Statistical Strategies for Long Term Clinical Trials in Psychopharmacology

Andrew C. Leon, Ph.D.

Weill Medical College of Cornell University

Funded, in part, by NIH Grant MH60447

Financial Disclosures

- NIH support
  - MH60447, MH068638, MH72878, MH64608, MH73175, MH62475, DA06534, DC007894
- Consulting relationships
  - Avera, Cortex, Cyberonics, and Eli Lilly
- DSMBs
  - Pfizer, Quintiles, and Vanda
- No equity interests
- No speakers bureaus

FDA PDAC Meeting: Long Term Efficacy for Psychiatric Drugs

- Most psychiatric disorders are chronic
- FDA does not require long term data for initial approval
- FDA typically asks for long term data after approval

- **Question to PDAC**: Is it a reasonable expectation that a sponsor would have accumulated data for both acute and longer-term efficacy trials at the time of filing of an application for a drug for the treatment of MDD?

TP Laughren. Sept 26, 2005 Memo for Oct 25, 2005 PDAC Meeting

Long Term RCTs in Psychopharmacology

- Conventional Approaches to Data Analysis
- Consider Challenges of Longitudinal Data
- Strategies for Analyzing Incomplete Data
- Recommendations

Goals of Long Term RCT Design

- Minimize bias in estimate effect of treatment over time
- Maintain type I error
- Sufficient statistical power
- Feasible and Applicable

Features of RCT Design

- Randomized group assignment
- Double-blinded assessments
- Control or comparison groups

Attrition Interferes with Goals of Long-term RCTs

Long term trials are more vulnerable to attrition.

- **Attrition can compromise randomization & group balance**.
  - Biased estimate of the treatment effect
- Smaller N reduces statistical power
- Limited generalizability

Leon et al, Biological Psychiatry (2006)

Attrition Rates

- Antipsychotic RCT's submitted to FDA (Khan, 2001, *AJP*)
- 7 RCTs; 1920 subjects
- Mean dropout rate at 6 weeks:
  - Placebo: **64%**
  - Investigational: **50%**
  - Active comparator: **56%**

Attrition Rates

Antipsychotic RCT's

(Kemmler et al., 2005, *AGP*)

Placebo Controlled RCTs		
	Placebo Arm	**Active Arm**
2nd Generation (N=11 RCTs)	60%	48%
Other Anti-psychotics (N=5 RCTs)	63%	55%

RCTs 12 weeks or less

Attrition Rates in Active Cells

Antipsychotic RCT's

(Kemmler et al., 2005, *AGP*)

	Placebo Controlled	Active Controlled
2nd Generation	48%	28%
Other Anti-psychotics	55%	37%

RCTs 12 weeks or less

Attrition in Long-term RCTs

- Attrition is expected to be more pronounced.
- Attrition rates over time - not accessible in most papers.
- CATIE (Lieberman et al., 2005)
  - About 45% discontinued assigned treatment by 6 months; 74% by 18 months.
  - Attrition rates over time are less clear.
- InterSePT (Meltzer et al. AGP 2003)
  - At 18 months: Endpoint provided by about 80%.
  - At 2 years: about 39% discontinued.

Prevent Loss of Subjects

Design RCTs to minimize attrition

- 7 steps to minimize missing data (Wisniewski, BiolPsych 2006)

Reduce subject burden

- More accessible assessments
  - Telephone calls, IVR, Home visits

Ethical guidelines protect subjects

- Guarantee subject's right to exit

Choosing a Statistical Procedure for Long Term RCTs

- What is the form of the dependent variable?
- What is sample size?

* How many assessment times per subject?

