SESUG Paper 123-2025

Writing ESTIMATE statements to compare continuous scores over time overall and by group

Alexis C. Wardell, Allison M. Deal, UNC Lineberger Comprehensive Cancer Center

ABSTRACT

Research studies often involve continuous outcomes, for example survey instrument scores measured repeatedly over time. We see this in our cancer research when patients complete surveys before, during, and after an intervention. Our motivating example evaluates an intervention vs control, within a randomized trial, and compares scores between groups at 3 timepoints. PROC GENMOD is used to fit marginal linear models with generalized estimating equation methods to account for potential clustering within patients and allow for missingness. The longitudinal model includes time, group (intervention v control), and the interaction between time and group. There are many interesting estimates from this type of study, including 1) outcome scores at each time point, 2) differences in scores between groups at different timepoints, 3) changes in scores over time within groups, and 4) differences in the slopes of scores over time by group. Additionally, consideration must be made in treating time as continuous or categorical. PROC GENMOD is a useful tool to make these comparisons, but special considerations must be made in order to write all the necessary ESTIMATE statements. While tools like LSMEANS are useful, understanding how to write ESTIMATE statements for comparisons is helpful for cleaning output and making comparisons that are unable to be made via LSMEANS. We lay out the underlying model structure and equations to clearly illustrate how to write these ESTIMATE statements.

1 INTRODUCTION

Many studies use a continuous outcome measure, for example a survey instrument with continuous scores, and repeat the measure over time. The goal of these studies is generally to examine how an intervention might change a continuous outcome (patient self-efficacy, anxiety, depression, lab values, etc.) over time, separately for intervention and control groups. Additionally, studies aim to examine if the outcomes change differently over time and if there are group differences at various timepoints. Our motivating example is a pilot randomized control trial which evaluated the Families Addressing Cancer Together (FACT) intervention[1]. FACT is a web-based, individually tailored, psychoeducational intervention for parents with cancer to improve illness-related communication with their minor children. The outcome measure was the Communication Self-Efficacy Scale (CSES), which had scores ranging from 0-100, and was asked at baseline (t0; 0 weeks), mid-intervention (t1; 3 weeks), and post-intervention (t2; 6 weeks). Higher scores reflected more confidence in communication self-efficacy.

For analyzing longitudinal data with repeated measures, generalized estimating equation (GEE) methods developed by Liang and Zeger, is a common method for appropriately handling correlated outcome data collected over time[2]. This method requires specification of a working correlation matrix, a link function, and a distributional assumption. Common working correlation matrices are independent, exchangeable, autoregressive, and unstructured and can be chosen based on what assumptions can be made about the correlations. Many link functions may also be specified, such as logistic, linear, log-linear, etc. depending on the type of outcome data. The focus for this is a marginal linear model with GEE methods, using a linear link function with a normal distributional assumption and an exchangeable correlation matrix. PROC GENMOD is a common SAS procedure that is utilized to build this model type[3]. A common issue with PROC GENMOD for investigators is the mystification of ESTIMATE statements. We aim to add to the existing literature with specific examples that add explanation about the underlying model to develop ESTIMATE statements to produce results of interest in clinical practice. While PROC GENMOD has many useful statements, such as LSMEANS, that are able to provide differences in least squares means, it is important to understand how to build ESTIMATE statements for ease of interpretation, clean output production, and estimating comparisons that aren't available via LSMEANS.

2 DATA STRUCTURE

The data structure for the motivating example is a long data form. There is a row for each repeated measure containing the patient ID (index), the group (FACT intervention vs. control), the timepoint (baseline = t0, mid-intervention = t1, and post-intervention = t2), and their CSES score.

