

BMJ Open

To condition or not condition? Analyzing “change” in longitudinal randomized controlled trials

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2016-013096
Article Type:	Research
Date Submitted by the Author:	28-Jun-2016
Complete List of Authors:	Coffman, Cynthia; Duke University Medical Center, Biostatistics and Bioinformatics; Durham VA Medical Center, Health Services Research and Development Edelman, David; Duke University Medical Center, General Internal Medicine; Durham VA Medical Center, Health Services Research and Development Woolson, Robert; Duke University Medical Center, Biostatistics and Bioinformatics
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Health services research
Keywords:	Constrained longitudinal data analysis, ANCOVA, Longitudinal RCT

SCHOLARONE™
Manuscripts

1
2
3 **To condition or not condition? Analyzing “change” in longitudinal randomized**
4 **controlled trials**
5
6
7
8
9

10 Cynthia J. Coffman PhD^{1,2}, Edelman, David MD^{1,3} and Robert F. Woolson, PhD^{1,2}
11
12

13
14
15 ¹Health Services Research, Durham Veterans Affairs Medical Center,
16

17 ²Department of Biostatistics and Bioinformatics, Duke University Medical Center
18

19 ³Department of Medicine, Duke University Medical Center
20
21
22

23
24 Corresponding author: Cynthia J. Coffman, PhD,
25

26 Durham VA Medical Center, HSR&D(152)
27

28 508 Fulton St.
29

30 Durham, NC 27705,
31

32
33 Tel: 919-286-6936 Fax: 919-416-5836
34

35 Email: Cynthia.Coffman@duke.edu
36
37
38
39

40 **Running Title:** Analyzing “change” in longitudinal randomized controlled trials
41
42
43

44 Word Count: 3328
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ABSTRACT

Objective: The statistical analysis for a two-arm randomized controlled trial (RCT) with a baseline outcome followed by a few longitudinal assessments at fixed follow-up times typically invokes traditional data analytic methods (e.g., Analysis of Covariance (ANCOVA), longitudinal data analysis (LDA)). ‘Constrained’ longitudinal data analysis (cLDA) is a well-established unconditional technique that constrains the means of baseline to be equal between arms of the trial. We describe these analysis methods and use an analysis of fasting lipid profiles from the Group Medical Clinics (GMC) longitudinal RCT on diabetic patients to illustrate applications of ANCOVA, LDA and cLDA to demonstrate theoretical concepts of these methods including the impact of missing data.

Results: When analyzing subjects with complete data (n=195), 95% CI coverage are equivalent for ANCOVA and cLDA with an estimated 11.2 mg/dL (95%CI -19.2, -3.3; p=0.006) lower mean LDL-C in GMC compared to usual care. When analyzing all available data (n=233), applying the cLDA model yielded an LDL improvement of 8.9 mg/dl (95% CI -16.7,-1.0; p=0.03) for GMC compared to usual care. The less efficient, LDA analysis yielded an LDL improvement of 7.2 mg/dl (95% CI -17.2,2.8; p=0.15) for GMC compared to usual care.

Conclusions: Under reasonable missing data assumptions cLDA will yield efficient treatment effect estimates and robust inferential statistics. It may be regarded as the method of choice over ANCOVA and LDA.

Strengths and limitations

- Clarification of the statistical methods available for longitudinal RCTs as well as analysis recommendations is warranted.
- In many longitudinal RCTs subjects are measured at baseline, then at the same follow-up occasions with a small number of follow-ups e.g. 2-4. In this design, how should baseline values be handled?
- In practical applications, cLDA, an appropriate generalization of ANCOVA, is the most straightforward to implement and under reasonable missing data assumptions will yield robust estimates of treatment effect differences and valid inferential statistics.

INTRODUCTION

In a recent longitudinal randomized controlled trial (RCT) designed to examine the effect of Group Medical Clinics (GMC) on cardiovascular outcomes in diabetic patients, the statistical inference on the effect of the GMC intervention on low density lipoprotein (LDL) levels was dependent on the method of analysis applied (1). Estimates of mean difference in LDL between treatment arms ranged from 6.9 mg/dl to 11.2 mg/dl depending on the analysis method used. Reviewers questioned the primary analysis method which yielded an LDL improvement of 8.9 mg/dl for the GMC arm compared to usual care ($p=0.03$), and requested an alternative, and seemingly plausible but less powerful analysis that yielded an LDL improvement of 7.2 mg/dl for the GMC arm compared to usual care ($p=0.15$). Thus, the interpretation of the intervention effect varied significantly depending on the analytic technique used. Based on this experience and others of a similar nature, clarification of methods available as well as analysis recommendations is warranted and may be useful, not only to statistical analysts, but clinical researchers evaluating longitudinal RCTs.

Longitudinal Study Designs

In many longitudinal RCTs research participants are measured at the same follow-up measurement occasions and the number of follow-up occasions is small e.g. 2-4. Typically, baseline outcomes are measured prior to randomization and can be ignored, change scores can be calculated, baseline values can be used conditionally as covariates, or baseline values can be part of the outcome response vector. In this design, how should baseline values be handled? Is the question of interest conditional,

1
2
3 given one's baseline what is the difference in outcome between treatment arms, or is
4
5 the question unconditional, what is the treatment difference in change? Or does it
6
7 matter? These fundamental questions and their implications on the method of analysis
8
9 were introduced over 50 years ago in a sentinel paper by Lord (1967) subsequently
10
11 leading to the term 'Lord's paradox(2). In Lord's hypothetical example, the difference
12
13 between the conditional and unconditional questions and how analyses are applied
14
15 yielded different answers and statistical inference when applied to the same data set. In
16
17 spite of extensive literature (3-8) for this type of data, there remains a lack of clarity in
18
19 biomedical research applications regarding longitudinal statistical analysis.
20
21
22
23
24

25 As the issue is addressed here, a major complication requiring consideration in
26
27 the analysis strategy of a longitudinal RCT is missing data. Attrition (dropout) or
28
29 intermittent missingness occurs even in studies with few measurement occasions as we
30
31 are addressing. Reducing the quantity of missing data in an RCT enhances the
32
33 reliability of results (9), however, rarely is it possible to eliminate missing data
34
35 completely. The assumptions and methods applied for handling missing data when
36
37 analyzing "change" in a longitudinal RCT can greatly affect the robustness, validity, and
38
39 power of results and need to be clearly understood.
40
41
42

43 The general consensus in the statistical literature for a longitudinal RCT is that
44
45 the conditional approach, analysis of covariance (ANCOVA), is the most powerful and
46
47 robust method (4, 8) to address the fundamental questions of interest. We take this
48
49 view, and elaborate on its implementation, taking into account the additional
50
51 complication of missing data, and enumerate the benefits of unconditional constrained
52
53
54
55
56
57
58
59
60

1
2
3 longitudinal data analysis (cLDA) models that generalize the ANCOVA approach in the
4
5 framework of longitudinal trials.
6
7

8 **Analysis strategies**

9
10 In a longitudinal RCT of the type addressed here, the outcome of interest is
11 measured at each of the defined assessment periods of the trial (each subject in each
12 arm of the trial). In this setting define the outcome as Y_{it} , the assessed value at each
13 time point, t for each individual i (where $i = 1..n, t = 0..k,$). In its simplest form, a
14 longitudinal RCT is a pre-/post- type of study where there are assessments at two time
15 points, one pre-treatment baseline, Y_{i0} , and one post-treatment, Y_{i1} .
16
17
18
19
20
21
22
23

24 In the more general longitudinal RCT study design with multiple but few follow-up
25 time points, response profile modeling (6) is generally used as it allows for arbitrary
26 patterns in the mean responses over time (i.e. time is dummy coded) as well as for the
27 covariance of responses and can be used for fitting ANCOVA, cLDA, or longitudinal
28 data analysis (LDA) models. From a modeling perspective, the pre-treatment baseline
29 assessment can be viewed as either a fixed predictor of post-treatment outcome or as
30 an outcome assessment. For the former, the conditional ANCOVA approach yields
31 estimates of treatment differences over time given the observed baseline values where
32 post-treatment outcome measurements are the response variables. For the latter, an
33 unconditional model (LDA or cLDA) yields estimates of treatment differences
34 unconditionally, and in the context of a longitudinal RCT, both the pre- and post-
35 treatment outcome measurements are response variables.
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

52 *Conditional vs. Unconditional Models*

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33

The fundamental question in choosing between a conditional or unconditional analysis is how to treat the pre-treatment assessment. In unconditional analysis, baseline is part of the response vector requiring additional assumptions for modeling baseline. In an LDA, there are no modeling constraints on the baseline; separate baseline means are assumed and fit for each randomized group. The general test used for treatment difference over time in LDA is equivalent to a change score type of analysis; comparing change from baseline to follow-up between randomized groups. In contrast for cLDA baseline means are constrained to be equal between the randomized groups; a common baseline mean is assumed and fit across randomized groups. In an RCT, baseline precedes any treatment deliverance and under expectation the baseline means are equal. The test for treatment difference over time in the cLDA is essentially equivalent to a test for treatment difference in an ANCOVA when no outcome data are missing (5, 10-12).