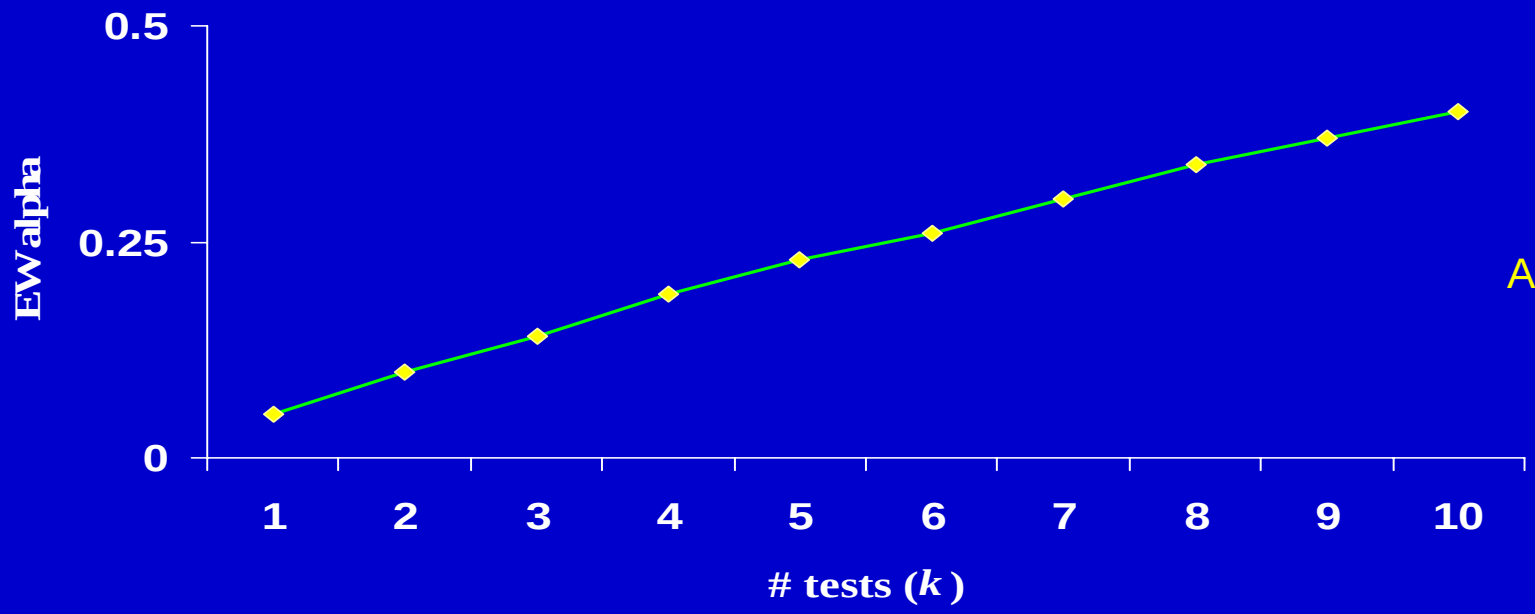
** Can S's with incomplete data be included in analyses?

Standard Cross-sectional Approaches

- χ^2 test H_0: $\pi_1 = \pi_2$
- *t*-test H_0: $\mu_1 = \mu_2$
- ANOVA H_0: $\mu_1 = \mu_2 = \mu_3$
- Logistic regression or ANCOVA to account for baseline severity

Cross-sectional Approaches

- Ignore the richness of multiple assessments in longitudinal studies.
- Fail to examine trends of change over time.
- Separate χ^2 tests (per month) elevates risk of Type I error:



Assumes: α=0.05 *per* test

$\alpha_{EW}=1-(1-\alpha)^k$

Cross-sectional Approaches: Incomplete Data

- Assume an endpoint was collected on each subject
  - Or an imputation procedure was specified for dropouts
- **Last Observation Carried Forward (LOCF) assumes**
  - Dropout unrelated to outcome
  - No change after dropout
- LOCF bias can go in either direction
  - Can favor either comparator or investigational agent

(Mallinckrodt, 2001, 2003)

Cross-sectional Approaches: Incomplete Data

- Decrease generalizability
- Decrease statistical power
- Decrease precision (i.e., widens 95% CI)
- Introduce attrition bias

