Longitudinal Data Analysis: A Conceptual Approach

Maren Olsen, Ph.D.

May 9, 2000

Outline

- Motivation
- Understanding longitudinal study data (steps BEFORE modeling)
- Change scores
- Repeated-measures ANOVA
- Marginal Models
- Linear mixed models
- Conclusions

Motivation, cont.

Definition: longitudinal data is characterized by repeated observations over time on the same set of individuals

Issue: repeated observations on the same individual tend to be *correlated*

$\Rightarrow$ Appropriate statistical analyses must take this correlation into account

Motivation, cont.

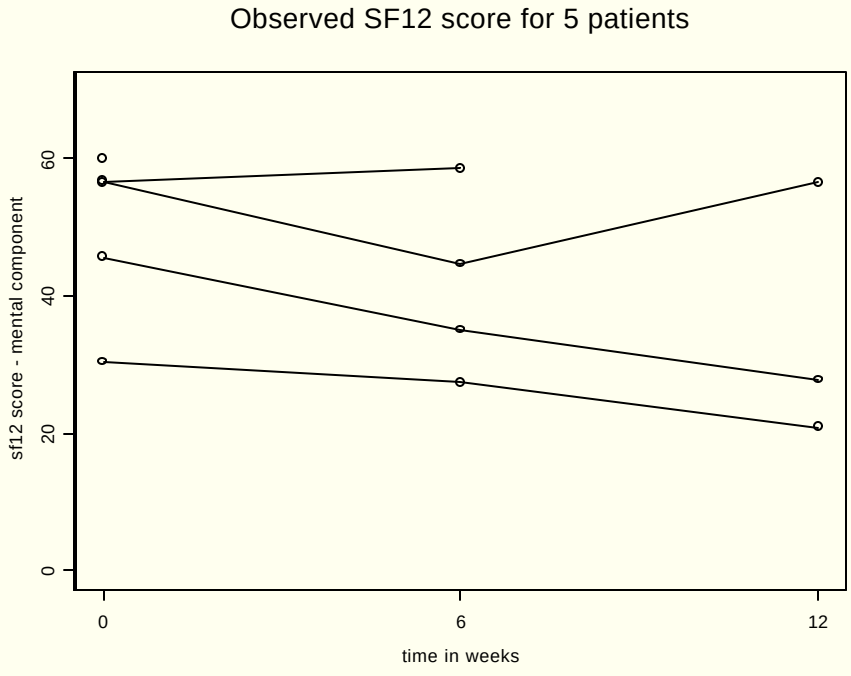
Examples of longitudinal study objectives:

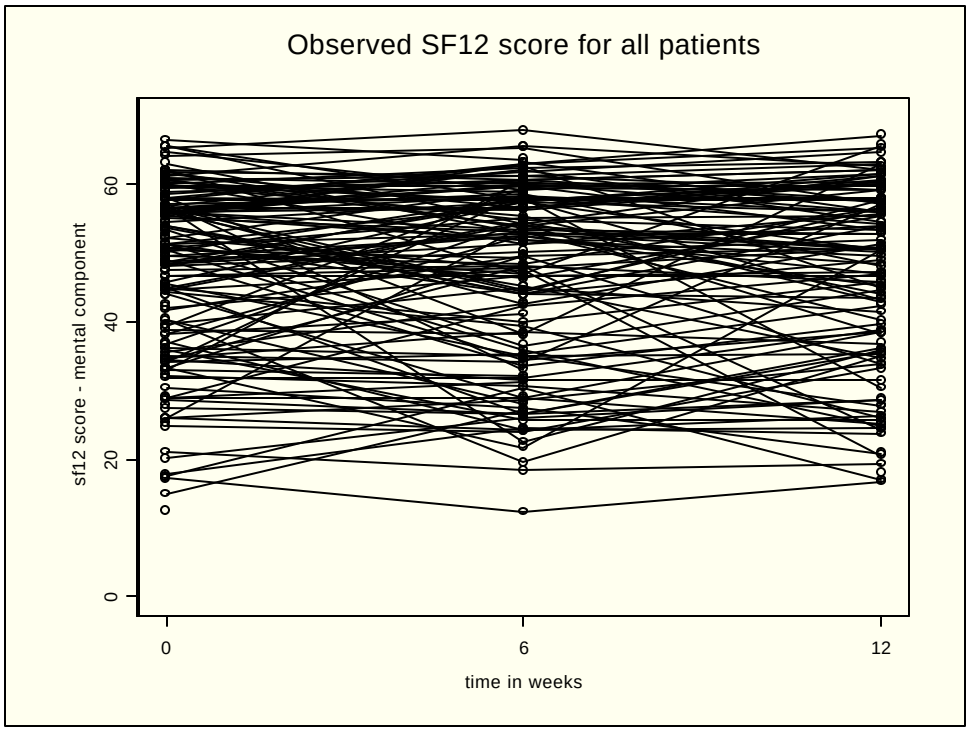
- Examine patients' general health quality of life (SF-12) to see if there is a change over time for ENT vs. non-ENT patients. Also examine how baseline comorbidities and baseline disease severity influence the SF-12 trend over time.
- The proportion of veterans with BP control who receive the physician-directed continuous quality improvement (CQI) intervention will be increased by 15% as compared to usual care over the two years.