34
35

Conditional Model:

36
37
38
39
40
41
42
43
44
45

An ANCOVA model in a 2-arm ($j=1, 2$) RCT with a pre-/post- design will have study outcome measures at baseline and one follow-up time ($t=0,1$). However, there will only be one response variable per subject ($i=1, \dots, n_j$) as the baseline (Y_0) is a covariate in the model. The model is written as,

$$Y_{ij1} = \mu + \alpha Y_{ij0} + \gamma I(j = 2) + e_{ij1} \quad (1)$$

46
47
48
49
50
51
52
53
54
55

The marginal mean at follow-up time ($t=1$) is conditional on the baseline Y_{ij0} . The parameter α is the slope for the baseline, and γ is the effect of treatment ($j=2$) at time ($t=1$) compared to treatment ($j=1$) adjusted for the baseline effect.

56
57
58
59
60

Unconditional Models:

LDA model -

An LDA model in a 2-arm ($j=1, 2$) RCT with a pre-/post- design will have study outcome measures at both the baseline and follow-up time ($t=0,1$). The model is written as:

$$Y_{ijt} = \mu + \gamma I(j = 2) + \tau I(t = 1) + (\gamma\tau) I(j = 2, t = 1) + e_{ijt} \quad (2)$$

The parameter γ is the difference in baseline means at time ($t=0$) between arms, τ is the mean difference in change from baseline to follow-up for arm ($j=1$), and $(\gamma\tau)$ is the difference in change from baseline to follow-up between arms.

cLDA model -

A cLDA model in a 2-arm ($j=1, 2$) RCT with a pre-/post- design will have study outcome measures at both the baseline and follow-up time ($t=0,1$). The model is written as:

$$Y_{ijt} = \mu' + \tau' I(t = 1) + (\gamma\tau)' I(j = 2, t = 1) + e_{ijt} \quad (3)$$

The parameter τ' is the mean change in outcome from baseline to follow-up time in arm ($j=1$), and $(\gamma\tau)'$ is the mean difference in change from baseline to follow-up between arms, however, since baseline means are assumed equivalent this is the mean difference between arms at the follow-up time.

Conditional and Unconditional Model Comparison

No missing baseline or follow-up data: When there is no missing baseline or follow-up data Liang and Zeger (5) and Liu, et al. (10) have shown that ANCOVA and cLDA models produce identical point estimates for treatment differences. Using our model notation we can state under the case of no missing baseline or follow-up data,

$$\gamma_{ANCOVA} = (\gamma\tau)'_{cLDA} \quad (4)$$

1
2
3 Similarly, for reasonably sized trials the variance of the treatment effect differences will
4
5 be approximately equivalent between ANCOVA and cLDA models (see Appendix A).
6
7

8 Frison and Pocock (3) have shown that estimated treatment differences from
9
10 ANCOVA or of POST treatment means or CHANGE from baseline from an LDA model
11
12 have the same expected value. However, they demonstrate the general superiority of
13
14 inferential power for ANCOVA over both POST and CHANGE from baseline analysis.
15
16 We can see from plotting equations of estimates of variance in Figure 1 that ANCOVA
17
18 and cLDA are superior to LDA-POST and LDA-CHANGE over the range of correlations
19
20 between pre and post measurements (see Appendix A for variance estimates). For
21
22 high correlations between pre- and post- measurements the variances of LDA-CHANGE
23
24 and ANCOVA/cLDA are similar.
25
26
27
28

29 The conditional and unconditional models described above are easily extendible
30
31 from the pre-/post type design with only one follow-up time point to multiple follow-up
32
33 time points by additional dummy variables to represent follow-up times. Similarly,
34
35 comparison of models and estimates from the pre-/post- design apply to the follow-up
36
37 time points. In the longitudinal RCTs with few measurement occasions, the comparison
38
39 of interest is generally the treatment difference at the last follow-up time point (T) – the
40
41 theory as described above for the post time measurement in a pre-/post- design applies
42
43 to analyzing the differences between treatment arms at any specific follow-up time.
44
45
46
47

48 **Motivating Example: Group Medical Clinics Study**

49 We present an analysis of fasting lipid profiles from the Group Medical Clinics
50
51 (GMC) longitudinal RCT on diabetic patients to illustrate applications of ANCOVA, LDA
52
53 and cLDA models and demonstrate theoretical concepts described above including the
54
55
56
57
58
59
60

1
2
3 impact of missing data (1). This study randomized 239 participants at a 3:2 ratio to the
4
5 GMC arm (n=133) and usual care arm (n=106). Fasting lipid profiles are secondary
6
7 outcomes from the GMC study measured at baseline, midpoint (median follow-up 6.8
8
9 months) and end of study (median follow-up 12.8 months)(13).
10
11

12
13 For illustration, we focus analysis on the baseline and 12-month LDL-C
14
15 measurements. For notation, the 2-arms will be denoted with $j=G$ for GMC or $j=U$ for
16
17 Usual Care and time with $t=0$ for baseline and $t=12$ for 12-month. All analyses were
18
19 conducted using SAS 9.2 (SAS Institute, Inc., Cary, NC)
20
21

22 *Methods*

23
24 For completeness of analyses approaches, in addition to the ANCOVA, LDA and
25
26 cLDA models, we also conducted simple post only (SPO) and simple change score
27
28 analysis (SACS). For SPO, we compared mean 12-month LDL-C ($\hat{Y}_{G,12}$) for GMC ($j=G$)
29
30 to mean usual care ($j=U$) 12-month LDL-C ($\hat{Y}_{U,12}$) and for SACS, we compared mean
31
32 LDL-C change from baseline ($\hat{C}_{G,12}$) for GMC ($j=G$) to the mean change from baseline
33
34 LDL-C ($\hat{C}_{U,12}$) for usual care ($j=U$) using two-sample t-tests. We then applied the
35
36 ANCOVA, LDA-CHANGE, and cLDA models to, $Y_{j,0}$ and $Y_{j,12}$, the assessed value of
37
38 LDL-C. All three types of models were fit using linear mixed models (14) and the
39
40 response profile modeling approach. Models were fit using PROC MIXED with restricted
41
42 maximum likelihood (REML) estimation.
43
44
45
46
47
48

49 *Completers Analysis*

50
51 The first set of analyses was applied to participants that had both baseline and
52
53 12-month measurements (completers) to demonstrate theoretical comparisons of
54
55 models as described without the added complication of missing data.
56
57
58
59
60

Results

Among the 239 patients randomized in the study, 195 participants had LDL-C measurements at both baseline and 12-months. All methods yield statistically significant differences in LDL-C between arms (see Table 1) with change score analysis methods (SACS and LDA) right at $p=0.05$. SPO had the largest estimated difference in mean LDL-C between GMC and usual care. As expected, estimated mean treatment differences at 12-months and 95% CI coverage are equivalent for the ANCOVA and cLDA with an estimated 11.2 mg/dL lower mean LDL-C in GMC compared to usual care. The SACS and LDA estimated differences and 95% CI coverage are equivalent with an estimated 10.1 mg/dL lowering between baseline and 12-months of mean LDL-C for GMC compared to usual care. LDA and SACS analyses yield a wider 95% CI than ANCOVA and cLDA methods. For completers, the ANCOVA and cLDA yield the most robust results and illustrate their superiority (as shown in Figure 1) over SPO and change score methods (both SACS and LDA), as the correlation between baseline and 12-month LDL-C values was estimated in the 0.50 range.

All Participants Analysis

The second set of analyses was applied to all participants (i.e. including those with either missing baseline or 12-month measurements) to compare methods and illustrate the impact of missing data. SPO participants with missing 12-month LDL-C are deleted and SACS participants with either missing baseline or end of study measurements are deleted. Similarly, for ANCOVA with only two time points, participants with missing baseline or 12-month measurements are deleted. For LDA and cLDA, all available data were used; no participants were deleted due to missing

1
2
3 data. The estimation procedure used in the mixed model framework for longitudinal
4
5 analysis yields unbiased estimates of parameters when missing outcomes are assumed
6
7 to be ignorable, i.e., when missing values are related to either observed covariates or
8
9 response variables but not to unobserved variables (15, 16).
10
11