Longitudinal Data

ID	0	1	2	3	4
1001	29	27	22	18	17
1002	22	22	18	11	11
1003	21	22	19	18	16
1004	26	22	22	22	19
1005	24	18	17	16	15
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...
1100	25	22	21	19	18

- **Repeated Measures ANOVA**

  Assumptions
  - Complete data
  - Compound symmetry (equal variances & covariances)

- **Multivariate Analysis of Variance** (MANOVA)
  - Assumption: complete data

Complete Case Analyses

Assumptions

- Completers are representative of:
  - Randomized subjects
  - Patient population to which inferences are to be made

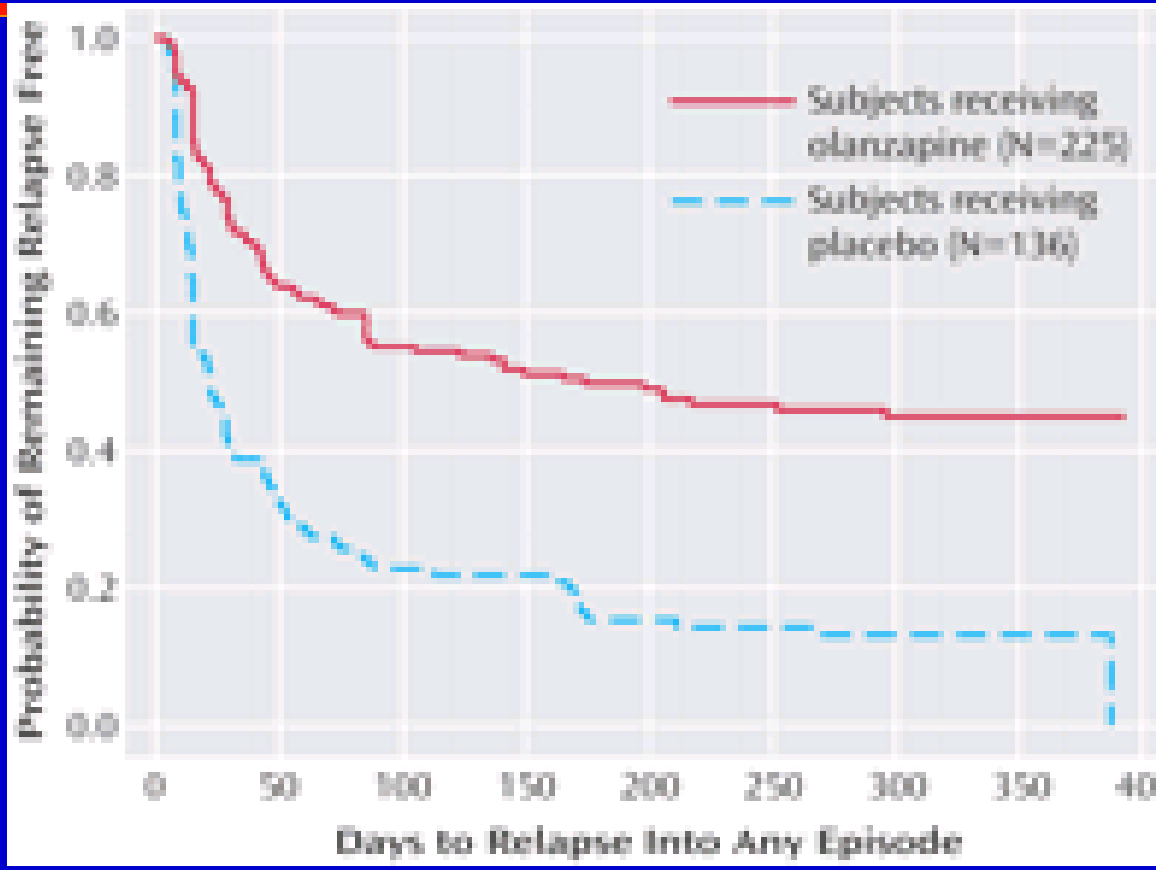
- Data are Missing Completely at Random (MCAR):
  - Missingness does not depend on the dependent variable (*observed* or *unobserved* measures)
    - Difficult to verify