See the sample data structure below:

Index	Group	Timepoint (t)	CSES
1	Intervention (FACT)	0	95.5
1	Intervention (FACT)	1	97.8
1	Intervention (FACT)	2	100
4	Control	0	69.4
4	Control	1	66.7
4	Control	2	76.1

2.1 MODEL TYPE AND TIME SPECIFICATION

Marginal models were used with generalized estimating equation methods to account for potential clustering within patients. A normal distribution with an identity link function was used to build the model. All longitudinal models included the following covariates: time (baseline (t0), mid-intervention (t1; 3 weeks), post-intervention (t2; 6 weeks)), the grouping variable (FACT intervention vs. control), and the interaction between time and the grouping variable. The time variable can be treated as continuous or categorical. It may be beneficial to treat time categorically if the outcome is not monotonically increasing or decreasing over time. The first step should be to graph the raw data and connect lines over time for each patient. If the graph suggests, for example, that scores dip or rise at t1 compared to t0 and t2, then treating time as categorical is likely more appropriate. If the graphs show fairly straight lines for each patient, treating time as continuous is likely appropriate.

3 CONTINUOUS TIME

When treating time as a continuous variable, the model can be written as:

$$g(\mu_{it}) = a + \beta_1 group + \beta_2 time + \beta_3 group * time$$

Where g(.) is the specified link function, μ_{it} is the mean CSES score at time t for subject i (a continuous score representing communication self-efficacy), α is the intercept, β_1 is the parameter estimate for the grouping variable which represents the difference in the FACT intervention vs. control at time 0, β_2 is the parameter estimate for the time variable representing the slope of the control group from time 0 to time 1, as well as time 1 to time 2, β_3 is the parameter estimate for the interaction term representing the difference in slope between the two study groups. This model framework can be used to check ESTIMATE statements and identify the needed parameters.

The following SAS code represents the model where index denotes the subject's identifier, and the REPEATED statement allows for GEE use. Reference coding was used for this study.

```
proc genmod data=dataset;
      class group(ref="Control (waitlist)") index;
      model outcome = group t group*t/ dist=normal link=identity;
      repeated subject=index / type=exch;
run;
```

The general form of the ESTIMATE statements is as follows, where terms are included and excluded based on comparisons of interest:

```
ESTIMATE "Description" int group t group*t;
```

When running the code above, the following output is produced detailing the parameter information and the parameter estimates (Figure 1 and Figure 2). Using the equation derived in the next section, the parameter estimates can be used to check the ESTIMATE statement results.

Figure 1. Parameter information detailing the order of the parameters in the model.

Parameter Information		
Parameter	**Effect**	**group**
Prm1	Intercept	
Prm2	group	Treatment (FACT)
Prm3	group	Control (waitlist)
Prm4	t	
Prm5	t*group	Treatment (FACT)
Prm6	t*group	Control (waitlist)

Figure 2. Parameter estimates after running the PROC GENMOD procedure.

Analysis Of GEE Parameter Estimates							
Empirical Standard Error Estimates							
Parameter		**Estimate**	**Standard Error**	**95% Confidence Limits**		**Z**	**Pr > |Z|**
Intercept		79.5292	3.1913	73.2743	85.7841	24.92	<.0001
group	Treatment (FACT)	1.2277	4.6461	-7.8784	10.3339	0.26	0.7916
group	Control (waitlist)	0.0000	0.0000	0.0000	0.0000	.	.
t		0.3362	1.5799	-2.7604	3.4327	0.21	0.8315
t*group	Treatment (FACT)	4.4158	2.2666	-0.0268	8.8583	1.95	0.0514
t*group	Control (waitlist)	0.0000	0.0000	0.0000	0.0000	.	.

3.1 COMPARISONS OF INTEREST

For a study with two groups and multiple timepoints, there are multiple comparisons one might be interested in. These are the typical research questions our investigators often want to answer:

1. What are the **model estimated** outcome measure scores at each time point for each group?
2. Are the outcome measure **scores different by group** at any timepoint?
3. Is there a significant change over time in outcome measure **within group** (separately for control and FACT intervention)?
4. Does the outcome measure **change differently** over time by group? (i.e. is there a difference in slopes between the FACT intervention and control groups)?

For each of these comparisons, we will outline the appropriate equations that can be written and the accompanying SAS code needed to form the corresponding ESTIMATE statements.

