Accelerating Impact: Immersive Summer Bootcamp in Implementation Science and Biostatistics

Georgian Implementation Science Fogarty Training (GIFT) Program

Ilia State University & Yale University

Yale Center for Analytical Sciences

Repeated Measures

Behavior Therapy for Children With Tourette Disorder
A Randomized Controlled Trial

Table 2. Baseline, Week 5, and Week 10 Scores on Key Outcome Measures[a]

	Mean (95% CI) Behavioral Intervention (n = 61)	Mean (95% CI) Control (n = 65)	Group Difference at Week 10 (95% CI)	Effect Size[b]
Yale Global Tic Severity Scale				
Total tic score				
Baseline	24.7 (23.1-26.3)	24.6 (23.2-26.0)		
Week 5	19.7 (17.6-21.7)	22.8 (20.7-24.9)	3.3 (1.2-5.4)	0.54
Week 10	17.1 (15.1-19.1)	21.1 (19.2-23.0)	4.1 (2.0-6.2)	0.68
Total motor				
Baseline	14.6 (13.5-15.7)	14.6 (13.8-15.4)		
Week 5	12.2 (10.8-13.6)	13.6 (12.4-14.7)	1.3 (0.1-2.7)	0.34
Week 10	10.7 (9.3-12.1)	12.5 (11.5-13.5)	1.9 (0.4-3.3)	0.49
Total vocal				
Baseline	10.1 (9.0-11.2)	10.0 (8.9-11.1)		
Week 5	7.4 (6.2-8.6)	9.3 (7.9-10.6)	2.0 (0.7-3.3)	0.43
Week 10	6.5 (5.4-7.6)	8.6 (7.4-9.8)	2.2 (0.9-3.6)	0.50
Impairment				
Baseline	25.0 (22.6-27.4)	23.4 (21.6-25.2)		
Week 5	16.8 (14.0-19.5)	20.1 (17.6-22.7)	3.8 (0.7-6.8)	0.47
Week 10	12.2 (9.8-14.6)	16.4 (13.8-19.0)	4.7 (1.6-7.8)	0.57
Parent Tic Questionnaire total score				
Baseline	34.2 (29.5-38.9)	35.7 (30.3-41.1)		
Week 5	25.8 (21.5-30.1)	33.7 (27.8-39.6)	7.3 (1.5-13.1)	0.28
Week 10	20.0 (16.3-23.7)	27.6 (23.0-32.2)	7.8 (1.9-13.8)	0.30
Children's Global Assessment Scale[c]				
Baseline	59.0 (57.1-60.9)	59.3 (57.3-61.3)		
Week 10	69.4 (66.9-71.9)	64.1 (59.5-68.7)	5.8 (2.8-8.8)	0.64

Abbreviation: CI, confidence interval.
[a] Data are presented as least square mean values and 95% CIs and standard deviations for each assessment point. Group differences with 95% CIs are also presented.
[b] Effect sizes were calculated as follows: change from baseline in behavior therapy minus change in the control group divided by the pooled standard deviation for the entire study sample at baseline.
[c] Children's Global Assessment Scale administered only at baseline and week 10.

Repeated Measures

A. Longitudinal – assess an outcome at different timepoints in a study with follow-up
   - Example: Measure tic severity in RCT at baseline 5 and 10 weeks

B. Testing multiple conditions on same experimental unit
   - Example: Assess craving to several different foods

C. Assessing response to a stimulus
   - Example: OGTT - Measure glucose every 30 minutes following a consumption of 75 g glucose drink

Assumption of Independence

- The value of one observed outcome does not affect the value of another observed outcome
  - Not met in repeated measures studies
    - Example: In a weight loss study with multiple weight assessments per individual, those with higher baseline weights tend to have higher follow-up weights