Analysis of Incomplete Long-term RCTs Data

- Reduce attrition bias with data analytic procedure that does not completely exclude subjects with some missing data

Survival Analysis:
Time to Relapse

- K-M examines <u>cumulative</u> response or recurrence during trial (Kaplan & Meier, 1959)
- Logrank test compares survival across groups (Peto & Peto, 1972).
- No need to impute data for dropouts



(logrank p<0.001)

Tohen et al. RCT of Olanzapine as Maintenance Therapy in Patients With Bipolar I Disorder Responding to Acute Treatment With Olanzapine. AJP, 2006.

Survival Analysis

Time to relapse, response, or med. discontinuation.

Assumptions

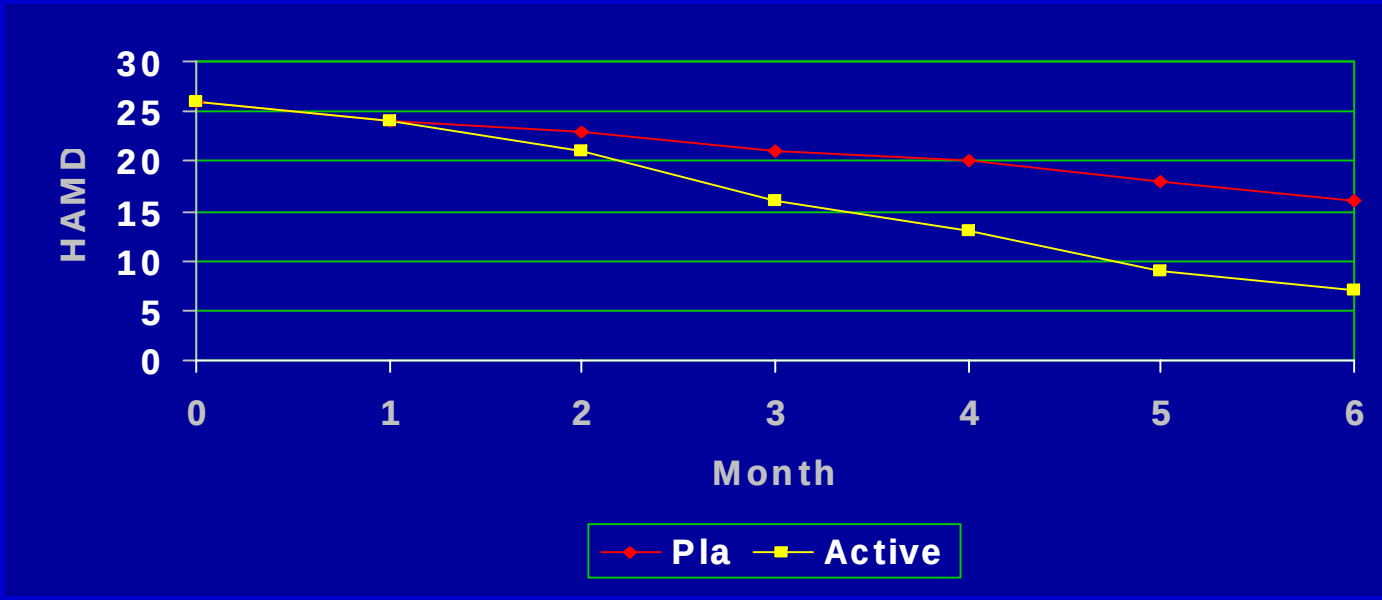
- Once classified a *responder*, a subject will not revert to *non-responder* status (during trial) – unlike a terminal event

* Dropout is not related to outcome

Mixed-Effects Linear Regression:
Severity over Course of RCT

Continuous dependent variable

- **HAMD**
- **YMRS**
- **PANSS**



Time 0 is point of randomization in

Acute, Maintenance, Randomized withdrawal designs

Mixed-Effects Models
(Laird & Ware, 1982)

- Account for within-subject change over time

- Unit of analysis: monthly assessments
- Include varying number of observations per subject
- Use available data. Do not exclude subjects with some missing data.