12 *Results*

13
14 Among the 239 patients randomized, 6 participants had no baseline or 12-month
15
16 LDL-C so they are excluded yielding 233 patients for the all participants analysis. The
17
18 estimated treatment differences diverge across analysis methods as well as statistical
19
20 significance using a 0.05 alpha (see Table 2). ANCOVA and SACS are equivalent to
21
22 completers analyses, as the 38 cases with either a missing baseline or 12-month LDL-C
23
24 are deleted. ANCOVA and cLDA methods no longer have equivalent estimates of
25
26 treatment differences. All 233 participants contribute at least one measurement to cLDA
27
28 analysis so that participants missing either a baseline or 12-month contribute to
29
30 estimated treatment difference of an 8.9 mg/dL lower mean LDL-C at 12-months in
31
32 GMC compared to usual care (see Table 2). The discrepancy in the estimated
33
34 treatment difference between ANCOVA and cLDA is due to missing data and
35
36 assumptions that are made about missing data. As discussed previously, mixed-effect
37
38 models yield unbiased estimates of treatment effects under the assumption that the
39
40 missing data are conditional on observed quantities. Although this assumption cannot
41
42 be specifically tested, the discrepancy in estimates between ANCOVA and cLDA model
43
44 yields some evidence that data is missing conditional on observed quantities. If data
45
46 were missing completely at random (not conditional on observed quantities) then
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 estimated treatment differences between ANCOVA with case deletion and cLDA should
4
5 be similar.
6
7

8 As shown for ANCOVA and cLDA, LDA and SACS methods estimated
9
10 differences are discrepant due to missing data and assumptions that are made about
11
12 missing data. LDA is using all available data and fit using mixed-effects model, whereas
13
14 in SACS participants with missing data are deleted. Similar to completers analysis, LDA
15
16 methods yield wider 95% CI than both cLDA and ANCOVA.
17
18
19

20 **Missing Data Implications: The Case for cLDA**

21
22 The illustrative example presented above, clearly demonstrates the statistical
23
24 theory described for the analysis methods and the impact of missing data (3, 4, 15) on
25
26 the inference and interpretations of results. In general, the implications of how missing
27
28 data are handled in a trial range from sample size and statistical power loss to
29
30 potentially biased estimates of treatment effects. Olsen, et al. (17) discuss principled
31
32 methods for handling missing data in longitudinal sleep disorder trials that are relevant
33
34 to the types of longitudinal RCTs we are discussing. The default method for handling
35
36 missing data in many software packages is case deletion; if any variable or outcome is
37
38 missing for a subject then the subject is deleted from analysis. In order for estimates of
39
40 treatment effect differences to be unbiased when case deletion occurs, we have to
41
42 assume that the subset of participants with complete data is a random sample of the
43
44 entire study sample. Therefore, for ANCOVA, SACS and SPO analysis above, we are
45
46 making this assumption. For many longitudinal RCTs it is likely that the probability of
47
48 dropout is related to some observable characteristic or quantity (e.g., treatment arm
49
50 assignment) and therefore the subset of participants with complete data is not a random
51
52
53
54
55
56
57
58
59
60

1
2
3 sample of the entire study. Under this assumption unbiased estimates of treatment
4
5 effects can be achieved with mixed-effect models or by performing multiple
6
7 imputation(17). Methods for handling missing data when we cannot assume that
8
9 missing data can be characterized by observable quantities are beyond the scope of
10
11 this paper.
12
13

14
15 The choice of whether to use multiple imputation is up to the analyst and based
16
17 on the untestable assumption that all predictors of missing data can be included in the
18
19 mixed-effect model. In many longitudinal RCTs it is a reasonable assumption that
20
21 treatment group assignment, time of assessment and available outcome assessments
22
23 along with potential baseline stratification variables would be the predictors of missing
24
25 outcome data. If the number of potential predictor variables of missing data is
26
27 expansive and unreasonable to include in the mixed-effects model then it may be
28
29 necessary to perform a multiple imputation analysis.
30
31
32

33
34 Multiple imputation is a more complicated process for handling missing data that
35
36 is also based on the untestable assumption that all predictors of missing data are
37
38 included in the imputation model (16). In the multiple imputation framework,
39
40 comparisons of conditional and unconditional models would be equivalent to what was
41
42 discussed above for no missing baseline or follow-up data. Once imputed data sets are
43
44 created, there are no missing outcome data; the uncertainty from estimating imputed
45
46 values is accounted for in the standard error estimates.
47
48
49

50
51 For ease of implementation and robustness of results (14, 15) mixed-effect
52
53 models with maximum likelihood estimates are a good first-line choice for handling
54
55 missing data. Even when applying mixed-effects models, the treatment of baseline
56
57
58
59
60

1
2
3 values can impact results because of differences in case deletion. In the pre-/post-
4
5 design if the baseline is a covariate (ANCOVA), participants with the missing covariate
6
7 will be deleted. Similarly, with only one follow-up time, if the follow-up time is missing
8
9 then there is no outcome data for the subject to be included. However, with
10
11 unconditional models since both baseline and follow-up are part of the outcome vector,
12
13 if a subject is missing either measurement their available data will be included and
14
15 subject is not deleted. When missing data occur and mixed-effect models are used,
16
17 cLDA models are the optimal choice for providing the most precise estimate of
18
19 treatment differences under a reasonable assumption that missing data is related to
20
21 observable characteristics.
22
23
24
25

26 27 **Summary**

28
29 It does not appear widely appreciated that cLDA can often be regarded as the
30
31 method of choice for the analysis of a longitudinal RCT with few measurement
32
33 occasions. Except for small samples it is equivalent to ANCOVA when there are no
34
35 missing data, and cLDA is an appropriate generalization of ANCOVA under reasonable
36
37 missing data assumptions. Without question, potential baseline imbalance between
38
39 treatment arms has implications for the analysis of longitudinal RCTs and is a source of
40
41 confusion. In an RCT, baseline differences can be attributed to random chance
42
43 assuming there were no problems or issues with the randomization process and no
44
45 significant measurement error issues. LDA or Change score analysis is sometimes
46
47 viewed as a more intuitive analysis; however, for an RCT most often the best method is
48
49 an ANCOVA as far as bias, precision, and power (18). In a longitudinal RCT, when
50
51 within subject correlations are not high, change score or LDA (an unconditional
52
53
54
55
56
57
58
59
60

1
2
3 analysis) are not the most powerful analyses. Frison and Pocock (3) found that
4
5 ANCOVA treatment difference estimates have small bias under the assumption that
6
7 measurement error is a small proportion of the between subject variance, however, bias
8
9 in a change score analysis may be somewhat larger especially if correlations between
10
11 pre- and post -measurements are small.
12
13

14
15 cLDA is sometimes erroneously viewed as more problematic when there is baseline
16
17 imbalance in outcomes between treatment arms. However, cLDA and ANCOVA are
18
19 equivalent when analyzing complete data. cLDA generalizes the ANCOVA approach
20
21 and both are superior to an LDA in many cases. Therefore, the primary analytic issue is
22
23 not necessarily whether or not to perform conditional analysis. ANCOVA is a
24
25 conditional analysis and cLDA is an unconditional analysis, yet both are powerful
26
27 methods that can be applied to examine treatment differences over time in a
28
29 longitudinal RCT. In practical applications, cLDA is the most straightforward to
30
31 implement and under reasonable missing data assumptions will yield robust estimates
32
33 of treatment effect differences and inferential statistics. In most cases, it is the method
34
35 of choice.
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 a. Contributorship statement
4

5 The primary author is a senior statistician in the Biostatistics Unit of the Durham VA
6 Health Services Research and Development group and an Associate Professor in the
7 Department of Biostatistics and Bioinformatics at Duke University Medical Center. All
8 listed authors have contributed to the design and preparation of the manuscript (C.C.,
9 D.E. and R.W.) all data analysis was conducted by C.C.
10
11
12
13
14
15
16

17 b. Competing interests
18

19 We have no conflicts of interest.
20

21 c. Funding
22

23 The study was funded by the US Department of Veterans Affairs Health Services
24 Research and Development Service, which had no direct role in the study's design,
25 conduct, and reporting beyond approval of the scientific protocol in peer review for
26 funding.
27
28
29
30
31
32
33