Weight Loss Example

Subject	No Diet		Diet		Difference
1	168	(4)	163	(4)	-5
2	140	(2)	134	(2)	-6
3	175	(5)	169	(5)	-6
4	194	(6)	189	(6)	-5
5	203	(8)	200	(7.5)	-3
6	201	(7)	200	(7.5)	-1
7	165	(3)	160	(3)	-5
8	134	(1)	130	(1)	-4
$\bar{x}$	172.5		168.13		-4.38
s	26.3		27.2		1.69

r=0.999

*Pooled SD from Independent Sample T-test = 26.7

Weight Loss Example

$$Ho: \mu_{before} = \mu_{after} \text{ or } \mu_{before} - \mu_{after} = 0$$

$$Ha: \mu_{before} \neq \mu_{after} \text{ or } \mu_{before} - \mu_{after} \neq 0$$

- This problem becomes a one-sample t-test

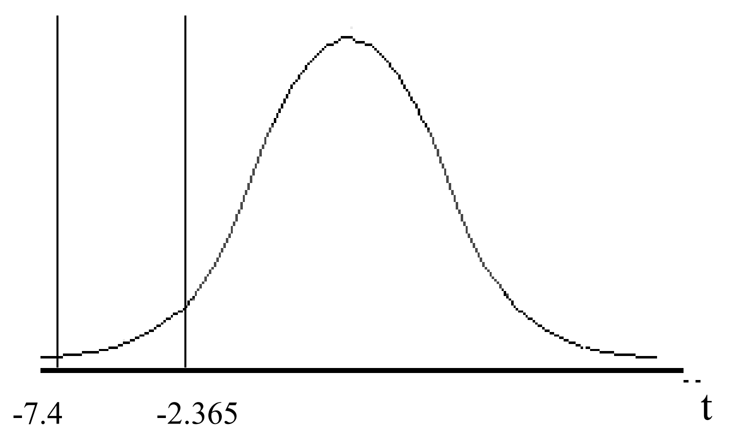
$$t = \frac{\bar{x}_{diff} - 0}{s / \sqrt{n}} = \frac{-4.38 - 0}{1.69 / \sqrt{8}} = -7.4$$

*Denominator from Independent Sample T-test = 13.4
t-statistic from Independent T-test = 0.33 (14 df)

Weight Loss Example

- Compare t to t_{crit}

$$t_{0.05/2,7} = -2.365$$ Excel: tinv(.05,7)



  – -7.4<-2.365 so Reject the Null

Assumptions for Paired t-test

- The population of differences from which the sample arose are distributed approximately normal
- The difference observations are independent

Standard Deviation for the Difference in Two Measures

$$\sigma^2{}_{diff} = \sigma^2{}_{pre} + \sigma^2{}_{post} - 2\rho\sigma_{pre}\sigma_{post}$$

$$\sigma^2{}_{diff} = 2\sigma^2(1 - \rho)$$

$s^2{}_{diff}$=26.3^2 + 27.2^2 – 2*0.999*26.3*27.2 = 2.84

s_{diff}=$\sqrt{2.84} = 1.69$

Advantage of the Paired Design
(One Group Pre-test/Post-test Design)

- Often permits the easier detection of change by omitting extraneous between subject variation
  - Numerator of the paired t-statistic is the same as the two-sample t-statistic (same mean difference)
  - The difference is in the denominator which is reduced when pre and post measurements are correlated

Disadvantages of the Pre-test/ Post-test Design

- Individuals may change without treatment
  - History
  - Maturation
  - Testing

- Change occurs because of measurement error
  - Regression to the mean
  - Instrumentation

