# Growth Curve Example with Time-Varying Covariate

For all of the examples below, the health variable has been centered so that poor = -2, fair = -1, good = 0, very good = 1, and excellent = 2, using a theoretically chosen point. Grand-mean or within-person centering (i.e., centering within context or group-mean centering) also could be used.

## SPSS

```
*center health at the middle value (theoretical based centering).
recode health (1=-2) (2=-1) (3=0) (4=1) (5=2).

*genlinmixed requires string id variable.
STRING id (A4).
COMPUTE id = STRING(rid, F4.0).

*time was a nominal variable, convert it to scale.
variable level time (scale).

GENLINMIXED
/DATA_STRUCTURE SUBJECTS=id
/FIELDS TARGET= depress
/TARGET_OPTIONS DISTRIBUTION=NORMAL LINK=IDENTITY
/BUILD_OPTIONS DF_METHOD=SATTERTHWAITE COVB=ROBUST
/FIXED  EFFECTS= time health USE_INTERCEPT=TRUE
/RANDOM EFFECTS=time health USE_INTERCEPT=TRUE SUBJECTS=id
COVARIANCE_TYPE=UNSTRUCTURED.

*this is the syntax for nonrobust estimation with Kenward-Roger SEs.
*MIXED depress WITH time health
  /METHOD = REML
  /PRINT = SOLUTION TESTCOV HISTORY
  /FIXED = time health | SSTYPE(3)
  /RANDOM = INTERCEPT  time health | SUBJECT(rid) COVTYPE(UN)
 /CRITERIA=DFMETHOD(KENWARDROGER).
```

### Generalized Linear Mixed Models

**Model Summary**

| | | |
|---|---|---|
| Target | | depress Summed CESD score |
| Probability Distribution | | Normal |
| Link Function | | Identity |
| Information Criterion | Akaike Corrected | 4853.299 |
| | Bayesian | 4884.984 |

Information criteria are based on the -2 log likelihood (4839.137) and are used to compare models. Models with smaller information criterion values fit better.

**Coefficients of Determination**

| | | |
|---|---|---|
| Pseudo-R Square Measures | Marginal | .100 |
| | Conditional | .611 |

*Note: the marginal value is for the approximate variance accounted for by the fixed effects and the conditional is both fixed and random effects together.*

**Fixed Coefficients** [a]

| | | | | | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| Model Term | Coefficient | Std. Error | t | Sig. | Lower | Upper |
| Intercept | 14.562 | .6763 | 21.531 | <.001 | 13.227 | 15.896 |
| time | -2.061 | .2851 | -7.230 | <.001 | -2.623 | -1.499 |
| health | -2.442 | .3871 | -6.309 | <.001 | -3.206 | -1.679 |

Probability distribution: Normal
Link function: Identity

a. Target: Summed CESD score

### Covariance Parameters

**Random Effect**

| Random Effect Covariance | Estimate | Std. Error | Z | Sig. | 95% Confidence Interval Lower | 95% Confidence Interval Upper |
|---|---|---|---|---|---|---|
| UN (1,1) | 58.679 | 10.128 | 5.793 | <.001 | 41.836 | 82.301 |
| UN (2,1) | -7.322 | 3.777 | -1.938 | .053 | -14.725 | .082 |
| UN (2,2) | 1.874 | 2.490 | .752 | .452 | .139 | 25.348 |
| UN (3,1) | -9.760 | 4.117 | -2.371 | .018 | -17.829 | -1.691 |
| UN (3,2) | .500 | 1.744 | .286 | .775 | -2.919 | 3.918 |
| UN (3,3) | 5.756 | 2.855 | 2.016 | .044 | 2.178 | 15.214 |