34 d. Data sharing statement
35

36 No additional data available.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REFERENCES

1. Edelman D, Fredrickson SK, Melnyk SD, Coffman CJ, Jeffreys AS, Datta S, et al. Medical Clinics Versus Usual Care for Patients With Both Diabetes and Hypertension A Randomized Trial. *Annals of Internal Medicine*. 2010;152(11):689-96.
2. Lord FM. A paradox in the interpretation of group comparisons. . *Psychological Bulletin*. 1967;68(5):304-5.
3. Frison L, Pocock SJ. Repeated measures in clinical trials: Analysis using mean summary statistics and its implications for design. *Statistics in Medicine*. 1992;11(13):1685-704.
4. Senn S. Change from baseline and analysis of covariance revisited. *Statistics in Medicine*. 2006;25(24):4334-44.
5. Liang K-Y, Zeger SL. Longitudinal Data Analysis of Continuous and Discrete Responses for Pre-Post Designs. *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)*. 2000;62(1):134-48.
6. Fitzmaurice GM, Laird NM, Ware JH. *Applied longitudinal analysis*. Hoboken, N.J.: Wiley; 2011.
7. Cramer MR. Analysis of Covariance in Parallel-Group Clinical Trials with Pretreatment Baselines. *Biometrics*. 1987;43(4):895-901.
8. Vickers AJ, D.G. Analysing controlled trials with baseline and follow up measurements. *BMJ*. 2001;323(10 November):1123-4.
9. Fleming TR. Addressing Missing Data in Clinical Trials. *Annals of Internal Medicine*. 2011;154(2):113-7.
10. Liu GF, Lu K, Mogg R, Mallick M, Mehrotra DV. Should baseline be a covariate or dependent variable in analyses of change from baseline in clinical trials? *Statistics in Medicine*. 2009;28(20):2509-30.
11. Lu K, Mehrotra DV, Liu G. Sample size determination for constrained longitudinal data analysis. *Statistics in Medicine*. 2009;28(4):679-99.
12. Kenward MG, White IR, Carpenter JR. Should baseline be a covariate or dependent variable in analyses of change from baseline in clinical trials? by G. F. Liu, K. Lu, R. Mogg, M. Mallick and D. V. Mehrotra, *Statistics in Medicine* 2009; 28:2509–2530. *Statistics in Medicine*. 2010;29(13):1455-6.
13. Crowley MJ, Melnyk SD, Ostroff JL, Fredrickson SK, Jeffreys AS, Coffman CJ, et al. Can Group Medical Clinics Improve Lipid Management in Diabetes? *The American Journal of Medicine*(0).
14. Verbeke G, Molenberghs G. *Linear mixed models for longitudinal data*. New York: Springer; 2000.
15. Little RJA, Rubin DB. *Statistical analysis with missing data*: Wiley; 2002.
16. Schafer JL. *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall; 1997.
17. Olsen MK, Stechuchak KM, Edinger JD, Ulmer CS, Woolson RF. Move over LOCF: Principled methods for handling missing data in sleep disorder trials. *Sleep Medicine*. 2012;13(2):123-32.

- 1
2
3 18. Egbewale BE, Lewis M, Sim J. Bias, precision and statistical power of analysis of
4 covariance in the analysis of randomized trials with baseline imbalance: a
5 simulation study. BMC Medical Research Methodology. 2014;14(1):1-12.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

Table 1. Completers only (n=195 participants) – Pre/Post analyses

Model	Outcome (Y_t and/or C_t)	GMC (n=117)	Usual Care (n=78)	GMC vs. Usual Care (95%CI)	P-value
Post-only	12-month (Y_{12})	81.9	94.1	-12.1(-21.5,-2.7)	0.01
SACS	12-month – Baseline (C_{12})	-12.9	-2.8	-10.1(-20.2,0.0)	0.05
ANCOVA	12-month (Y_{12})	82.3	93.5	-11.2(-19.2,-3.3)	0.006
LDA	Baseline (Y_0)	94.8	96.9		
	12-month(Y_{12})	81.9	94.1	-10.1(-20.2,0.0)	0.05
cLDA	Baseline (Y_0)	95.7	95.7		
	12-month (Y_{12})	82.3	93.5	-11.2(-19.2,-3.3)	0.006

Table 2. All available data (n=233 participants) - Pre/Post analyses

Model	Outcome (Y _t and/or C _t)	N	GMC	Usual Care	GMC vs. Usual Care (95%CI)	P-value
Post-only	12-month (Y ₁₂)	204	89.7	96.7	-6.9(-14.2,0.4)	0.07
SACS	12-month – Baseline (C ₁₂)	195	-12.9	-2.8	-10.1(-22.0,-0.8)	0.05
ANCOVA	12-month (Y ₁₂)	195	83.4	94.6	-11.2(-19.2,-3.3)	0.006
LDA ¹	Baseline (Y ₀)	233	96.7	99.6		
	12-month(Y ₁₂)		83.5	93.6	-7.2(-17.2,2.8)	0.15
cLDA ¹	Baseline (Y ₀)	233	98.0	98.0		
	12-month (Y ₁₂)		84.0	92.9	-8.9(-16.7,-1.0)	0.03

¹Baseline LDL-C is missing for 9 participants and 12-month LDL-C is missing for 29 participants

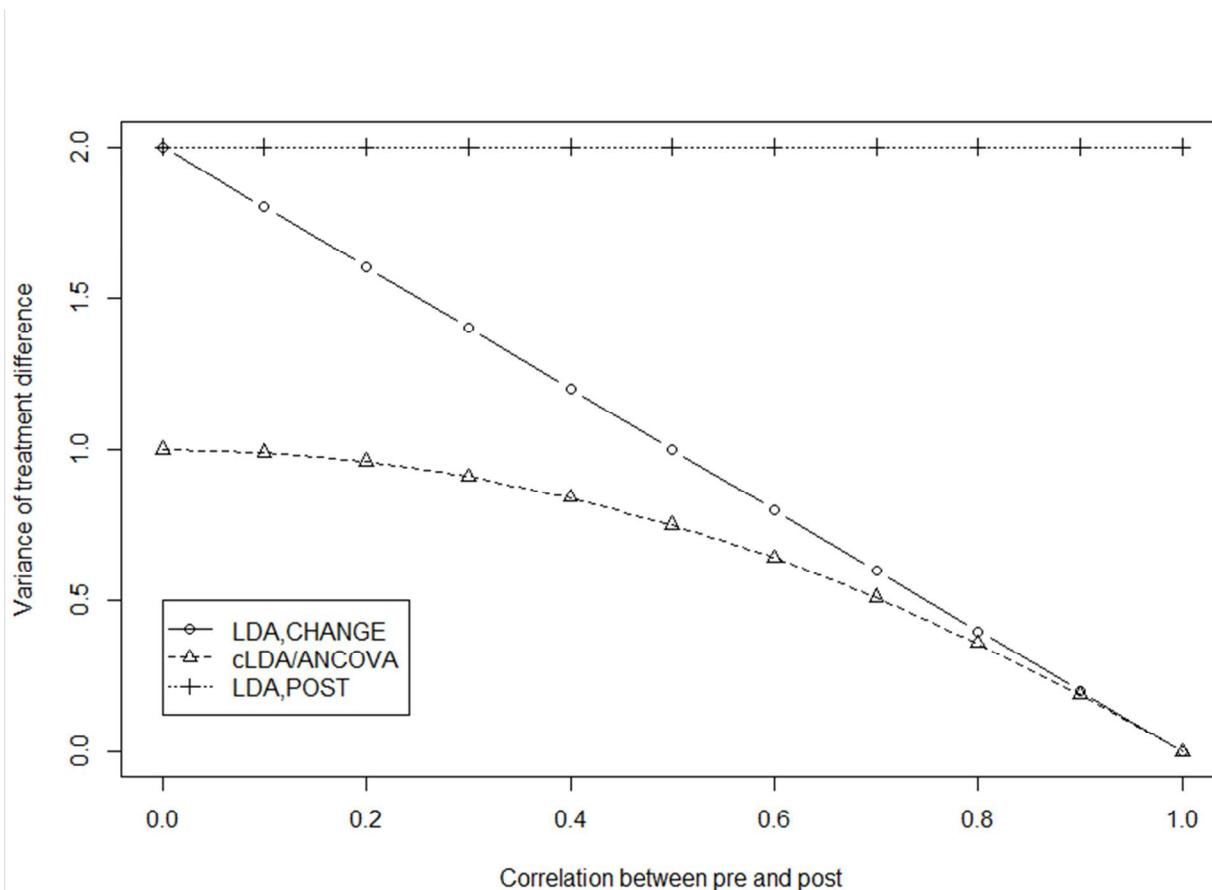


Figure 1. Comparison of variance of treatment difference estimates over the range of correlations between pre- and post- outcome measurements for LDA, cLDA, ANCOVA and SPO methods; plot generated from variance estimate formulas given in Appendix A.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Appendix A.
4
5
6

7
8 Variance Estimates
9

10
11
12 ANCOVA
13
14

15
16 For a reasonably sized RCT, the variance of $\hat{\gamma}$ from equation (1), is approximately
17

$$\left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\} \sigma^2 (1 - \rho_{01}^2)$$

18
19 where ρ_{01} is the within subject correlation between baseline and follow-up
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
measurements.