Regression to the Mean

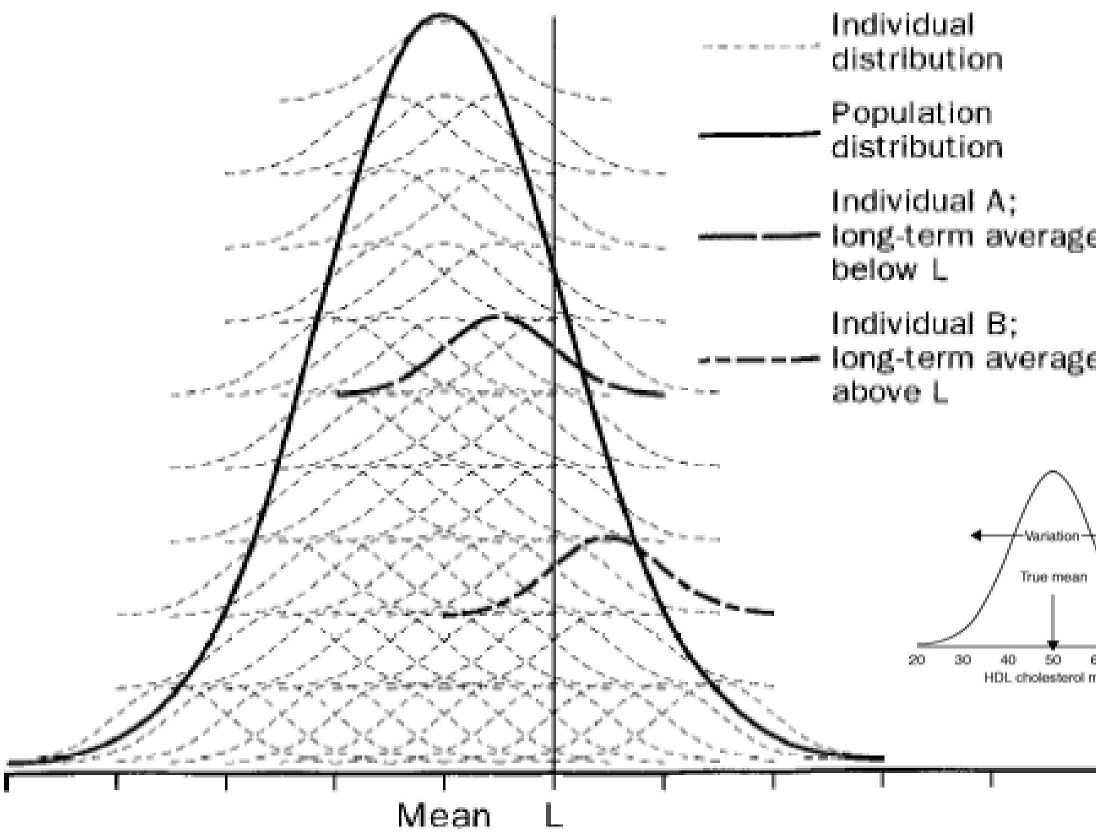
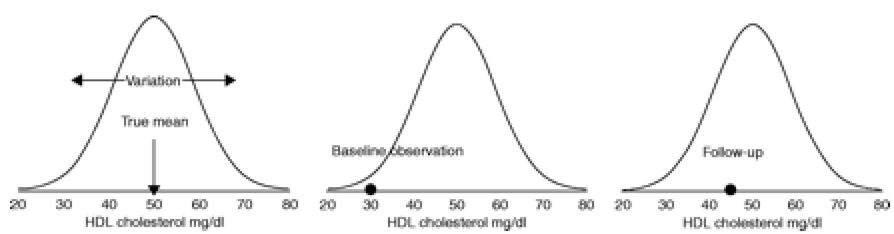


Figure: **Schematic distributions of serum cholesterol in populations and in individuals**



Lancet 1996; **347:** 241–43

Extending the Paired Design

- Paired design can compare two dependent samples. How can we draw inference when greater than 2?

Food Intake Example

Animal studies have demonstrated that compression and distension of the stomach trigger nerves that signal the brain to turn off the desire to eat. To test this investigators limited expansion of the stomach by placing a large inflatable cuff around the abdomen of 7 experimental subjects and inflating it to 3 incremental pressures. At each specified pressure, the subjects food intake was measured.