Mixed-Effects Models

- Applied in observational studies and RCTs
- Used too seldom in psychopharmacology trials. When used, often as secondary. (Laird and Ware was 1982.)
- Only 2 of 13 recent long-term RCTs for bipolar disorder used mixed-effects models for the primary analysis

- Recommend mixed-effects models for the primary analysis in regulatory submission

TABLE 2. Treatment Response of Subjects With First-Episode Psychosis in the 12-Week Acute Treatment Phase of a Long-Term Comparison of Olanzapine and Haloperidol

	Observed Case Means				Least Squares Means				Difference Between Treatments (p)	
	Baseline		Endpoint		Baseline		Endpoint			
Measure and Treatment Group	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mixed-Model Analysis[a]	Last-Observation-Carried-Forward Analysis[b]
Positive and Negative Syndrome Scale										
Total score									<0.02	0.58
Olanzapine (N=126)	75.90	18.07	55.85	16.52	75.56	14.90	59.33	19.41		
Haloperidol (N=125)	73.57	17.50	59.35	16.63	76.07	14.90	65.40	19.42		
Negative scale score									<0.04	0.89
Olanzapine (N=126)	19.02	6.51	16.07	6.00	18.91	5.44	16.64	5.89		
Haloperidol (N=126)	18.77	6.61	17.56	5.95	19.05	5.44	18.29	5.87		
Positive scale score									0.50	0.76
Olanzapine (N=126)	19.53	5.78	12.12	4.14	19.63	4.37	13.39	5.59		
Haloperidol (N=126)	19.37	5.33	12.31	4.50	19.67	4.37	13.90	5.59		
General scale score									<0.003	0.25
Olanzapine (N=132)	37.51	9.42	27.66	8.09	37.07	7.78	29.14	9.50		
Haloperidol (N=130)	35.72	8.91	29.48	8.34	37.40	7.78	33.04	9.51		
CGI severity score									0.07	0.46
Olanzapine (N=127)	4.35	0.87	3.01	1.09	4.25	0.61	3.24	1.18		
Haloperidol (N=127)	4.20	0.77	3.18	1.00	4.27	0.61	3.54	1.18		
Montgomery-Åsberg Depression Rating Scale score									<0.02	0.07
Olanzapine (N=125)	9.53	7.26	6.95	7.01	10.12	5.84	8.49	8.68		
Haloperidol (N=126)	10.21	6.65	8.38	8.21	10.35	5.84	11.27	8.68		

[a] F statistic testing equality of linear slope coefficients from a mixed linear random regression coefficients model with a common intercept, a linear slope for each treatment, and a common quadratic slope.

[b] Type III F statistic testing equality of last-observation-carried-forward change scores in an analysis of variance model including terms for investigator, treatment, and investigator-by-treatment interaction.

Lieberman et al., RCT of Olanzapine vs. Haloperidol, AJP 2003, 160:1396-1404

RCT of Olanzapine vs. Haloperidol

Lieberman et al., AJP 2003

" ... possible explanations for the differences in results between the LOCF and random regression analysis.

LOCF analyses use only baseline and endpoint values, while random regression analyses use all data points.

LOCF analyses assume that the patients who drop out would continue to maintain their last score and that dropouts occur randomly and are not influenced by the treatment assignment or the level of symptoms. Neither assumption is necessarily justified."

Linear Regression Model

$$y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 G_i + \varepsilon_{ij}$$

y_{ij} = dependent variable for subject *i* at time *j*

Fixed effects

T_{ij} = Time

G_i = treatment Group (0=Comparator; 1=Investigational)

* Assumes errors ε_{ij} are uncorrelated.
  Not plausible with repeated measures within-subject.