LDA model

The variance of this estimate ($\widehat{\gamma\tau}$) from equation (2) when assuming that the baseline and follow-up treatment variance is equal ($\sigma_0^2 = \sigma_1^2 = \sigma^2$) can be written as,

$$\left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\} 2\sigma^2 (1 - \rho_{01})$$

We could also estimate the treatment difference at follow-up (POST only) between the treatment arms which is ($\hat{\gamma} + \widehat{\gamma\tau}$). The variance of this estimate ($\hat{\gamma} + \widehat{\gamma\tau}$) is then

$$\left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\} 2\sigma_1^2$$

cLDA Model

The variance of this estimate ($\widehat{\gamma\tau}$)' from equation (3) is then (10, 11)

$$\left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\} \sigma_1^2 (1 - \rho_{01}^2)$$

BMJ Open

To condition or not condition? Analyzing “change” in longitudinal randomized controlled trials

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2016-013096.R1
Article Type:	Research
Date Submitted by the Author:	06-Oct-2016
Complete List of Authors:	Coffman, Cynthia; Duke University Medical Center, Biostatistics and Bioinformatics; Durham VA Medical Center, Health Services Research and Development Edelman, David; Duke University Medical Center, General Internal Medicine; Durham VA Medical Center, Health Services Research and Development Woolson, Robert; Duke University Medical Center, Biostatistics and Bioinformatics
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Health services research
Keywords:	Constrained longitudinal data analysis, ANCOVA, Longitudinal RCT

SCHOLARONE™
Manuscripts

1
2
3 **To condition or not condition? Analyzing “change” in longitudinal randomized**
4 **controlled trials**
5
6
7
8
9

10 Cynthia J. Coffman PhD^{1,2}, Edelman, David MD^{1,3} and Robert F. Woolson, PhD^{1,2}
11
12

13
14
15 ¹Health Services Research, Durham Veterans Affairs Medical Center,
16

17 ²Department of Biostatistics and Bioinformatics, Duke University Medical Center
18

19 ³Department of Medicine, Duke University Medical Center
20
21
22

23
24 Corresponding author: Cynthia J. Coffman, PhD,
25

26 Durham VA Medical Center, HSR&D(152)
27

28 508 Fulton St.
29

30 Durham, NC 27705,
31

32
33 Tel: 919-286-6936 Fax: 919-416-5836
34

35 Email: Cynthia.Coffman@duke.edu
36
37
38
39

40 **Running Title:** Analyzing “change” in longitudinal randomized controlled trials
41
42
43

44 Word Count: 3439
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ABSTRACT

Objective: The statistical analysis for a two-arm randomized controlled trial (RCT) with a baseline outcome followed by a few assessments at fixed follow-up times typically invokes traditional analytic methods (e.g., Analysis of Covariance (ANCOVA), longitudinal data analysis (LDA)). ‘Constrained’ longitudinal data analysis (cLDA) is a well-established unconditional technique that constrains means of baseline to be equal between arms. We use an analysis of fasting lipid profiles from the Group Medical Clinics (GMC) longitudinal RCT on diabetic patients to illustrate applications of ANCOVA, LDA and cLDA to demonstrate theoretical concepts of these methods including the impact of missing data.

Methods: For the analysis of the illustrated example, all models were fit using linear mixed models to subjects with only complete data and to subjects using all available data.

Results: With complete data (n=195), 95% CI coverage are equivalent for ANCOVA and cLDA with an estimated 11.2 mg/dL (95%CI -19.2, -3.3; p=0.006) lower mean LDL-C in GMC compared to usual care. With all available data (n=233), applying the cLDA model yielded an LDL improvement of 8.9 mg/dl (95% CI -16.7,-1.0; p=0.03) for GMC compared to usual care. The less efficient, LDA analysis yielded an LDL improvement of 7.2 mg/dl (95% CI -17.2,2.8; p=0.15) for GMC compared to usual care.

Conclusions: Under reasonable missing data assumptions cLDA will yield efficient treatment effect estimates and robust inferential statistics. It may be regarded as the method of choice over ANCOVA and LDA.

Strengths and limitations

- Clarification of the statistical methods available for longitudinal RCTs as well as analysis recommendations is warranted.
- In many longitudinal RCTs subjects are measured at baseline, then at the same follow-up occasions with a small number of follow-ups e.g. 2-4. In this design, how should baseline values be handled?
- In practical applications, cLDA, an appropriate generalization of ANCOVA, is the most straightforward to implement and under reasonable missing data assumptions will yield robust estimates of treatment effect differences and valid inferential statistics.

INTRODUCTION

In a recent longitudinal randomized controlled trial (RCT) designed to examine the effect of Group Medical Clinics (GMC) on cardiovascular outcomes in diabetic patients, the statistical inference on the effect of the GMC intervention on low density lipoprotein (LDL) levels was dependent on the method of analysis applied (1). Estimates of mean difference in LDL between treatment arms ranged from 6.9 mg/dl to 11.2 mg/dl depending on the analysis method used. Reviewers questioned the primary analysis method which yielded an LDL improvement of 8.9 mg/dl for the GMC arm compared to usual care ($p=0.03$), and requested an alternative, and seemingly plausible but less powerful analysis that yielded an LDL improvement of 7.2 mg/dl for the GMC arm compared to usual care ($p=0.15$). Thus, the interpretation of the intervention effect varied significantly depending on the analytic technique used. Based on this experience and others of a similar nature, clarification of methods available as well as analysis recommendations is warranted and may be useful, not only to statistical analysts, but clinical researchers evaluating longitudinal RCTs.

Longitudinal Study Designs

In many longitudinal RCTs research participants are measured at the same follow-up measurement occasions and the number of follow-up occasions is small e.g. 2-4. Typically, baseline outcomes are measured prior to randomization. In analyses, baseline outcomes can be ignored, used to calculate change scores, used conditionally as covariates, or can be part of the outcome response vector. In this design, how should baseline values be handled? Is the question of interest conditional, given one's

1
2
3 baseline what is the difference in outcome between treatment arms, or is the question
4
5 unconditional, what is the treatment difference in change? Or does it matter? These
6
7 fundamental questions and their implications on the method of analysis were introduced
8
9 over 50 years ago in a sentinel paper by Lord (1967) subsequently leading to the term
10
11 'Lord's paradox(2). In Lord's hypothetical example, the difference between the
12
13 conditional and unconditional questions and how analyses are applied yielded different
14
15 answers and statistical inference when applied to the same data set. In spite of
16
17 extensive literature (3-8) for this type of data, there remains a lack of clarity in
18
19 biomedical research applications regarding longitudinal statistical analysis.
20
21
22
23

24
25 As the issue is addressed here, a major complication requiring consideration in
26
27 the analysis strategy of a longitudinal RCT is missing data. Attrition (dropout) or
28
29 intermittent missingness occurs even in studies with few measurement occasions as we
30
31 are addressing. Reducing the quantity of missing data in an RCT enhances the
32
33 reliability of results (9); however, rarely is it possible to eliminate missing data
34
35 completely. The assumptions and methods applied for handling missing data when
36
37 analyzing "change" in a longitudinal RCT can greatly affect the robustness, validity, and
38
39 power of results and need to be clearly understood.
40
41
42

43
44 The general consensus in the statistical literature for a longitudinal RCT is that
45
46 the conditional approach, i.e. analysis of covariance (ANCOVA), is the most powerful
47
48 and robust method (4, 8) to address the fundamental questions of interest. We take this
49
50 view, and elaborate on its implementation, taking into account the additional
51
52 complication of missing data, and enumerate the benefits of unconditional constrained
53
54
55
56
57
58
59
60

1
2
3 longitudinal data analysis (cLDA) models, illustrating that they generalize the ANCOVA
4
5 approach in the longitudinal trials' setting.
6
7

8 **Analysis strategies**

9
10 In a longitudinal RCT of the type addressed here, the outcome of interest is
11 measured at each of the defined assessment periods of the trial (each subject in each
12 arm of the trial). In this setting define the outcome as Y_{it} , the assessed value at each
13 time point, t for each individual i (where $i = 1..n, t = 0..k$). In its simplest form, a
14 longitudinal RCT is a pre-/post- type of study where there are assessments at two time
15 points, one pre-treatment baseline, Y_{i0} , and one post-treatment, Y_{i1} .
16
17
18
19
20
21
22
23

24 In the more general longitudinal RCT study design with multiple but few follow-up
25 time points, response profile modeling (6) also known as mixed model repeated
26 measures (MMRM) (10) is generally used as it allows for arbitrary patterns in the mean
27 responses over time (i.e. time is dummy coded) as well as for the covariance of
28 responses and can be used for fitting ANCOVA, cLDA, or longitudinal data analysis
29 (LDA) models. From a modeling perspective, the pre-treatment baseline assessment
30 can be viewed as either a fixed predictor of post-treatment outcome or as an outcome
31 assessment. For the former, the conditional ANCOVA approach yields estimates of
32 treatment differences over time given the observed baseline values where post-
33 treatment outcome measurements are the response variables. For the latter, an
34 unconditional model (LDA or cLDA) yields estimates of treatment differences
35 unconditionally, and in the context of a longitudinal RCT, both the pre- and post-
36 treatment outcome measurements are response variables.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54

55 *Conditional vs. Unconditional Models*

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33

A fundamental difference between a conditional or unconditional analysis is in the modeling of the pre-treatment assessment. In unconditional analysis, baseline is part of the response vector requiring additional assumptions for modeling baseline. In an LDA, there are no modeling constraints on the baseline; separate baseline means are assumed and fit for each randomized group. The general test used for treatment difference over time in LDA is equivalent to a change score type of analysis; comparing change from baseline to follow-up between randomized groups. In contrast for cLDA baseline means are constrained to be equal between the randomized groups; a common baseline mean is assumed and fit across randomized groups. In an RCT, baseline precedes any treatment deliverance and under expectation the baseline means are equal. The test for treatment difference over time in the cLDA is essentially equivalent to a test for treatment difference in an ANCOVA when no outcome data are missing (5, 11-13).