Subject	Food Intake (ml) at abdominal pressure		
	0 mm Hg	10 mm Hg	20 mm Hg
1	448	470	292
2	472	424	390
3	631	538	508
4	634	496	560
5	734	547	602
6	820	578	508
7	643	711	724

Randomized Block Design
(Repeated Measures ANOVA)

- Extension of the paired t-test

	Treatments			
Blocks	1	2	. . .	p
1	X_{11}	X_{12}	. . .	X_{1p}
2	X_{21}	X_{22}	. . .	X_{2p}
3	X_{31}	X_{32}	. . .	X_{3p}
.	.	.	. . .	
.	.	.	. . .	
b	X_{b1}	X_{b2}	. . .	X_{3b}

Randomized Block Design

- TSS
  Total Sum of Squares
  - SST
    Sum of Squares Treatment
  - SSE
    Sum of Squares Error
    - SSB
      Between Blocks
    - SSE
      Within Blocks

$$TSS = SST + SSB + SSE$$

Randomized Block Design

- Sources of Variation

$$TSS = \sum_{i=1}^{p} \sum_{j=1}^{b} (x_{ij} - \bar{\bar{x}})^2$$

Where

$\bar{x}_j$ = the mean for the j[th] block

$\bar{x}_i$ = the mean for the i[th] treatment

$$SSB = \sum_{j=1}^{b} p(\bar{x}_j - \bar{\bar{x}})^2$$

$$SST = \sum_{i=1}^{p} b(\bar{x}_i - \bar{\bar{x}})^2$$

ANOVA Table for Randomized Block Design

Source	d.f.	SS	MS	F
Treatment	k-1	SST	MST= $SST/(k-1)$	MST/MSE
Blocks	b-1	SSB	MSB= $SSB/(b-1)$	MSB/MSE
Error	(k-1)(b-1)	SSE	MSE= $SSE/((k-1)(b-1))$	

Food Intake Example

Tests of Between-Subjects Effects

Dependent Variable: INTAKE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	258988.857[a]	8	32373.607	5.904	.003
Intercept	6552042.857	1	6552042.857	1194.896	.000
SUBJECT	208935.143	6	34822.524	6.351	.003
HG	50053.714	2	25026.857	4.564	.034
Error	65800.286	12	5483.357		
Total	6876832.000	21			
Corrected Total	324789.143	20			

a. R Squared = .797 (Adjusted R Squared = .662)

INTAKE

Student-Newman-Keuls[a,b]

HG	N	Subset 1	Subset 2
20.00	7	512.0000	
10.00	7	537.7143	
.00	7		626.0000
Sig.		.528	1.000

Means for groups in homogeneous subsets are displayed.
Based on Type III Sum of Squares
The error term is Mean Square(Error) = 5483.357.

a. Uses Harmonic Mean Sample Size = 7.000.

b. Alpha = .05.

Assumptions for Randomized Block Design

- Normality
- Independence of Subjects
- Sphericity – similar to homogeneity of variance assumption

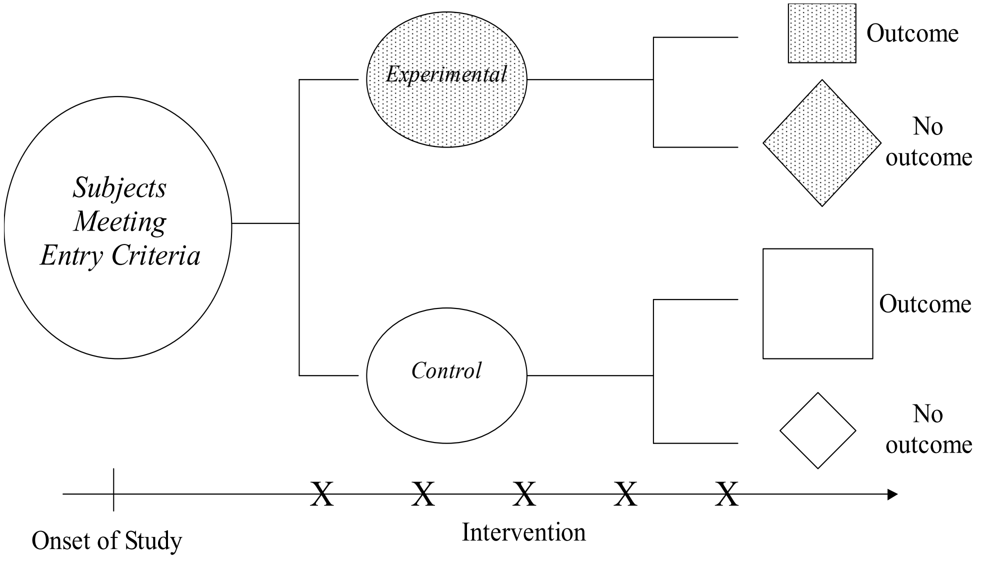
$$s^2_{0-10} = s^2_{0-20} = s^2_{10-20}$$