Hedeker and Gibbons. *Longitudinal Data Analysis*. Wiley, 2006

Mixed-Effects Linear Regression

$$y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 G_i + \nu_{0i} + \varepsilon_{ij}$$

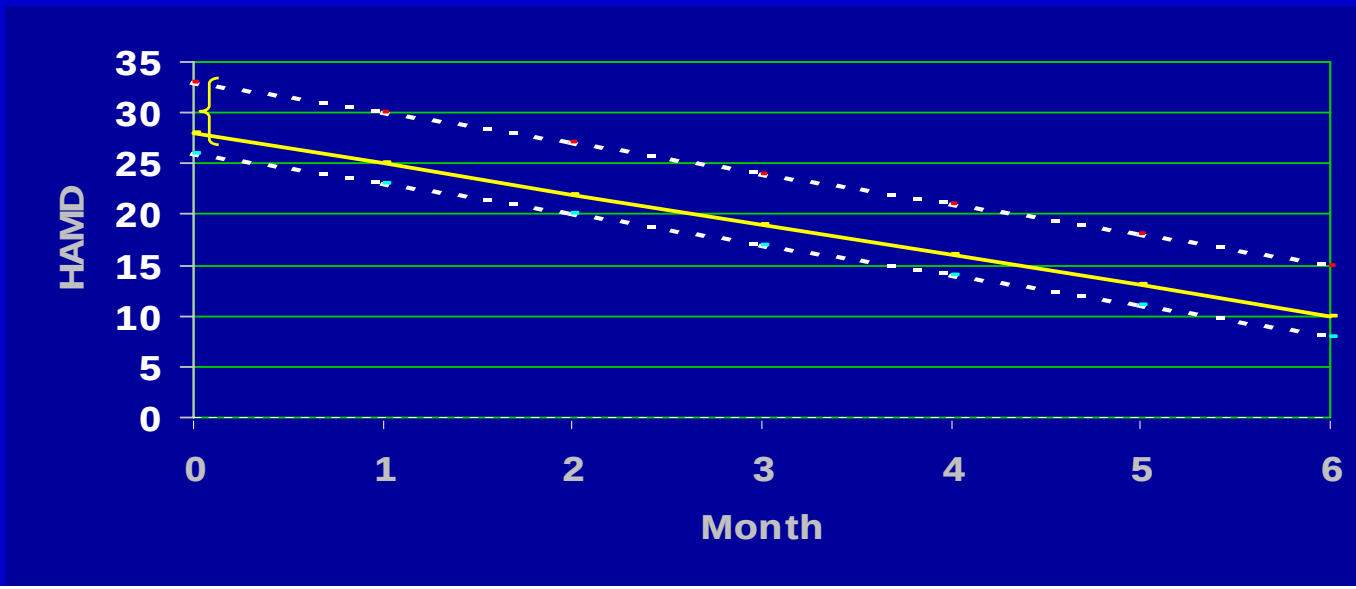
Random effect

ν_{01} = subject-specific **random intercept**

Accounts for each subject's deviation from group intercept

Subject's baseline value Influences subsequent values

Allows for correlated data. ε_{ij} are now conditionally independent



Mixed-Effects Linear Regression

$$y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 G_i + \nu_{0i} + \varepsilon_{ij}$$

y_{ij} = dependent variable for subject *i* at time *j*

Fixed effects

T_{ij} = Time

G = treatment Group

β_0 = intercept term (baseline severity for Group 0)

β_1 = coefficient for time (slope for Group 0)

β_2 = coefficient for treatment Group (group difference at time 0)

Mixed-Effects Linear Regression

$$y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 G_i + \nu_{0i} + \nu_{1i} + \varepsilon_{ij}$$

