 2 )
                Estimate StdErr     z    P  2.5% 97.5%
(Intercept)        57.04   2.17 26.26 0.00 52.78 61.30
G2Placebo.2         6.87   2.78  2.47 0.01  1.42 12.32
G2Acupuncture.2    20.82   2.87  7.24 0.00 15.18 26.45
Acp-eff            13.94   3.72  3.75 0.00  6.66 21.23
```

— and we see that in this dataset it makes very little difference whether we fit the baseline difference or not. But this is no surprise since this is randomized study and the baseline means are therefore *expected* be identical.

Treatment effect from model:	Estimate	s.e.	FU|BL
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	

From the table we see that allowing for different baseline gives the same s.e. as the conditional model but an estimate that deviates about 0.5 s.e., whereas the random effects model with identical baseline between the groups has a slightly smaller s.e. and an estimate that deviates about a third s.e., as well as a s.e. which is 15% smaller,

A fair summary would be that the three approaches in this case produces pretty much the same results.

We also see that the random effects model allowing for different baseline between groups produces an estimate which is identical to the analysis of the change-scores data. This is because both models essentially are saturated interaction models, and they also produce the same standard error of the effect.

The standard error of the intervention effect from the random effects model is substantially smaller that the other ones. This is because the s.e. is based on the residual sd. and the between-person variation is separated out.

2.2 Where do the differences come from?

Note that we had different baseline means in the two groups:

```
> gmn <- with( acp, tapply(bl,gr,mean) )
> c( gmn, df <- diff(gmn) )
    Placebo Acupuncture Acupuncture
  53.925926   60.400000    6.474074
```

Now suppose for the sage of the argument that the means had been exacly identical. We can fix this by adding the difference between means to the basline in the placebo group, and leaving everything else:

```
> acpx <- transform( acp, bl = bl + df*(gr=="Placebo") )
> gmn <- with( acpx, tapply(bl,gr,mean) )
> round( c( gmn, diff(gmn) ), 5 )
    Placebo Acupuncture Acupuncture
       60.4        60.4         0.0
```

What do we then get from the various approaches?

```
> ci.lin( lm( fu    ~ gr     , data=acpx ), subset="gr" )
             Estimate   StdErr        z            P   2.5%    97.5%
grAcupuncture  17.3037 4.872285 3.551455 0.0003831068 7.7542 26.85321
> ci.lin( lm( fu-bl ~ gr     , data=acpx ), subset="gr" )
```

```
              Estimate   StdErr        z            P     2.5%    97.5%
grAcupuncture  17.3037 4.251638 4.069891 4.703514e-05 8.970646 25.63676
> ci.lin( lm( fu    ~ bl + gr, data=acpx ), subset="gr" )
              Estimate   StdErr        z            P     2.5%    97.5%
grAcupuncture  17.3037 4.158268 4.161277 3.164731e-05 9.153649 25.45376
```

So we see that the estimated treatment difference is precisely the sam in the three cases, but the estimated s.e. is smallest for the ANCOVA approach. When we explicit contril for the confoundr (baseline) which is not really a counfounder in this example.

And we can do the same with the random effects models:

```
> lgx <- reshape( acpx, varying=1:2, v.names="score", direction="long" )
> lgx <- transform( lgx, G2 = Relevel( interaction( gr, time ),
+                                      list( B=1:2 ) ) )
> round( ci.lin( lmer( score ~ gr*factor(time) + (1|id), data=lgx ) ), 4 )
                           Estimate StdErr       z      P    2.5%   97.5%
(Intercept)                 60.4000 2.9907 20.1960 0.0000 54.5383 66.2617
grAcupuncture                0.0000 4.3132  0.0000 1.0000 -8.4538  8.4538
factor(time)2                1.8963 2.9480  0.6433 0.5201 -3.8816  7.6742
grAcupuncture:factor(time)2 17.3037 4.2516  4.0699 0.0000  8.9706 25.6368
> round( ci.lin( lmer( score ~ G2 + (1|id), data=lgx ) ), 4 )
               Estimate StdErr       z      P    2.5%   97.5%
(Intercept)     60.4000 2.1416 28.2036 0.0000 56.2026 64.5974
G2Placebo.2      1.8963 2.7623  0.6865 0.4924 -3.5178  7.3104
G2Acupuncture.2 19.2000 2.8554  6.7242 0.0000 13.6036 24.7964
> round( ci.lin( lmer( score ~ G2 + (1|id), data=lgx ),
+                ctr.mat=rbind(diag(3),c(0,-1,1)) ), 4 )
     Estimate StdErr       z      P    2.5%   97.5%
[1,]  60.4000 2.1416 28.2036 0.0000 56.2026 64.5974
[2,]   1.8963 2.7623  0.6865 0.4924 -3.5178  7.3104
[3,]  19.2000 2.8554  6.7242 0.0000 13.6036 24.7964
[4,]  17.3037 3.6858  4.6947 0.0000 10.0796 24.5278
```

Again we see that the random effects model allowing for baseline difference is equivalent to analysis of change scores, whereas the model assuming equl baseline produces the same estimate, but with a substantial smaller s.e. also as before. And for the same reason.

So we could say that including the baseline in the model as predictor is controlling for base-line imbalance.

2.3 A simulation example

We set up a function to generate a wide dataset, and a subsequent function to make it long:

```
> gen.data <-
+ function( tau = 15,
+          sigma = 5,
+             n1 = 50,
+             n2 = 50,
+             mu = 50,
+           beta = -5,
+          delta = 0,
+          gamma = 10 )
+ {
+ mB <- mu + rep( c(0,delta), c(n1,n2) )
+ mF <- mB + rep( c(0,gamma), c(n1,n2) ) + beta
+ ai <- rnorm( n1+n2, 0, tau )
+ yB <- rnorm( n1+n2, mB+ai, sigma )
+ yF <- rnorm( n1+n2, mF+ai, sigma )
```

```
+ data.frame( yF, yB, bb=factor( rep(0:1,c(n1,n2)),
+                                labels=c("Pl","Tr") ) )
+ }
> wd2long <-
+ function( ss )
+ {
+ n1 <- table( ss$bb )[1]
+ n2 <- table( ss$bb )[2]
+ data.frame( yy = with( ss, c( yB, yF ) ),
+             ii = rep( 1:(n1+n2), 2 ),
+             tt = factor( rep( 1:2, each=(n1+n2) ),
+                          labels=c("Bl","FU") ),
+             bb = factor( rep ( rep(0:1,c(n1,n2)), 2 ),
+                          labels=c("Pl","Tr") ) )
+ }
```

Once we have set up the functions we can simulate two data sets, both generated by the same random effects model, but one restricted to only contain baseline measurements above a certain quantile:

```
> set.seed( 724368 )
> n1 <- 50
> n2 <- 50
> ss <- gen.data( n1=n1, n2=n2, delta=0 )
> ll <- wd2long( ss )
> ff <- 4
> SS <- gen.data( n1=ff*n1, n2=ff*n2, delta=0 )
> SS <- subset( SS, yB > quantile( yB, 1-1/ff ) )
> LL <- wd2long( SS )
> cbind( ss=with( ss, table( bb ) ),
+        SS=with( SS, table( bb ) ) )
   ss SS
Pl 50 52
Tr 50 48
```

Here is a graphical display of the two data sets

```
> par( mfrow=c(1,2), mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> with( ss, plot( yB, yF,
+                 pch=16, col=c("limegreen","red")[bb],
+                 xlim=c(0,100), ylim=c(0,100) ) )
> with( SS, plot( yB, yF,
+                 pch=16, col=c("limegreen","red")[bb],
+                 xlim=c(0,100), ylim=c(0,100) ) )
```

We can take a look at the results from the analyses of the two data sets, simulated with a treatment effect of 10, a time effect of -5, and a baseline imbalance of 0.