- <u>Crossover factor</u> – all subjects are measured at each level

- <u>Nested factor</u> – each subject is assessed at one factor level

Two Group (or more) Repeated Measures

Subject	Time 1	Time 2	Time 3
Trt A			
1	X	X	X
2	X	X	X
3	X	X	X
.	.	.	.
.	.	.	.
b_A	X	X	X
Trt B			
1	X	X	X
2	X	X	X
3	X	X	X
.	.	.	.
.	.	.	.
b_B	X	X	X



Possible Analytic Strategies

- Endpoint Analysis – use only the last (or most important timepoint) – compare using a ttest, ANCOVA (adjusting for baseline) or multiple regression

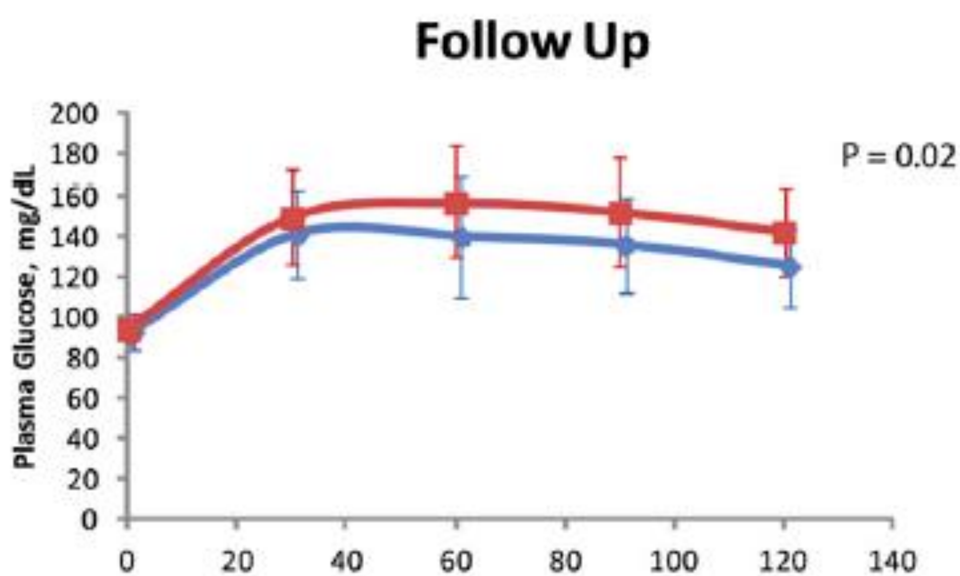
- Summary statistics – estimate a summary for each subject (e.g. mean, slope) compare using simple methods

- Repeated Measures Mixed Model – uses all available data

- Repeated Measures ANOVA – uses all subjects that have a complete set of data

Summary Statistics

- Can generate summary outcomes from repeated measures and compare using simple inferential procedures
  - change, mean, peak, slope, area under the curve



Repeated Measures Analysis

Data collected from the same participants over multiple time points.

