Regression to the mean and RCTs for continuous outcomes

Niels Henrik Bruun

Research data and statistics, AaUH

The why and what you get

- Barnett, Pols, and Dobson (2005) describes RTM and how to remedy the RTM effects
  - RTM is a statistical phenomenon that occurs when repeated measurements are made on the same subject or unit of observation. *It happens because values are observed with random error.*
  - The effect of RTM can also be compounded by categorizing subjects into groups based on their baseline measurement(s).
  - Solution 2 is baseline adjustment with baseline values
- Twisk et al. (2018) presents the 3 ways of analyzing RCTs
  - recommend using "longitudinal analysis of covariance or repeated measures without baseline treatment effect"
- Matheson (2019) argues for need of accounting for the reliability when designing new studies
  - Highlight the need to use previous Test-retest studies in planning new RCTs
  - Demonstrates what extra information that can be gained

I will

- Give a short introduction to the regression-to-the-mean
- Present approaches to analyzing RCTs through an example
- demonstrate that baseline adjustment of the outcome is important
- argue that every RCT with baseline adjustment is in fact also a Test-retest study
- highlight the importance of the ICC in RCTs
- propose that reporting the Test-retest results should be part of every RCT with baseline adjustment

Table of content

From abstract[1][2]

- Can medication reduce the blood pressure for patients with diabetes and kidney disease?
- One week randomised single blind trial of captopril versus placebo

[1] Hommel et al. (1986)
[2] Matthews (2006)

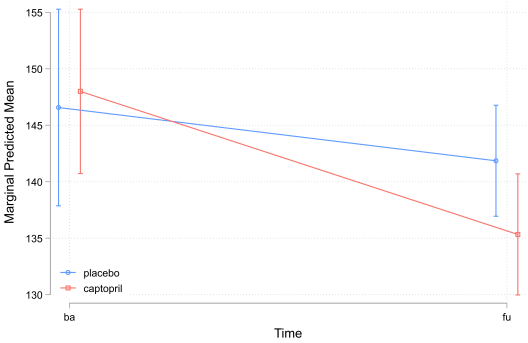
Summary[3][4]

Baseline characteristics

Columns by: Treatment	placebo	captopril
n	7	9
Sex (female), n	2	0
Age (years), mean (sd)	32.4 (9.0)	30.6 (9.5)
Duration of diabetes (years), mean (sd)	23.7 (8.8)	18.1 (4.3)
Retinopathy (simplex), n	4	3
Insulin dose (U/kg/day), mean (sd)	0.6 (0.1)	0.7 (0.2)

Non-significant baseline effect

Mean and CI of mean of outcome at baseline and at follow-up



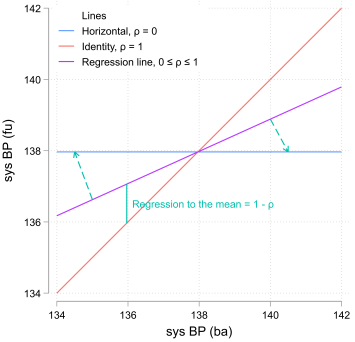
[3] Hommel et al. (1986)
[4] Matthews (2006)

Regression to the mean (RTM), continuous outcomes[5][6]

- OLS slope from regressing FU on BA

$$\beta = \frac{\rho \cdot \sigma_{FU}}{\sigma_{BA}} = \rho, \textit{if } \sigma_{BA} = \sigma_{FU}$$

  - between 0 (No relation) and 1 (Perfection/identity)
- the intersection between the regression line and the identity line is where $E[BA] = E[FU]$
- Regression to the mean is perfection minus correlation
  - the higher correlation the lesser regression to the mean
- The regression line is the true adjustment effect for the baseline values
- The correlation squared is the consistency ICC (test-retest), Nakagawa and Schielzeth (2010)



[5] Campbell and Kenny (1999)
[6] Barnett, Pols, and Dobson (2005)

The means and variances for the model approaches

- Sampling bias imply imbalance between the treatment group means
  - Mean estimates are never their true value
  - Splitting into, e.g., two groups, one group mean is the higher
- Due to imbalance the model approaches may lead to biased estimates, Matthews (2006) p.84:
- The (ADJ) model has the lowest variance: $Var[FU|BA = ba] = Var[FU] \cdot (1 - \rho^2)$, Matthews (2006) p.83
  - Note the importance of ρ^2