34
35

Conditional Model:

36
37
38
39
40
41
42
43
44
45

An ANCOVA model in a 2-arm ($j=1, 2$) RCT with a pre-/post- design will have study outcome measures at baseline and one follow-up time ($t=0,1$). However, there will only be one response variable per subject ($i=1, \dots, n_j$) as the baseline (Y_0) is a covariate in the model. The model is written as,

$$Y_{ij1} = \mu + \alpha Y_{ij0} + \beta I(j = 2) + e_{ij1} \quad (1)$$

46
47
48
49
50
51
52
53
54
55

The marginal mean at follow-up time ($t=1$) is conditional on the baseline Y_{ij0} . The parameter α is the slope for the baseline, and β is the effect of treatment ($j=2$) at time ($t=1$) compared to treatment ($j=1$) adjusted for the baseline effect.

56
57
58
59
60

Unconditional Models:

LDA model -

An LDA model in a 2-arm ($j=1, 2$) RCT with a pre-/post- design will have study outcome measures at both the baseline and follow-up time ($t=0,1$). The model is written as:

$$Y_{ijt} = \mu + \gamma I(j = 2) + \tau I(t = 1) + \delta I(j = 2, t = 1) + e_{ijt} \quad (2)$$

The parameter γ is the difference in baseline means at time ($t=0$) between arms, τ is the mean difference in change from baseline to follow-up for arm ($j=1$), and δ is the difference in change from baseline to follow-up between arms.

cLDA model -

A cLDA model in a 2-arm ($j=1, 2$) RCT with a pre-/post- design will have study outcome measures at both the baseline and follow-up time ($t=0,1$). The model is written as:

$$Y_{ijt} = \mu' + \tau' I(t = 1) + \delta' I(j = 2, t = 1) + e_{ijt} \quad (3)$$

The parameter τ' is the mean change in outcome from baseline to follow-up time in arm ($j=1$), and δ is the mean difference in change from baseline to follow-up between arms, however, since baseline means are assumed equivalent this is the mean difference between arms at the follow-up time.

Conditional and Unconditional Model Comparison

No missing baseline or follow-up data: When there is no missing baseline or follow-up data Liang and Zeger (5) and Liu, et al. (11) have shown that ANCOVA and cLDA models produce identical point estimates for treatment differences. Using our model notation we can state under the case of no missing baseline or follow-up data,

$$\beta_{ANCOVA} = \delta'_{cLDA} \quad (4)$$

1
2
3 Similarly, for reasonably sized trials the variance of the treatment effect differences will
4
5 be approximately equivalent between ANCOVA and cLDA models (see Appendix A).
6
7

8 Frison and Pocock (3) have shown that estimated treatment differences from
9
10 ANCOVA or of POST treatment means or CHANGE from baseline from an LDA model
11
12 have the same expected value. However, they demonstrate the general superiority of
13
14 inferential power for ANCOVA over both POST and CHANGE from baseline analysis.
15
16 We can see from plotting equations of estimates of variance in Figure 1 that ANCOVA
17
18 and cLDA are superior to LDA-POST and LDA-CHANGE over the range of correlations
19
20 between pre and post measurements (see Appendix A for variance estimates). For
21
22 high correlations between pre- and post- measurements the variances of LDA-CHANGE
23
24 and ANCOVA/cLDA are similar.
25
26
27
28

29 The conditional and unconditional models described above are easily extendible
30
31 from the pre-/post type design with only one follow-up time point to multiple follow-up
32
33 time points by additional dummy variables to represent follow-up times. Similarly,
34
35 comparison of models and estimates from the pre-/post- design apply to the follow-up
36
37 time points. In the longitudinal RCTs with few measurement occasions, the comparison
38
39 of interest is generally the treatment difference at the last follow-up time point (T) – the
40
41 theory as described above for the post time measurement in a pre-/post- design applies
42
43 to analyzing the differences between treatment arms at any specific follow-up time.
44
45
46
47

48 **Motivating Example: Group Medical Clinics Study**

49
50 We present an analysis of fasting lipid profiles from the Group Medical Clinics
51
52 (GMC) longitudinal RCT on diabetic patients to illustrate applications of ANCOVA, LDA
53
54 and cLDA models and demonstrate theoretical concepts described above including the
55
56
57
58
59
60

1
2
3 impact of missing data (1). This study randomized 239 participants at a 3:2 ratio to the
4
5 GMC arm (n=133) and usual care arm (n=106). Fasting lipid profiles are secondary
6
7 outcomes from the GMC study measured at baseline, midpoint (median follow-up 6.8
8
9 months) and end of study (median follow-up 12.8 months)(14).
10
11

12 For illustration, we focus analysis on the baseline and 12-month LDL-C
13
14 measurements. For notation, the 2-arms will be denoted with $j=G$ for GMC or $j=U$ for
15
16 Usual Care and time with $t=0$ for baseline and $t=12$ for 12-month. All analyses were
17
18 conducted using SAS 9.2 (SAS Institute, Inc., Cary, NC)
19
20

21 *Methods*

22 For completeness of analyses approaches, in addition to the ANCOVA, LDA and
23
24 cLDA models, we also conducted simple post only (SPO) and simple change score
25
26 analysis (SACS). For SPO, we compared mean 12-month LDL-C ($\hat{Y}_{G,12}$) for GMC ($j=G$)
27
28 to mean usual care ($j=U$) 12-month LDL-C ($\hat{Y}_{U,12}$) and for SACS, we compared mean
29
30 LDL-C change from baseline ($\hat{C}_{G,12}$) for GMC ($j=G$) to the mean change from baseline
31
32 LDL-C ($\hat{C}_{U,12}$) for usual care ($j=U$) using two-sample t-tests. We then applied the
33
34 ANCOVA, LDA-CHANGE, and cLDA models to, $Y_{j,0}$ and $Y_{j,12}$, the assessed value of
35
36 LDL-C. All three types of models were fit using linear mixed models (15) and the
37
38 response profile modeling approach. Models were fit using PROC MIXED with restricted
39
40 maximum likelihood (REML) estimation and unstructured covariance.
41
42
43
44
45
46
47
48

49 *Completers Analysis*

50 The first set of analyses was applied to participants that had both baseline and
51
52 12-month measurements (completers) to demonstrate theoretical comparisons of
53
54 models as described without the added complication of missing data.
55
56
57
58
59
60

Results

Among the 239 patients randomized in the study, 195 participants had LDL-C measurements at both baseline and 12-months. All methods yield statistically significant differences in LDL-C between arms (see Table 1) with change score analysis methods (SACS and LDA) right at $p=0.05$. SPO had the largest estimated difference in mean LDL-C between GMC and usual care. As expected, estimated mean treatment differences at 12-months and 95% CI coverage are equivalent for the ANCOVA and cLDA with an estimated 11.2 mg/dL lower mean LDL-C in GMC compared to usual care. The SACS and LDA estimated differences and 95% CI coverage are equivalent with an estimated 10.1 mg/dL lowering between baseline and 12-months of mean LDL-C for GMC compared to usual care. LDA and SACS analyses yield a wider 95% CI than ANCOVA and cLDA methods. For completers, the ANCOVA and cLDA yield the most robust results and illustrate their superiority (as shown in Figure 1) over SPO and change score methods (both SACS and LDA), as the correlation between baseline and 12-month LDL-C values was estimated in the 0.50 range.

All Participants Analysis

The second set of analyses was applied to all participants (i.e. including those with either missing baseline or 12-month measurements) to compare methods and illustrate the impact of missing data. SPO participants with missing 12-month LDL-C are deleted and SACS participants with either missing baseline or end of study measurements are deleted. Similarly, for ANCOVA with only two time points, participants with missing baseline or 12-month measurements are deleted. For LDA and cLDA, all available data were used; no participants were deleted due to missing

1
2
3 data. The estimation procedure used in the mixed model framework for longitudinal
4
5 analysis yields unbiased estimates of parameters when missing outcomes are assumed
6
7 to be ignorable, i.e., when missing values are related to either observed covariates or
8
9 response variables but not to unobserved variables (16, 17).
10
11

12 *Results*

13
14 Among the 239 patients randomized, 6 participants had no baseline or 12-month
15
16 LDL-C so they are excluded yielding 233 patients for the all participants analysis. The
17
18 estimated treatment differences diverge across analysis methods as well as statistical
19
20 significance using a 0.05 alpha (see Table 2). ANCOVA and SACS are equivalent to
21
22 completers analyses, as the 38 cases with either a missing baseline or 12-month LDL-C
23
24 are deleted. ANCOVA and cLDA methods no longer have equivalent estimates of
25
26 treatment differences. All 233 participants contribute at least one measurement to cLDA
27
28 analysis so that participants missing either a baseline or 12-month contribute to
29
30 estimated treatment difference of an 8.9 mg/dL lower mean LDL-C at 12-months in
31
32 GMC compared to usual care (see Table 2). The discrepancy in the estimated
33
34 treatment difference between ANCOVA and cLDA is due to missing data and
35
36 assumptions that are made about missing data. As discussed previously, mixed-effect
37
38 models yield unbiased estimates of treatment effects under the assumption that the
39
40 missing data are conditional on observed quantities. Although this assumption cannot
41
42 be specifically tested, the discrepancy in estimates between ANCOVA and cLDA model
43
44 yields some evidence that data is missing conditional on observed quantities. If data
45
46 were missing completely at random (not conditional on observed quantities) then
47
48 estimated treatment differences between ANCOVA with case deletion and cLDA should
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 be similar. Similarly, differences in estimates from the POST-only as well as LDA
4
5 models between completers only and participants using all available data, provide
6
7 evidence that subjects with missing observations may not be a completely random
8
9 sample from the study population. Therefore, any analyses based on completers only
10
11 (including the ANCOVA model) may be biased.
12
13