Mixed-Effects Models

- Linear Mixed-Effects Models: Analyzes data with both fixed and random effects.
- Generalized Mixed-Effects Models: Extends mixed-effects models to non-normal outcomes.

Fixed effect: e.g. treatment assignment

Random effect: e.g. baseline values

Example

Assessing the effect of a treatment (A or B) on blood pressure over time

Measurements are taken at three different time points:

- Baseline (T1)
- Week 4 (T2)
- Week 8 (T3)

How does time and treatment influence blood pressure?

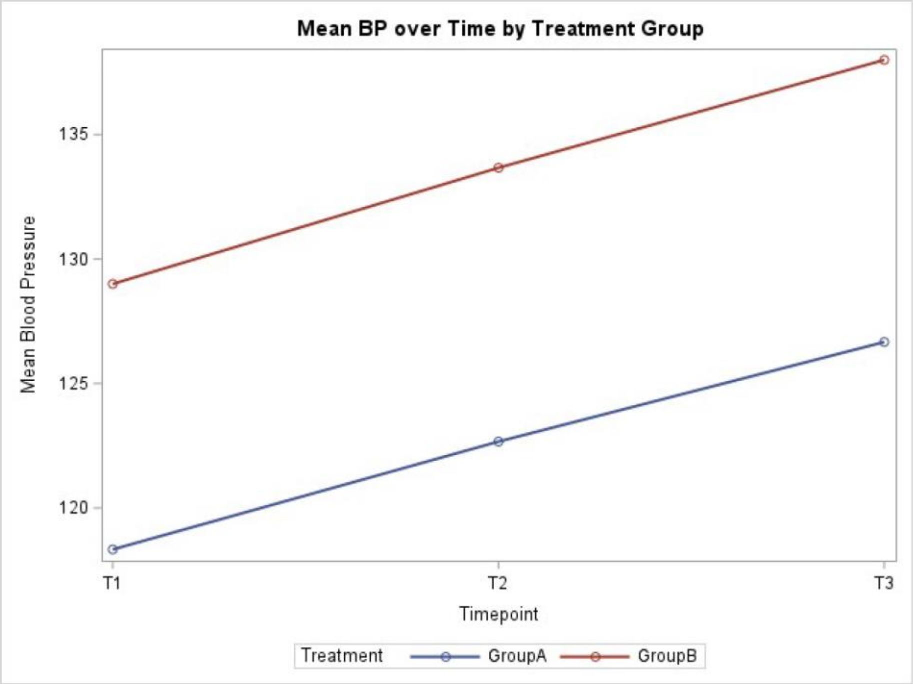
The SAS System

The Mixed Procedure

Model Information	
Data Set	WORK.BP_LONG
Dependent Variable	BP
Covariance Structure	Compound Symmetry
Subject Effect	ID
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Kenward-Roger
Degrees of Freedom Method	Kenward-Roger

Class Level Information

Class	Levels	Values
ID	6	1 2 3 4 5 6
Treatment	2	GroupA GroupB
Timepoint	3	T1 T2 T3



Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Treatment	1	4	57.65	0.0016
Timepoint	2	8	312.15	<.0001
Treatment*Timepoint	2	8	0.46	0.6461

Two Group Repeated Measures ANOVA Example

- Trial comparing multiple daily insulin injections (MDI) to insulin pump infusion (pump) in children with type I diabetes
  - Two factors
    - Group – 2 levels – MDI vs Pump
    - Time – 5 levels – 0, 4, 8, 12, 16 weeks
  - Response Variable
    - HbA1c