Covariance Structure: Unstructured
Subject Specification: id

## R

Special code is needed to suppress the default estimation of multiple intercepts in the `lmer` function in the `lme4` package whenever there is more than one random slope. Use 0+ before the name of all slope variable random effects after the first one. Here, `time`, which is the first random slope mentioned, does not need a preceding 0+.

```
> library(lme4)
> #health as a time-varying covariate
> model1 <- lmer(depress ~ time + health + (time|rid) + (0+health|rid), data = mydata,REML=TRUE)
> summary(model1)
Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: depress ~ time + health + (time | rid) + (0 + health | rid)
   Data: mydata

REML criterion at convergence: 4851.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1883 -0.4793 -0.1445  0.3575  5.0283

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 rid      (Intercept) 46.843   6.844
          time         1.225   1.107    -0.73
 rid.1    health       2.647   1.627
 Residual             34.673   5.888
Number of obs: 702, groups:  rid, 234

Fixed effects:
            Estimate Std. Error       df t value             Pr(>|t|)
(Intercept)  14.8266     0.6251 274.1977  23.720 < 0.0000000000000002 ***
time         -2.0417     0.2852 232.4308  -7.158      0.0000000000106 ***
health       -2.6376     0.3669 218.3959  -7.189      0.0000000000102 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
       (Intr) time
time   -0.578
health -0.368  0.045
> VarCorr(model1)
 Groups   Name        Std.Dev. Corr
 rid      (Intercept) 6.8442
          time        1.1067   -0.734
 rid.1    health      1.6270
 Residual             5.8884
> VarCorr(model1) #provides variances in variance form
 Groups   Name        Std.Dev. Corr
 rid      (Intercept) 6.8442
          time        1.1067   -0.734
 rid.1    health      1.6270
 Residual             5.8884
> rand(model1)  #LR test compared to empty using mixture distribution
boundary (singular) fit: see help('isSingular')
ANOVA-like table for random-effects: Single term deletions

Model:
depress ~ time + health + (time | rid) + (0 + health | rid)
                          npar  logLik    AIC    LRT Df Pr(>Chisq)
<none>                       8 -2425.8 4867.5
time in (time | rid)         6 -2428.1 4868.2 4.7024  2    0.09525 .
health in (0 + health | rid)    8 -2426.7 4869.3 1.7729  0
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> confint(model1) #profile likelihood intervals for better random effects tests
Computing profile confidence intervals ...
                  2.5 %     97.5 %
.sig01       5.75785821  7.9607605
.sig02      -1.00000000 -0.1857486
.sig03       0.06876013  2.4634925
.sig04       0.00000000  2.7303801
.sigma       5.32338002  6.3835116
(Intercept) 13.59506238 16.0578266
time        -2.60211296 -1.4804328
health      -3.37708994 -1.8986010
> library(MLMusingR)
> robust_mixed(model1) #get robust SE estimates

Standard error type = CR2
Degrees of freedom = Satterthwaite

            Estimate  mb.se robust.se t.stat  df              Pr(>t)
(Intercept)   14.827  0.625     0.678 21.870 195 <0.0000000000000002 ***
time          -2.042  0.285     0.285 -7.162 232 <0.0000000000000002 ***
health        -2.638  0.367     0.410 -6.426 148 <0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The warning `boundary (singular) fit` that occurred after using the `rand()` function suggests there may be something wrong with the estimation. Prior to requesting random slope tests, `lme4` does not provide any hints that something might be wrong in this model. You also may sometimes see a warning message after the profile likelihood confidence intervals that mentions a “non-monotonic profile” or “bad spline fit.”[1] Both warning messages are hinting at a problem estimating the random effects. The boundary singularity message means that the within-person variance estimate is problematic (and one or more values in the matrix have been set to 0). The problem stems from too many random effects in the model with only three time points, a problem which becomes clearer by attempting to estimate the model in HLM.

## HLM

This model differs from the SPSS and R models because the HLM program indicated that it did not have enough degrees of freedom to estimate the original model allowing for the `HEALTH` slope to vary. Consequently, I tested the model in which there is not a random effect for the covariate—`HEALTH` is assumed to have the same relationship to the outcome across respondents.

There are only three time points (the same as only three cases per group) and this puts restrictions on the number of random effects that can be estimated (either theoretically sometimes or practically) because of the within-person covariance matrix which will have six elements total (each of the three variances for intercept and two slopes and their three covariances) cannot be estimated in HLM. The fact that the parameterization in HLM will not allow for the estimation of an additional slope makes me suspicious that SPSS and R perhaps should not be producing solutions at all in this circumstance.

**Summary of the model specified**
**Level-1 Model**

$$DEPRESS_{ti} = \pi_{0i} + \pi_{1i}*(TIME_{ti}) + \pi_{2i}*(HEALTH_{ti}) + e_{ti}$$

**Level-2 Model**

$$\pi_{0i} = \beta_{00} + r_{0i}$$
$$\pi_{1i} = \beta_{10} + r_{1i}$$
$$\pi_{2i} = \beta_{20}$$

**Mixed Model**

$$DEPRESS_{ti} = \beta_{00} + \beta_{10}*TIME_{ti} + \beta_{20}*HEALTH_{ti} + r_{0i} + r_{1i}*TIME_{ti} + e_{ti}$$

$\sigma^2 = 36.15791$

[1] This is the message I received on another example.