14
15 As shown for ANCOVA and cLDA, LDA and SACS methods estimated
16
17 differences are discrepant due to missing data and assumptions that are made about
18
19 missing data. LDA is using all available data and fit using mixed-effects model, whereas
20
21 in SACS participants with missing data are deleted. Similar to completers analysis, LDA
22
23 methods yield wider 95% CI than both cLDA and ANCOVA.
24
25

26 27 **Discussion**

28 29 *Missing Data Implications: The Case for cLDA*

30
31 The illustrative example presented above, clearly demonstrates the statistical
32
33 theory described for the analysis methods and the impact of missing data (3, 4, 16) on
34
35 the inference and interpretations of results. In general, the implications of how missing
36
37 data are handled in a trial range from sample size and statistical power loss to
38
39 potentially biased estimates of treatment effects. Olsen, et al. (18) discuss principled
40
41 methods for handling missing data in longitudinal sleep disorder trials that are relevant
42
43 to the types of longitudinal RCTs we are discussing. The default method for handling
44
45 missing data in many software packages is case deletion; if any variable or outcome is
46
47 missing for a subject then the subject is deleted from analysis. In order for estimates of
48
49 treatment effect differences to be unbiased when case deletion occurs, we have to
50
51 assume that the subset of participants with complete data is a random sample of the
52
53
54
55
56
57
58
59
60

1
2
3 entire study sample. Therefore, for ANCOVA, SACS and SPO analysis above, we are
4
5 making this assumption. For many longitudinal RCTs it is likely that the probability of
6
7 dropout is related to some observable characteristic or quantity (e.g., treatment arm
8
9 assignment) and therefore the subset of participants with complete data is not a random
10
11 sample of the entire study. Under this assumption unbiased estimates of treatment
12
13 effects can be achieved with mixed-effect models or by performing multiple
14
15 imputation(18). Methods for handling missing data when we cannot assume that
16
17 missing data can be characterized by observable quantities are beyond the scope of
18
19 this paper.
20
21
22
23
24

25 The choice of whether to use multiple imputation is up to the analyst and based
26
27 on the untestable assumption that all predictors of missing data can be included in the
28
29 mixed-effect model. In many longitudinal RCTs it is a reasonable assumption that
30
31 treatment group assignment, time of assessment and available outcome assessments
32
33 along with potential baseline stratification variables would be the predictors of missing
34
35 outcome data. If the number of potential predictor variables of missing data is
36
37 expansive and unreasonable to include in the mixed-effects model then it may be
38
39 necessary to perform a multiple imputation analysis.
40
41
42
43

44 Multiple imputation is a more complicated process for handling missing data that
45
46 is also based on the untestable assumption that all predictors of missing data are
47
48 included in the imputation model (17). In the multiple imputation framework,
49
50 comparisons of conditional and unconditional models would be equivalent to what was
51
52 discussed above for no missing baseline or follow-up data. Once imputed data sets are
53
54
55
56
57
58
59
60

1
2
3 created, there are no missing outcome data; the uncertainty from estimating imputed
4
5 values is accounted for in the standard error estimates.
6
7

8 For ease of implementation and robustness of results (15, 16) mixed-effect
9
10 models with maximum likelihood estimates are a good first-line choice for handling
11
12 missing data. Even when applying mixed-effects models, the treatment of baseline
13
14 values can impact results because of differences in case deletion. In the pre-/post-
15
16 design if the baseline is a covariate (ANCOVA), participants with the missing covariate
17
18 will be deleted. Similarly, with only one follow-up time, if the follow-up time is missing
19
20 then there is no outcome data for the subject to be included. However, with
21
22 unconditional models since both baseline and follow-up are part of the outcome vector,
23
24 if a subject is missing either measurement their available data will be included and
25
26 subject is not deleted. When missing data occur and mixed-effect models are used,
27
28 cLDA models are the optimal choice for providing the most precise estimate of
29
30 treatment differences under a reasonable assumption that missing data is related to
31
32 observable characteristics. Finally, it should also be noted that these methods will not
33
34 diverge for every data set to which they may be applied to – however as certain missing
35
36 data assumptions cannot specifically be tested, cLDA models are the optimal choice
37
38 across a range of conditions and data sets.
39
40
41
42
43
44

45 *Summary*

46
47
48 It does not appear widely appreciated that cLDA can often be regarded as the
49
50 method of choice for the analysis of a longitudinal RCT with few measurement
51
52 occasions. Except for small samples it is equivalent to ANCOVA when there are no
53
54 missing data, and cLDA is an appropriate generalization of ANCOVA under reasonable
55
56
57
58
59
60

1
2
3 missing data assumptions. Without question, potential baseline imbalance between
4
5 treatment arms has implications for the analysis of longitudinal RCTs and is a source of
6
7 confusion. In an RCT, baseline differences can be attributed to random chance
8
9 assuming there were no problems or issues with the randomization process and no
10
11 significant measurement error issues. LDA or Change score analysis is sometimes
12
13 viewed as a more intuitive analysis; however, for an RCT most often the best method is
14
15 an ANCOVA as far as bias, precision, and power (19). In a longitudinal RCT, when
16
17 within subject correlations are not high, change score or LDA (an unconditional
18
19 analysis) are not the most powerful analyses. Frison and Pocock (3) found that
20
21 ANCOVA treatment difference estimates have small bias under the assumption that
22
23 measurement error is a small proportion of the between subject variance, however, bias
24
25 in a change score analysis may be somewhat larger especially if correlations between
26
27 pre- and post -measurements are small.
28
29
30
31
32

33
34 cLDA is sometimes erroneously viewed as more problematic when there is baseline
35
36 imbalance in outcomes between treatment arms. However, cLDA and ANCOVA are
37
38 equivalent when analyzing complete data. cLDA generalizes the ANCOVA approach
39
40 and both are superior to an LDA in many cases. Therefore, the primary analytic issue is
41
42 not necessarily whether or not to perform conditional analysis. ANCOVA is a
43
44 conditional analysis and cLDA is an unconditional analysis, yet both are powerful
45
46 methods that can be applied to examine treatment differences over time in a
47
48 longitudinal RCT. In practical applications, cLDA is the most straightforward to
49
50 implement and under reasonable missing data assumptions will yield robust estimates
51
52
53
54
55
56
57
58
59
60

1
2
3 of treatment effect differences and inferential statistics. In most cases, it is the method
4
5 of choice.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

1
2
3 a. Contributorship statement
4

5 The primary author is a senior statistician in the Biostatistics Unit of the Durham VA
6 Health Services Research and Development group and an Associate Professor in the
7 Department of Biostatistics and Bioinformatics at Duke University Medical Center. All
8 listed authors have contributed to the design and preparation of the manuscript (C.C.,
9 D.E. and R.W.) all data analysis was conducted by C.C.
10
11
12
13
14
15
16

17 b. Competing interests
18

19 We have no conflicts of interest.
20

21 c. Funding
22

23 The study was funded by the US Department of Veterans Affairs Health Services
24 Research and Development Service, which had no direct role in the study's design,
25 conduct, and reporting beyond approval of the scientific protocol in peer review for
26 funding.
27
28
29
30
31
32
33