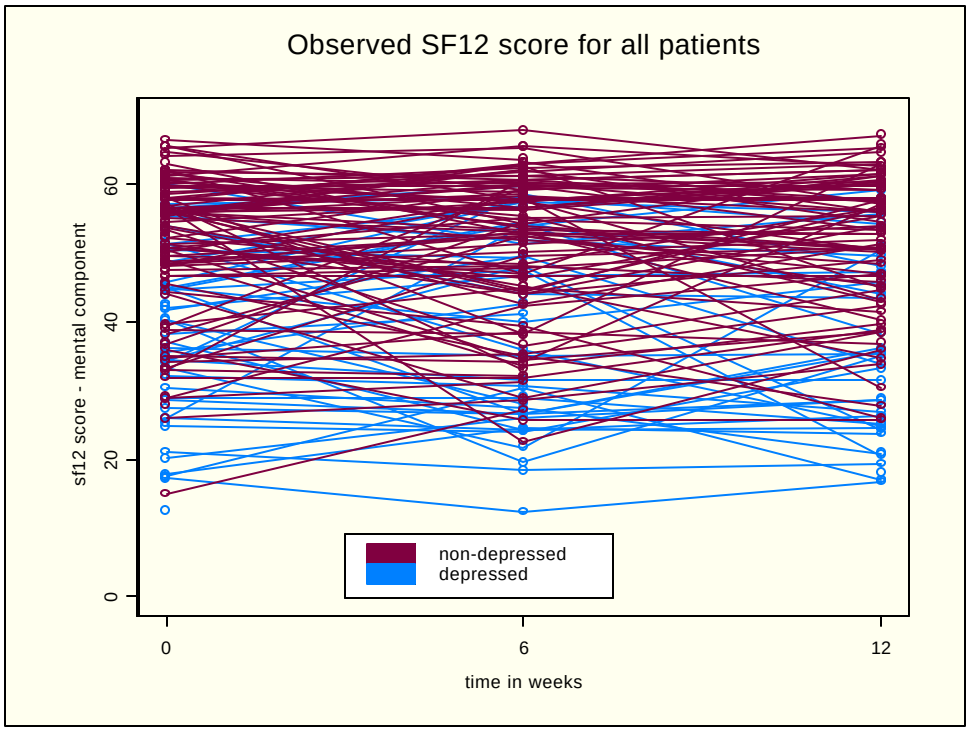
2. Understanding longitudinal study data

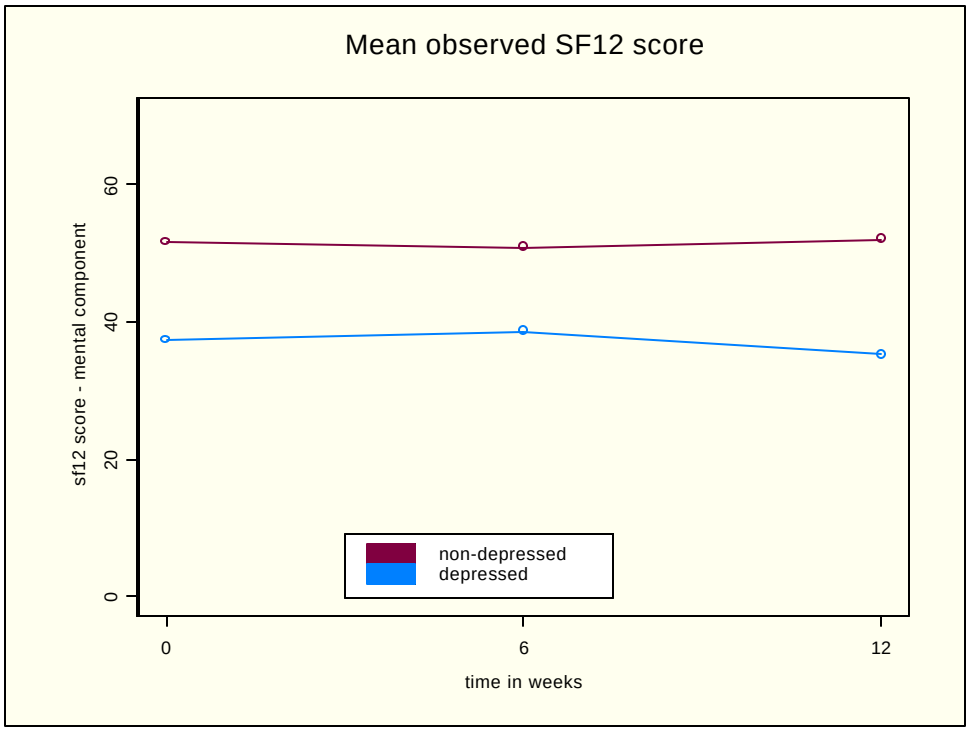
Plot observed individual trajectories of the response:

- helps to identify general trends within individuals over time
- may be able to detect non-linear change over time
- provides some information about the amount of inter-individual variability
- include different subgroups on same graph to explain inter-individual variability




Observed SF12 score for all patients
sf12 score - mental component
0
20
40
60
0
6
12
time in weeks


Observed SF12 score for all patients
sf12 score - mental component
0
20
40
60
non-depressed
depressed
0
6
12
time in weeks



What do we want from a longitudinal data model?

- address the longitudinal study objectives

Examine patients' general health quality of life (SF-12) to see if there is a change over time for ENT vs. non-ENT patients. Also examine how baseline comorbidities and baseline disease severity influence the SF-12 trend over time.

What do we want from a longitudinal data model?

- address the longitudinal study objectives
- characterize the process of change
    1. individual profiles of change
    2. subgroup and population-average profiles of change

What do we want from a longitudinal data model?

- address the longitudinal study objectives
- characterize the process of change
- lead to precise estimates with good power
- use **all** available information

Change Scores, cont.

Using the difference between **two** points in time to understand change.

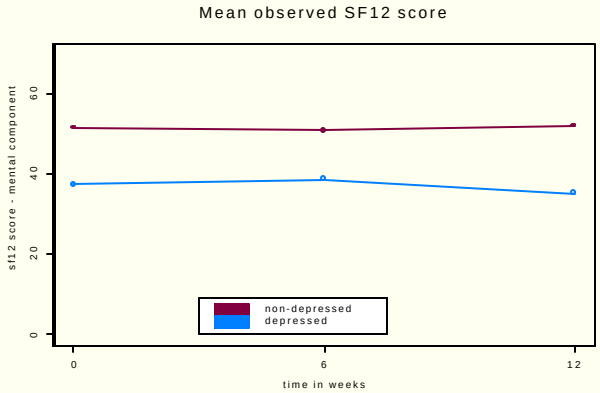
Limitations:

• tells the researcher **nothing** about the shape of each patient's trajectory. Did the change occur immediately after time 1 or was progress steady over the time interval?

• only using two time points leads to less precise information

• can only include data observed at **both** time points

4. Repeated-Measures ANOVA



Idea:

• analyze differences between observed means

• individual profiles of change are *noise*

Repeated-Measures ANOVA, cont.

Limitations:

- Data must be **balanced**
  - measurements occur at same times for all subjects
  - each subject has the same number of observations
- no distinction between unequally-spaced time points and equally-spaced time points
- does not produce a parameter that estimates rate of change over time for different treatment groups
- correlation among all observations within a subject is the same

5. Marginal Models

Idea:

- describe relationship between response variable and explanatory variables with a **population-average regression**
- account for within-subject correlation
  - more flexible correlation structures than ANOVA
  - examples include: first-order auto-regressive (AR1), unspecified

5. Marginal Models

Idea:

- describe relationship between response variable and explanatory variables with a **population-average regression**
- account for within-subject correlation

Two parts are modeled separately using a semi-parametric method called Generalized Estimating Equations (GEE).

5. Marginal Models

Idea:

- describe relationship between response variable and explanatory variables with a **population-average regression**
- account for within-subject correlation

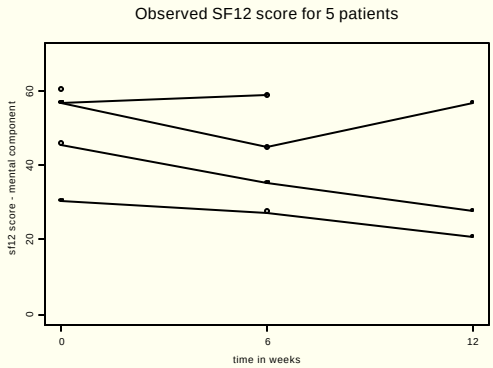
Comments:

• Popular method for non-normal responses

• robust method: estimates and standard errors are statistically consistent even if correlation structure is misspecified

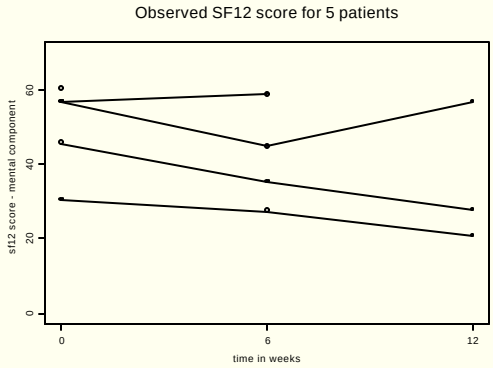
• models can be fit with PROC GENMOD

A linear mixed model has **two** parts.