```
Warning messages:
1: In FUN(X[[i]], ...) : non-monotonic profile for .sig02
2: In confint.thpr(pp, level = level, zeta = zeta) :
  bad spline fit for .sig02: falling back to linear interpolation
```

τ

| | | |
|---|---|---|
| INTRCPT1,$\pi_0$ | 47.53823 | -5.13561 |
| TIME,$\pi_1$ | -5.13561 | 1.10917 |

τ (as correlations)

| | | |
|---|---|---|
| INTRCPT1,$\pi_0$ | 1.000 | -0.707 |
| TIME,$\pi_1$ | -0.707 | 1.000 |

| Random level-1 coefficient | Reliability estimate |
|---|---|
| INTRCPT1,$\pi_0$ | 0.612 |
| TIME,$\pi_1$ | 0.058 |

The value of the log-likelihood function at iteration 856 = -2.425749E+003

**Final estimation of fixed effects:**

| Fixed Effect | Coefficient | Standard error | *t*-ratio | Approx. *d.f.* | *p*-value |
|---|---|---|---|---|---|
| For INTRCPT1, $\pi_0$ | | | | | |
| INTRCPT2, $\beta_{00}$ | 14.839168 | 0.624616 | 23.757 | 233 | <0.001 |
| For TIME slope, $\pi_1$ | | | | | |
| INTRCPT2, $\beta_{10}$ | -2.018598 | 0.286710 | -7.041 | 233 | <0.001 |
| For HEALTH slope, $\pi_2$ | | | | | |
| INTRCPT2, $\beta_{20}$ | -2.644731 | 0.348884 | -7.581 | 233 | <0.001 |

**Final estimation of fixed effects (with robust standard errors)**

| Fixed Effect | Coefficient | Standard error | *t*-ratio | Approx. *d.f.* | *p*-value |
|---|---|---|---|---|---|
| For INTRCPT1, $\pi_0$ | | | | | |
| INTRCPT2, $\beta_{00}$ | 14.839168 | 0.678546 | 21.869 | 233 | <0.001 |
| For TIME slope, $\pi_1$ | | | | | |
| INTRCPT2, $\beta_{10}$ | -2.018598 | 0.283545 | -7.119 | 233 | <0.001 |
| For HEALTH slope, $\pi_2$ | | | | | |
| INTRCPT2, $\beta_{20}$ | -2.644731 | 0.415843 | -6.360 | 233 | <0.001 |

**Final estimation of variance components**

| Random Effect | Standard Deviation | Variance Component | *d.f.* | $\chi^2$ | *p*-value |
|---|---|---|---|---|---|
| INTRCPT1, $r_0$ | 6.89480 | 47.53823 | 233 | 598.16785 | <0.001 |
| TIME slope, $r_1$ | 1.05317 | 1.10917 | 233 | 243.98745 | 0.297 |
| level-1, $e$ | 6.01314 | 36.15791 | | | |

**Statistics for current covariance components model**
Deviance = 4851.497101

## Write Up

*(These results are from the SPSS output—I am using these results simply to show an example in which the covariate has a random slope, but knowing that R and HLM had difficulties estimating the model, raises concerns about their validity. After seeing an error in HLM or R, I would rerun the model, first eliminating the random slopes for health. Diagnostics plots would also advisable to explore whether residuals were nonnormal or heteroscedastic, or whether nonlinear effects might be present).*

A growth curve model with a time-varying covariate was tested to investigate the change in depression over time controlling changes in health over time. Random slopes for the time and the health variable were estimated in the model. The average depression score at baseline was 14.56 and these values varied significantly across participants, $\tau_0^2$ = 58.68, $z$ = 5.79, $p$ < .001. There was a significant decline in depression over time after controlling for changes in health, $\gamma_{10}$ = -2.05, $t$ = -7.23, $p$ < .001, which indicates that depression scores decreased by approximately two points every six months. Growth curves did not vary significantly, $\tau_1^2$ = 1.87, $z$ = .75, $p$ = .23, however, suggesting that widows declined in depression at the similar rates. Health was significant related to depression at each time point, $\gamma_{20}$ = -2.44, $t$ = 6.31, p < .001, indicating that those in better health had lower depression scores. The relationship between health and depression also varied across participants, $\tau_2^2$ = 5.76, $z$ = 2.02, $p$ = .02.