34 d. Data sharing statement
35

36 No additional data available.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REFERENCES

1. Edelman D, Fredrickson SK, Melnyk SD, Coffman CJ, Jeffreys AS, Datta S, et al. Medical Clinics Versus Usual Care for Patients With Both Diabetes and Hypertension A Randomized Trial. *Annals of Internal Medicine*. 2010;152(11):689-96.
2. Lord FM. A paradox in the interpretation of group comparisons. . *Psychological Bulletin*. 1967;68(5):304-5.
3. Frison L, Pocock SJ. Repeated measures in clinical trials: Analysis using mean summary statistics and its implications for design. *Statistics in Medicine*. 1992;11(13):1685-704.
4. Senn S. Change from baseline and analysis of covariance revisited. *Statistics in Medicine*. 2006;25(24):4334-44.
5. Liang K-Y, Zeger SL. Longitudinal Data Analysis of Continuous and Discrete Responses for Pre-Post Designs. *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)*. 2000;62(1):134-48.
6. Fitzmaurice GM, Laird NM, Ware JH. *Applied longitudinal analysis*. Hoboken, N.J.: Wiley; 2011.
7. Crager MR. Analysis of Covariance in Parallel-Group Clinical Trials with Pretreatment Baselines. *Biometrics*. 1987;43(4):895-901.
8. Vickers AJaA, D.G. Analysing controlled trials with baseline and follow up measurements. *BMJ*. 2001;323(10 November):1123-4.
9. Fleming TR. Addressing Missing Data in Clinical Trials. *Annals of Internal Medicine*. 2011;154(2):113-7.
10. Mallinckrod CH, Lane PW, Schnell D, Peng Y, Mancuso JP. Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials. *Drug Information Journal*. 2008;42(4):303-19.
11. Liu GF, Lu K, Mogg R, Mallick M, Mehrotra DV. Should baseline be a covariate or dependent variable in analyses of change from baseline in clinical trials? *Statistics in Medicine*. 2009;28(20):2509-30.
12. Lu K, Mehrotra DV, Liu G. Sample size determination for constrained longitudinal data analysis. *Statistics in Medicine*. 2009;28(4):679-99.
13. Kenward MG, White IR, Carpenter JR. Should baseline be a covariate or dependent variable in analyses of change from baseline in clinical trials? by G. F. Liu, K. Lu, R. Mogg, M. Mallick and D. V. Mehrotra, *Statistics in Medicine* 2009; 28:2509–2530. *Statistics in Medicine*. 2010;29(13):1455-6.
14. Crowley MJ, Melnyk SD, Ostroff JL, Fredrickson SK, Jeffreys AS, Coffman CJ, et al. Can Group Medical Clinics Improve Lipid Management in Diabetes? *The American Journal of Medicine*(0).
15. Verbeke G, Molenberghs G. *Linear mixed models for longitudinal data*. New York: Springer; 2000.
16. Little RJA, Rubin DB. *Statistical analysis with missing data*: Wiley; 2002.
17. Schafer JL. *Analysis of Imcomplete Multivariate Data*. London: Chapman & Hall; 1997.

- 1
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8
 - 9
 - 10
 - 11
 - 12
 - 13
 - 14
 - 15
 - 16
 - 17
 - 18
 - 19
 - 20
 - 21
 - 22
 - 23
 - 24
 - 25
 - 26
 - 27
 - 28
 - 29
 - 30
 - 31
 - 32
 - 33
 - 34
 - 35
 - 36
 - 37
 - 38
 - 39
 - 40
 - 41
 - 42
 - 43
 - 44
 - 45
 - 46
 - 47
 - 48
 - 49
 - 50
 - 51
 - 52
 - 53
 - 54
 - 55
 - 56
 - 57
 - 58
 - 59
 - 60
18. Olsen MK, Stechuchak KM, Edinger JD, Ulmer CS, Woolson RF. Move over LOCF: Principled methods for handling missing data in sleep disorder trials. *Sleep Medicine*. 2012;13(2):123-32.
19. Egbewale BE, Lewis M, Sim J. Bias, precision and statistical power of analysis of covariance in the analysis of randomized trials with baseline imbalance: a simulation study. *BMC Medical Research Methodology*. 2014;14(1):1-12.

For peer review only

Table 1. Completers only (n=195 participants) – Pre/Post analyses

Model	Outcome (Y_t and/or C_t)	GMC (n=117)	Usual Care (n=78)	GMC vs. Usual Care (95%CI)	P-value
Post-only	12-month (Y ₁₂)	81.9	94.1	-12.1(-21.5,-2.7)	0.01
SACS	12-month – Baseline (C ₁₂)	-12.9	-2.8	-10.1(-20.2,0.0)	0.05
ANCOVA	12-month (Y ₁₂)	82.3	93.5	-11.2(-19.2,-3.3)	0.006
LDA	Baseline (Y ₀)	94.8	96.9		
	12-month(Y ₁₂)	81.9	94.1	-10.1(-20.2,0.0)	0.05
cLDA	Baseline (Y ₀)	95.7	95.7		
	12-month (Y ₁₂)	82.3	93.5	-11.2(-19.2,-3.3)	0.006

Table 2. All available data (n=233 participants) - Pre/Post analyses

Model	Outcome (Y _t and/or C _t)	N	GMC	Usual Care	GMC vs. Usual Care (95%CI)	P-value
Post-only	12-month (Y ₁₂)	204	89.7	96.7	-6.9(-14.2,0.4)	0.07
SACS	12-month – Baseline (C ₁₂)	195	-12.9	-2.8	-10.1(-22.0,-0.8)	0.05
ANCOVA	12-month (Y ₁₂)	195	83.4	94.6	-11.2(-19.2,-3.3)	0.006
LDA ¹	Baseline (Y ₀)	233	96.7	99.6		
	12-month(Y ₁₂)		83.5	93.6	-7.2(-17.2,2.8)	0.15
cLDA ¹	Baseline (Y ₀)	233	98.0	98.0		
	12-month (Y ₁₂)		84.0	92.9	-8.9(-16.7,-1.0)	0.03

¹Baseline LDL-C is missing for 9 participants and 12-month LDL-C is missing for 29 participants

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

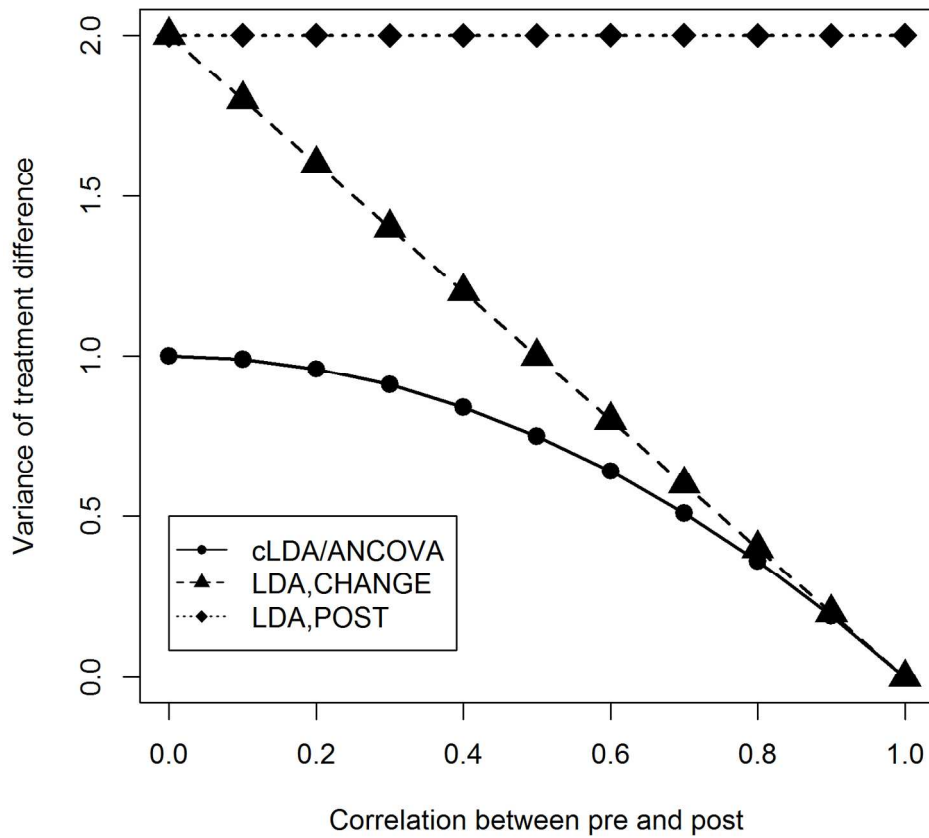


Figure 1. Comparison of variance of treatment difference estimates over the range of correlations between pre- and post- outcome measurements for LDA, cLDA, ANCOVA and SPO methods; plot generated from variance estimate formulas given in Appendix A.

Figure 1
152x152mm (300 x 300 DPI)

Appendix A

Variance Estimates assuming that baseline and follow-up treatment variance is equal

ANCOVA

For a reasonably sized RCT, the variance of $\hat{\beta}$ from equation (1), is approximately

$$\left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\} \sigma^2 (1 - \rho_{01}^2)$$

where ρ_{01} is the within subject correlation between baseline and follow-up measurements.

LDA model

The variance of this estimate $\hat{\delta}$ from equation (2) ($\sigma_0^2 = \sigma_1^2 = \sigma^2$) can be written as,

$$\left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\} 2\sigma^2 (1 - \rho_{01})$$

We could also estimate the treatment difference at follow-up (POST only) between the treatment arms which is $(\hat{\gamma} + \hat{\delta})$. The variance of this estimate $(\hat{\gamma} + \hat{\delta})$ is then

$$\left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\} 2\sigma_1^2$$

cLDA Model

The variance of this estimate $\hat{\delta}'$ from equation (3) is then (11, 12)

$$\left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\} \sigma_1^2 (1 - \rho_{01}^2)$$