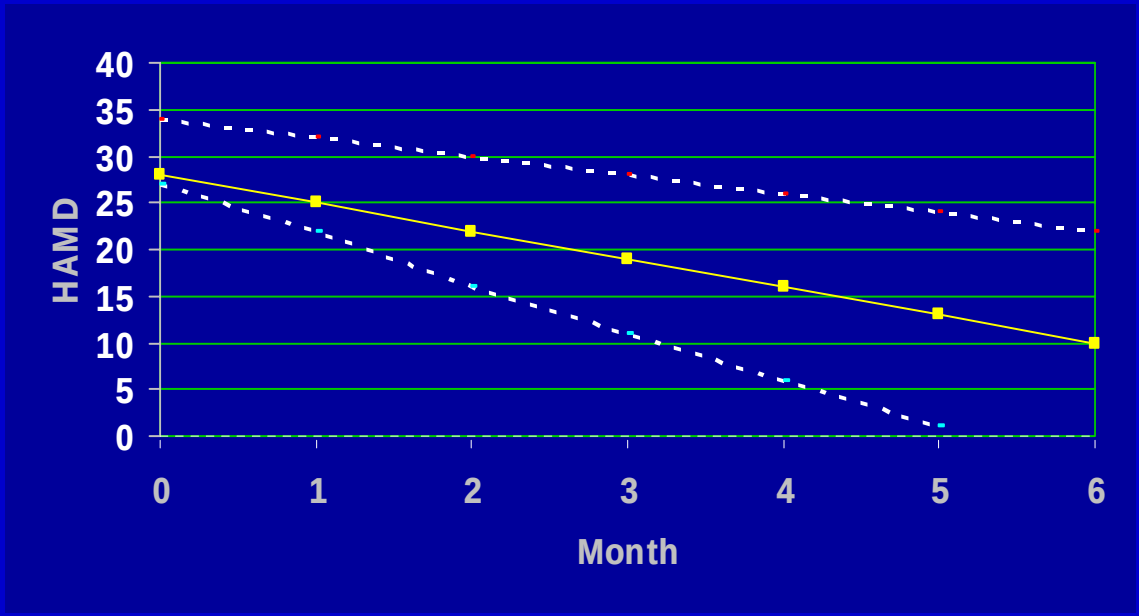
Random effects

ν_{01} = random intercept

ν_{i1} = subject-specific **random slope**

Accounts for each subject's deviation from group slope.

S's post-baseline value is influenced by S's baseline level and S's rate of change.



Quantify Group Differences in Rate of Change over Time

$$y_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 G_i + \beta_3 G_i T_{ij} + \nu_{0i} + \nu_{1i} + \varepsilon_{ij}$$

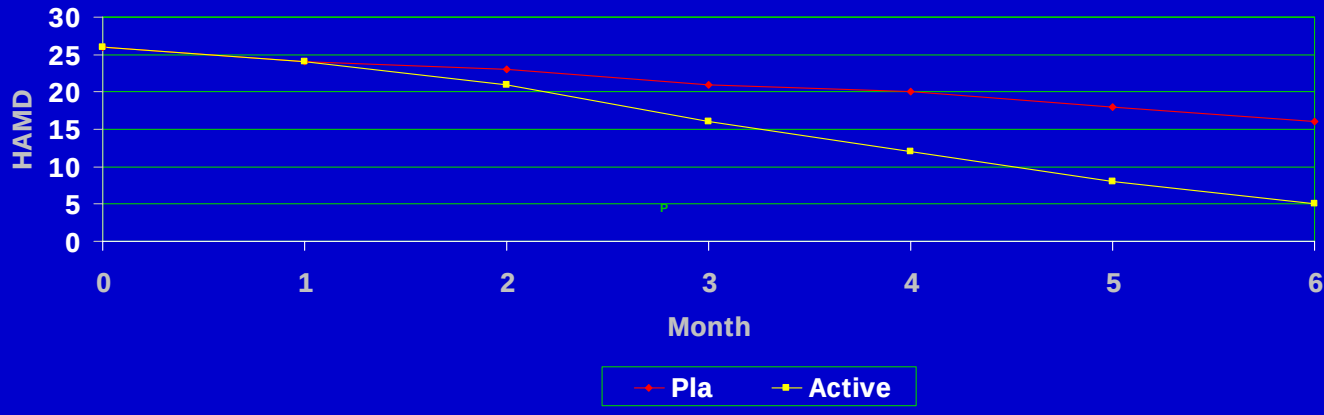
y_{ij} = dependent variable for subject *i* at time *j*