- **within-subject part**:
  - a subject's change over time described by a *regression model* with a population-level mean and slope
  - error term in regression is the within-subject variation

A linear mixed model has **two** parts.



- **between-subject part**:
  - models variation between individuals' intercepts and slopes
  - captures relationship between individuals' intercepts and slopes (e.g. positive relationship between intercept and slope => patients with higher baseline measure have faster rate of change over time)

Linear Mixed Models, cont.

Extremely flexible analysis tool:

- unbalanced data
- unequally-spaced time points
- subjects may be observed at different time points
- including covariates is straightforward (same as regression)
- emphasis on **both** individual and population-average components

$\Rightarrow$ possible to estimate individual-level and population-level growth curve parameters

$\Rightarrow$ individual variability explicitly modeled

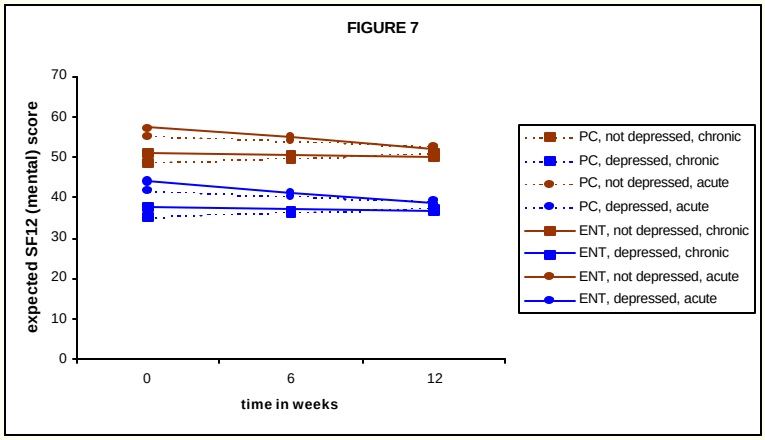
Linear Mixed Model: Example

Longitudinal study objective

Examine patients' general health quality of life (SF-12) to see if there is a change over time for ENT vs. non-ENT patients. Also examine how baseline comorbidities and baseline disease severity influence the SF-12 trend over time.

Possible predictors

- diagnosis site
- disease duration
- presence or absence of self-reported depression



At baseline:

• acute sinusitis patients had *higher* mental QOL than chronic sinusitis patients
• depressed patients had *lower* QOL scores than non-depressed patients

By 12 weeks:

• effects of baseline disease duration and diagnosis site have dissipated

Linear Mixed Model: Example, cont.

Results: between-subject part

Var (intercepts) = 8.28

• high variability among individuals' baseline SF-12 scores

Var (slopes) = 1.41

• low variability among individuals' change in SF-12 scores over time

Corr (intercepts, slopes) = 0.066

• no relationship between baseline level and rate of change over time

Linear Mixed Models, cont.

Extensions:

- additional levels of clustering

  $\Rightarrow$ longitudinal study where patients are also clustered within physicians

  $\Rightarrow$ software includes: HLM, MLn

Linear Mixed Models, cont.

Extensions:

- additional levels of nesting
- non-normal responses

  $\Rightarrow$ binary, count data

  $\Rightarrow$ software includes: MIXOR, HLM, MLn

  $\Rightarrow$ semicontinuous data

Linear Mixed Models, cont.

Extensions:

- additional levels of nesting
- non-normal responses
- latent response variables

  $\Rightarrow$ Structural Equation Modeling

  $\Rightarrow$ software includes: LISREL, EQS, Mplus

Linear Mixed Models, cont.

Extensions:

- additional levels of nesting
- non-normal responses
- latent response variables
- covariate measurement error

  $\Rightarrow$ Edwards (1994; 1995)

7. Conclusions

- Graphical approaches are a useful and necessary first step
- Marginal models and mixed effects models …
  - ✓ address the longitudinal study objectives
  - ✓ characterize the process of change
  - ✓ lead to precise estimates with good power
  - ✓ use **all** available information

7. Conclusions

- Graphical approaches are a useful and necessary first step
- Marginal models and mixed effects models
- Fitting these models takes **time** and **extra care**