Model approaches in wide datasets using Stata

Look at the estimated intercept

Do not adjust (FUwide) - only use follow-up

- The intercept (_cons) is the expected value for the control group at follow-up
- Baseline effects exists even if not measured
- Only model with power calculation in Stata

Analyze the change from baseline (CHGwide)

- Each individual has their own intercept (their baseline value)
- Adjust the effect from FU with the difference in baseline means
- Missing values at follow-up implies also removing the baseline values
- equation (3a) in Twisk et al. (2018)

Adjust with baseline regression/RTM (ADJwide), wide dataset

- The individual intercept is predicted by the baseline value (adjusting for RTM)
- Missing values at follow-up implies also removing the baseline values
- equation (1a) in Twisk et al. (2018)

Stata code

Do not adjust (FUwide) - only use follow-up

```
. glm sysfu i.treatment, vce(robust)
. estimates store FUwideide
```

Analyze the change from baseline (CHGwide)

```
. constraint 1 _b[sysba] = 1
. glm sysfu i.treatment c.sysba, vce(robust) constraint(1)
. estimates store CHGw
```

Adjust with baseline regression (ADJwide)

```
. glm sysfu i.treatment c.sysba, vce(robust)
. estimates store ADJw
```

Model approaches in long datasets using Stata

Look at the change from baseline mean for each treatment group

Do not adjust (FUlong) - only use follow-up
- equation (2a) in Twisk et al. (2018)

Analyze the change from baseline (CHGlong)
- Missing values at follow-up do NOT imply removing the baseline values
- equation (3a) in Twisk et al. (2018)

Handling RTM by costraint
- Same mean at baseline means no baseline treatment effect
- Missing values at follow-up do NOT imply removing the baseline values
- equation (2c) in Twisk et al. (2018)
- Note: Option *coefl* is nice when building constraints

Stata code

Making the dataset long
```
. reshape long sys, i(id) j(tm) string
. strtonum tm, base(0)
. label variable tm "Time"
```

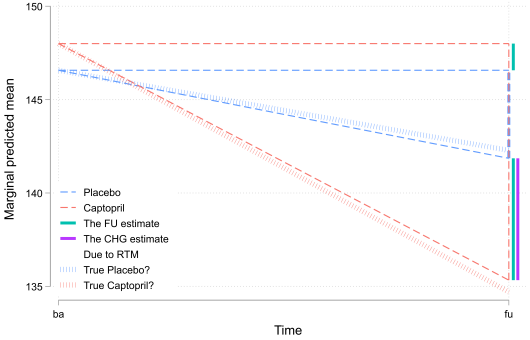
Doing the GLMM regression getting the FUlong estimate
```
. meglm sys i.treatment##i.tm || id:, vce(robust) ///
        noheader nolog
. xlincom (1.treatment=_b[1.treatment] ///
        + _b[1.treatment#1.tm]), post
. estimates store FUlong
```

Doing the GLMM regression getting the CHGw estimate
```
. meglm sys i.treatment##i.tm || id:, vce(robust) ///
        noheader nolog
. xlincom (1.treatment=_b[1.treatment#1.tm]), post
. estimates store CHGlong
```

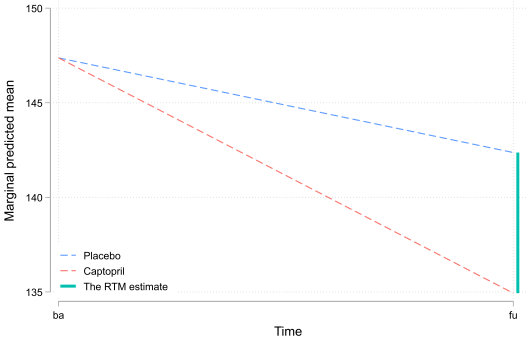
Using the constraint of no baseline treatment effect
```
. constraint 1  0.tm#0.0.treatment = 0.tm#1.treatment
. meglm sys i.tm i.treatment#i.tm || id:, vce(robust) ///
  noheader nolog constraint(1)
. xlincom (1.treatment=_b[1.treatment#1.tm]), post
. estimates store RTM2
```

Model approaches in long datasets visualized

- Imbalance at baseline means RTM effects
- Looking at the differences in CHGlong means a rescale to zero and hence ignore the full baseline effect



- one common baseline mean implies no RTM effect
- estimates are slightly different



Comparison of methods

	FUwide effect / SE	FUlong effect / SE	CHGwide effect / SE	CHGlong effect / SE	ADJwide effect / SE	RTM2 effect / SE
Treatment effect	-6.524	-6.524	-7.952	-7.952	-7.178	-7.434
	3.711	3.711	4.084	4.084	2.703	2.916
RTM adjustment			1.000		0.458	
					0.131	