Two Group Repeated Measures ANOVA Example

Tests of Within-Subjects Effects

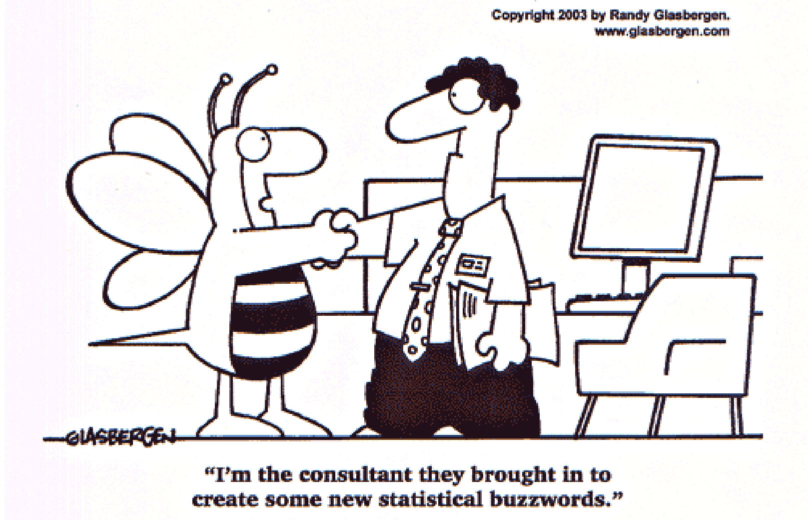
Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TIME	Sphericity Assumed	8.565	4	2.141	7.905	.000
	Greenhouse-Geisser	8.565	2.067	4.143	7.905	.001
	Huynh-Feldt	8.565	2.389	3.585	7.905	.001
	Lower-bound	8.565	1.000	8.565	7.905	.010
TIME * GROUP	Sphericity Assumed	2.002	4	.500	1.847	.127
	Greenhouse-Geisser	2.002	2.067	.968	1.847	.168
	Huynh-Feldt	2.002	2.389	.838	1.847	.161
	Lower-bound	2.002	1.000	2.002	1.847	.188
Error(TIME)	Sphericity Assumed	23.839	88	.271		
	Greenhouse-Geisser	23.839	45.482	.524		
	Huynh-Feldt	23.839	52.556	.454		
	Lower-bound	23.839	22.000	1.084		

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	6966.966	1	6966.966	1973.625	.000
GROUP	10.355	1	10.355	2.933	.101
Error	77.661	22	3.530		



Assumptions

- Normality
- Independence of subjects
- Multisample sphericity – not a problem if group sizes are equal (i.e. only check sphericity)

Mauchly's Test of Sphericity[b]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
TIME	.187	34.266	9	.000	.517	.597	.250

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in Tests of Within-Subjects Effects table.

b.

Design: Intercept+GROUP
Within Subjects Design: TIME

MAYBE AN R EXAMPLE HERE REPEATED MEASURES

Mixed model

Missing Data

Substantial instances of missing data are a serious problem that undermines the scientific credibility of causal conclusions from clinical trials. The assumption that analysis methods can compensate for such missing data are not justified, so aspects of trial design that limit the likelihood of missing data should be an important objective.

A major source of missing data in clinical trials is participants who discontinue the assigned treatment because of adverse events, lack of tolerability, lack of efficacy, or simple inconvenience. Too many investigators incorrectly equate treatment discontinuation with study dropout; that is, outcomes are not recorded for participants who discontinue treatment. However, enrollees committed to participating in the study, not just to receiving the assigned treatment. When a study treatment is discontinued, efforts should be made to obtain the participant's consent for the collection of data on treatments and outcomes. When such efforts are successful, gathering these data after treatment discontinuation preserves the ability to analyze end points for all participants who underwent randomization and thus to make possible intention-to-treat inferences, which are grounded in randomization. It also

limit the extent of missing data. Finally, in studies with missing data, analysis methods that are based on plausible scientific assumptions should be used. For example, this consideration often rules out simple fixes, such as imputation by the last observation carried forward.[10] Although there

Table 5. Approaches to handling and preventing missing data during trial design, planning, conduct and analysis.