```
> library( lme4 )
> library( Epi )
> round( ci.lin( lm( yF ~ yB + bb, data = ss ) ), 4 )
            Estimate StdErr       z      P    2.5%   97.5%
(Intercept)   1.9150 2.1147  0.9056 0.3652 -2.2297  6.0597
yB            0.8731 0.0366 23.8613 0.0000  0.8014  0.9448
bbTr          8.8177 1.3919  6.3352 0.0000  6.0897 11.5458
> round( ci.lin( lmer( yy ~ tt + I((bb=="Tr")*(tt=="FU")) + (1|ii), data=ll ) ), 4 )
                             Estimate StdErr       z P    2.5%   97.5%
(Intercept)                   49.6626 1.8598 26.7025 0 46.0173 53.3078
ttFU                          -4.4627 1.0246 -4.3556 0 -6.4708 -2.4545
I((bb == "Tr") * (tt == "FU"))  8.9698 1.4346  6.2524 0  6.1580 11.7816
> round( ci.lin( lmer( yy ~ tt + bb + I((bb=="Tr")*(tt=="FU")) + (1|ii), data=ll ) ), 4 )
                             Estimate StdErr       z      P     2.5%   97.5%
(Intercept)                   51.1978 2.6342 19.4360 0.0000  46.0349 56.3607
ttFU                          -4.5814 1.0347 -4.4278 0.0000  -6.6094 -2.5535
bbTr                          -3.0705 3.7253 -0.8242 0.4098 -10.3719  4.2309
I((bb == "Tr") * (tt == "FU"))  9.2073 1.4633  6.2922 0.0000   6.3394 12.0753
```

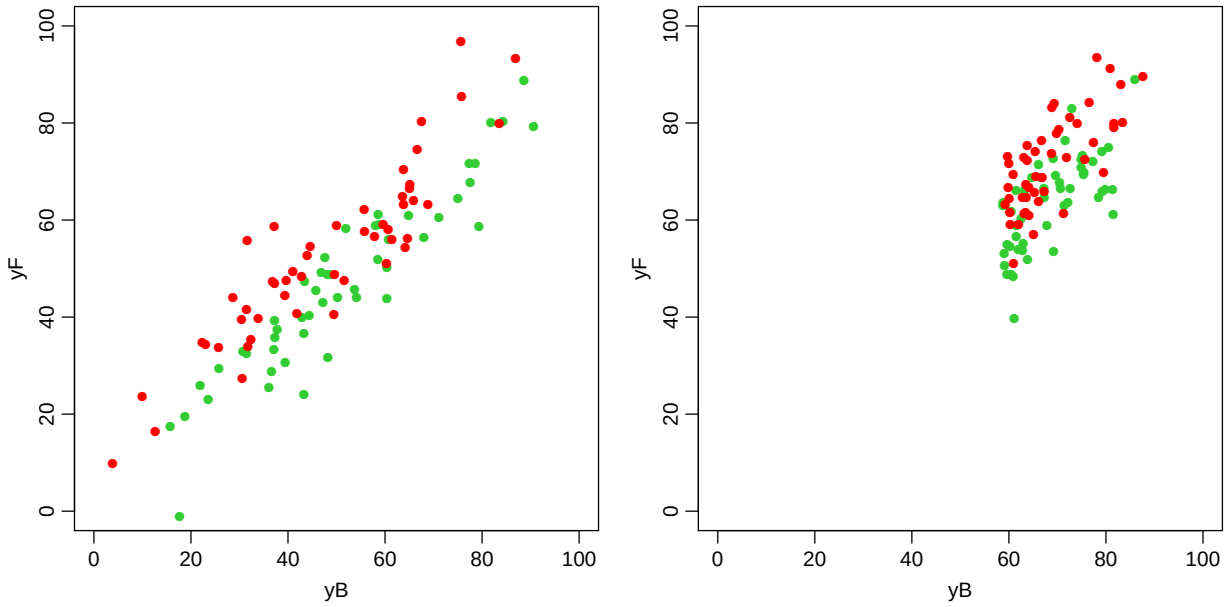

Figure 2.2: *Two data sets generated from the same base model, the latter restricted to those with baseline above a threshold.*

```
> round( ci.lin( lm( yF ~ yB + bb, data = SS ) ), 4 )
            Estimate StdErr      z      P    2.5%   97.5%
(Intercept)   2.9059 6.1598 0.4718 0.6371 -9.1670 14.9789
yB            0.8856 0.0891 9.9352 0.0000  0.7109  1.0603
bbTr          8.1436 1.3447 6.0559 0.0000  5.5079 10.7792
> round( ci.lin( lmer( yy ~ tt + I((bb=="Tr")*(tt=="FU")) + (1|ii), data=LL ) ), 4 )
                              Estimate StdErr       z P    2.5%   97.5%
(Intercept)                    68.4222 0.8573 79.8138 0 66.7420 70.1024
ttFU                           -4.9450 0.8987 -5.5027 0 -6.7064 -3.1837
I((bb == "Tr") * (tt == "FU"))  8.1881 1.2396  6.6052 0  5.7584 10.6177
> round( ci.lin( lmer( yy ~ tt + bb + I((bb=="Tr")*(tt=="FU")) + (1|ii), data=LL ) ), 4 )
                              Estimate StdErr       z      P    2.5%   97.5%
(Intercept)                    68.3122 1.1932 57.2503 0.0000 65.9735 70.6508
ttFU                           -4.9111 0.9346 -5.2546 0.0000 -6.7429 -3.0793
bbTr                            0.2293 1.7223  0.1331 0.8941 -3.1463  3.6048
I((bb == "Tr") * (tt == "FU"))  8.1173 1.3490  6.0172 0.0000  5.4733 10.7614
```

Bibliography

[1] D. G. Altman and C. J. Dore. Baseline comparisons in randomized clinical trials. *Stat Med*, 10(5):797–799, May 1991.

[2] S. Senn. Baseline comparisons in randomized clinical trials. *Stat Med*, 10(7):1157–1159, Jul 1991.

[3] S. J. Senn. Covariate imbalance and random allocation in clinical trials. *Stat Med*, 8(4):467–475, Apr 1989.

[4] Andrew J Vickers and Douglas G Altman. Analysing controlled trials with baseline and follow-up measurements. *British Medical Journal*, 323:1123–1124, November 2001.