- Biased estimates
  - Analyzing only at Follow-up (FU)
  - analyzing change (CHG)
- The ADJwide and then the RTM2 estimates has the lowest standard error

On RCTs and Test-retest reliability

- Every RCT is also a Test-retest reliability study for the control group
- The correlation squared is the consistency ICC, Nakagawa and Schielzeth (2010)
- The ICC is a quality measure of the RCT study
  - We cannot use a bathroom scale to reliably measure and compare the weight of bricks (low ICC)
  - The ICC is often much lesser than expected from to the instrument precision alone

- $\rho = ICC = \frac{Variation\ explained}{Variation\ explained + measurement\ error}$
- *measurement error* depends on
  - the measure instrument
  - the operator
  - the intra biological variation
  - the chosen model
- *Variation explained* depends on
  - the inter biological variation
  - the chosen model
- Correlation decreases over time (time series)

A power calculation example

Having baseline values included in the design lead to

- unbiased estimates
- more power in the study
- require smaller sample size to measure an effect

Classical power calculation

```
. power twomeans 145 150, sd(12)
Estimated sample sizes for a two-sample means test

t test assuming sd1 = sd2 = sd
H0: m2 = m1  versus  Ha: m2 != m1

Estimated sample sizes:
            N =       184
  N per group =        92
```

Using the correlation

```
. correlate sysfu sysba if !treatment
(obs=7)

             |    sysfu    sysba
-------------+------------------
       sysfu |   1.0000
       sysba |   0.8007   1.0000
```

And assuming the RTM baseline adjustments

```
. power twomeans 145 150, sd(`=12*sqrt(1-0.8^2)')
Estimated sample sizes for a two-sample means test

t test assuming sd1 = sd2 = sd
H0: m2 = m1  versus  Ha: m2 != m1

Estimated sample sizes:
            N =        68
  N per group =        34
```

Conclusions

- Due to measurement errors, there is always RTM effect in RCTs
- We can do better in RCTs than just analyzing FU values or change values
  - unbiased estimates
  - more power in the study
- To handle RTM effects, a baseline adjustment is necessary
  - The study becomes more powerful
  - There should be no baseline treatment effect

Reflections

- Every RCT study with continuous outcome and baseline adjustment should report the correlation between baseline and follow-up values / the consistency test-retest ICC
  - and the standard error of measurement (SEM)
- The correlation is a quality measure of the study (higher values better)
  - Should meta-analysis be stratified by correlations?
  - The correlations are the basis for better future power calculations

References

Barnett, Adrian G, Jolieke C van der Pols, and Annette J Dobson. 2005. "Regression to the mean: what it is and how to deal with it." *International Journal of Epidemiology* 34 (1): 215–20. https://doi.org/10.1093/ije/dyh299.

Campbell, D. T., and D. A. Kenny. 1999. *A Primer on Regression Artifacts*. Methodology in the Social Sciences. Guilford Publications. https://books.google.dk/books?id=mu1QzgEACAAJ.

Hommel, E, H H Parving, E Mathiesen, B Edsberg, M Damkjaer Nielsen, and J Giese. 1986. "Effect of Captopril on Kidney Function in Insulin-Dependent Diabetic Patients with Nephropathy." *BMJ* 293 (6545): 467–70. https://doi.org/10.1136/bmj.293.6545.467.

Matheson, Granville J. 2019. "We Need to Talk About Reliability: Making Better Use of Test-Retest Studies for Study Design and Interpretation." *PeerJ (San Francisco, CA)* 2019 (5): e6918–e6918.

Matthews, J. N. S. 2006. *Introduction to Randomized Controlled Clinical Trials*. Chapman & Hall/Crc Texts in Statistical Science. CRC Press. https://books.google.dk/books?id=gWXLBQAAQBAJ.

Nakagawa, Shinichi, and Holger Schielzeth. 2010. "Repeatability for Gaussian and Non-Gaussian Data: A Practical Guide for Biologists." *Biological Reviews of the Cambridge Philosophical Society* 85 (4): 935–56.

Twisk, J., L. Bosman, T. Hoekstra, J. Rijnhart, M. Welten, and M. Heymans. 2018. "Different Ways to Estimate Treatment Effects in Randomised Controlled Trials." *Contemporary Clinical Trials Communications* 10: 80–85. https://doi.org/https://doi.org/10.1016/j.conctc.2018.03.008.