<u>Fixed effects</u>

T_{ij} = Time

G_i = treatment Group

$G_i T_{ij}$ = **Group*Time**



β_3 = coefficient for **Group*Time**

Mixed-Effect Models

- Mixed-Effect Logistic Regression
  - Binary outcome: responder vs non-responder (monthly)
- Mixed-Effect Ordinal Logistic Regression (Hedeker & Gibbons, 1994)
  - Ordinal categories: Full vs partial vs non-responder (monthly) - REVAMP
- Mixed-Effects Nominal Regression
  - Euthymic vs. Depressed vs. Mania vs. Mixed/Cycling
- Mixed-Effect Grouped-time Survival Analysis (Hedeker et al., 2000)
  - Repeated intervals of time to event (observational data -- CDS)

Mixed-Effects Models

- Incorporating multiple assessment times:
  - More power
    (holding N and ICC constant)
  - Reduced sample size requirement
    (holding treatment effect and ICC constant)

Mixed-Effects Linear Regression:
Sample Size Requirements (per group)

effect size	# of assessment times			
	1 (t-test)	2	4	6
0.30	**176**	**123**	**96**	**88**
0.40	**100**	**69**	**54**	**50**
0.50	**64**	**44**	**35**	**32**
0.60	**45**	**31**	**24**	**22**
median reduction	**-**	**31%**	**46%**	**50%**

Assume: ICC=.40; 2-tailed α=.05, power=.80

Mixed-Effects Models

- **Assumption: <u>Ignorable dropout</u>**
  - Dropout is explained by covariates and/or measures of outcome prior to dropout

- With ignorable dropout, mixed-effects models can be used for valid inference.

Non-ignorable Dropout

- Dropout depends on unobserved outcome data
  - Dropout due to unobserved symptom severity, adverse events, or death
- Predictors of dropout are not known or unavailable

Strategies for Non-ignorable Dropout

- **Pattern-mixture Model**
- **Self-Assessed *Intent to Dropout***

Pattern-mixture model

(Little, *JASA*, 1993)

Stratify analyses by:

- **Pattern of missing data**
- **Phase of Dropout**
  - **Early, middle, completer**
- **Reason for Dropout**
  - **(Lack of efficacy, Adverse events,…)**

Pattern		Month			
		1	2	3	4
1		Obs	Obs	Obs	Obs
2		Obs	Obs	Obs	Missing
3		Obs	Obs	Missing	Obs
…					

Pool stratified results

Mixed-Effects Application of Pattern Mixture

- **6 week RCT for schizophrenia**
- **Mixed-effects pattern mixture model**
- **Covariate: *Dropout* vs. *Completer***

Hedeker D, Gibbons RD. *Psych Methods*, 1997

Predicting Dropout:
Intent to Attend

“How likely are you to attend the next assessment session?”

unlikely (0) **unsure (5)** **very likely (10)**

These data could change non-ignorable dropout to ignorable.

Ignorable dropout: mixed-effects models provide valid inference.

Demirtas H, Schafer JL. Statistics in Medicine, 2003

Improved Standards for Reporting Long Term RCT Results

Monthly (weekly) group-specific:

- **N's and retention rates over time**
- **Means (sd) & 95% CI: cross-sectional and change**
- **Response rates**

Effect sizes

ICC for each outcome: needed for power analyses

Summary

- Long term RCTs will have subjects with incomplete data.
- Use data analytic approach that does not require complete data.
- Mixed-effects models include subjects with incomplete data and can account for dropout.
  - Reduce bias
  - Increase generalizability, power and precision

Recommendations

- Design RCTs to minimize attrition

- Do not use cross-sectional approaches to longitudinal data.
- Do not use LOCF in RCTs.

- Designate mixed-effect models as primary data analytic approach.
- Collect data that predict dropout.