	Design Stage
Anticipate Expected Missing Data	1. Estimate the expected amount of missing data and likely reasons for it. 2. Account for missing data in the sample size calculations and develop a suitable pre-specified analytic plan.
Methods to Encourage Participant Retention	3. Limit burden to participant by reducing required visits and amount of data collected. 4. Adopt data collection methods that don't require face to face visits. 5. Utilize run-in periods, ascertainable treatment outcomes, shorter follow-up periods, randomized withdrawal designs where appropriate. 6. Budget for monetary incentives for participants that are weighted toward study completion.
	Planning Stage
Study Documentation	7. Develop detailed study documentation in the form of manual of operations addressing all aspects of the study including screening procedures, training requirements, methods of communication, delivery of treatment, schedule and windows for assessments, and data collection/entry/editing procedures.
Informed Consent	8. Develop an informed consent that distinguishes the difference between withdrawing from the treatment and withdrawing from the study.
Study Sites	9. Select study sites with strong track records for enrolling, following, and completing participants. 10. Adopt a reimbursement mechanism that encourages study completion.
Training Study Personnel	11. Train/certify study personnel for participant enrollment, data collection, data entry, delivery of treatment, etc. prior to enrollment with re-certification throughout trial if necessary. 12. Highlight the continued collection of data in participants that are not adherent to treatment but remain in the study.
Pilot Study	13. Test operational aspects of the trial (e.g., enrollment, retention, clarity of study manuals and data collection instruments, study burden on participants, randomization, treatment delivery).
	Conduct Stage
Create Monitoring Reports	14. Develop monitoring reports to regularly track amounts of missing data at the levels of the study site and study personnel. 15. Keep track of reasons for withdrawal from the study or intervention.
Enhance Participant Contact	16. Utilize approaches to keep the study participants engaged in the study including incentives, visit reminders, newsletters, and intermittent phone calls to monitor status. 17. Outline procedures for contacting individuals with missed visits in manual of operations. Identify and intervene in participants that are likely to drop out.
Data Entry and Management	18. Timely data entry allows earlier detection of problems with missing data. 19. Implement a verification process requiring fields to be checked for accuracy and all discrepancies resolved before data entry.
Communication	20. Devise an efficient method of communication with study personnel for identifying and resolving unanticipated issues that arise during the study.
	Analytic Stage
Explore Missing Data	21. The amount of missing data, missing data patterns and variables associated with missingness will help to inform the primary and sensitivity analyses.
Use All Available Data	22. For primary analysis, use methods that make use of all available data such as multiple imputation or likelihood-based approaches. These methods make weaker assumptions about the missing data compared to complete case analysis. 23. For primary analysis, avoid the use of ad-hoc solutions (e.g., last observation carried forward) as they make unreasonable assumptions about the mechanism that produced the missing data.
Perform sensitivity analysis	24. Use methods such as pattern mixture or selection models to examine robustness of conclusions to reasonable MNAR mechanisms.

Table 1. Common examples of the three missing data mechanisms in clinical trials.

Missing Mechanism	Examples
MCAR	Administrative censoring: follow-up is terminated because the study has ended. Migration-study participants move and are unable to complete visits. Random failure of the experimental instrument (e.g. test tube break, equipment failure)
MAR	Missing data caused by features of the study design such as participants being removed from the trial if their conditions are not controlled sufficiently well according to protocol criteria. Dropout based on recorded side-effects. Dropout based on known baseline characteristics.
MNAR	Dropout based on the unobserved response (e.g., a person not responding to treatment is more likely not to provide an observation). Participants miss a visit because they've had an outcome.

YALE JOURNAL OF BIOLOGY AND MEDICINE 86 (2013), pp.343-358.

3 Approaches to Dealing with Missing Data

- Complete Case Analysis – only subjects with all assessments – assumes MCAR

- Imputation
  - Crude – LOCF – assumes no change after dropout
  - Multiple Imputation – assumes MAR

- Analysis of Incomplete Data
  - Mixed model – assumes MAR
  - Pattern Mixture Model – for MNAR