3.1.1 ESTIMATED SCORES AT EACH TIMEPOINT

CONTROL: Model estimates for control at time =0, time=1, and time=2

- Estimate for control at time=0; Let group = 0, time = 0, so group*time=0

$$
\begin{aligned}
&= a + \beta_1 0 + \beta_2 0 + \beta_3 0 * 0 \\
&= a \\
&= 79.5
\end{aligned}
$$

- Estimate for control at time=1; Let group = 0, time = 1, so group*time=0

$$
\begin{aligned}
&= a + \beta_1 0 + \beta_2 1 + \beta_3 0 * 0 \\
&= a + \beta_2 \\
&= 79.5 + 0.3 \\
&= 79.8
\end{aligned}
$$

- Estimate for control at time=2; Let group = 0, time = 2, so group*time=0

$$
\begin{aligned}
&= a + \beta_1 0 + \beta_2 2 + \beta_3 0 * 0 \\
&= a + 2\beta_2 \\
&= 79.5 + 2(0.3) \\
&= 80.1
\end{aligned}
$$

The necessary ESTIMATE statements are:

```
ESTIMATE "model est control at t0" int 1 group 0 1 t 0 group*t 0 0;
ESTIMATE "model est control at t1" int 1 group 0 1 t 1 group*t 0 1;
ESTIMATE "model est control at t2" int 1 group 0 1 t 2 group*t 0 2;
```

To write the statement for control at time 0, we include the intercept, time, group, and the group*time interaction term in the ESTIMATE statements. The intercept is included as ‘int 1’, derived directly from the linear combinations in the equations above. In Figure 1, the parameter information is ordered, where the control group (reference) is listed second as Prm3 in comparison to Prm2 representing the FACT intervention group. Therefore, to indicate the control group, we use the second position of the group section, giving ‘group 0 1.’ Using continuous time, time will take the value of 0, resulting in ‘t 0’ as the required syntax. For the interaction term, insert the continuous time value into the second position to achieve ‘group*t 0 0’. The ESTIMATE statements for the control group at time 1 and 2 follow the same logic. When checking, we see that the ESTIMATE output in Figure 3 matches the calculations from the formulation above using the parameter estimates in Figure 2.

INTERVENTION: Model estimates for FACT intervention at time =0, time=1, and time=2

- Estimate for FACT at time=0; Let group = 1, time = 0, so group*time=0

$$
\begin{aligned}
&= a + \beta_1 1 + \beta_2 0 + \beta_3 0 * 0 \\
&= a + \beta_1 \\
&= 79.5 + 1.2 \\
&= 80.7
\end{aligned}
$$

- Estimate for FACT at time=1; Let group = 1, time = 1, so group*time=1

$$
\begin{aligned}
&= a + \beta_1 1 + \beta_2 1 + \beta_3 1 * 1 \\
&= a + \beta_1 + \beta_2 + \beta_3 \\
&= 79.5 + 1.2 + 0.3 + 4.4 \\
&= 85.4
\end{aligned}
$$

- Estimate for FACT at time=2; Let group = 1, time = 2, so group*time=2

$$
\begin{aligned}
&= a + \beta_1 1 + \beta_2 2 + \beta_3 1 * 2 \\
&= a + \beta_1 + 2\beta_2 + 2\beta_3 \\
&= 79.5 + 1.2 + 2(0.3) + 2(4.4) \\
&= 90.1
\end{aligned}
$$

The necessary ESTIMATE statements are as follows:

```
ESTIMATE "model est FACT at t0" int 1 group 1 0 t 0 group*t 0 0;
ESTIMATE "model est FACT at t1" int 1 group 1 0 t 1 group*t 1 0;
```

```
ESTIMATE "model est FACT at t2" int 1 group 1 0 t 2 group*t 2 0;
```

To write the statement for the FACT intervention at time 0, we include the intercept, time, group, and the group*time interaction term in the ESTIMATE statements. The intercept is included as ‘int 1’, derived directly from the linear combinations in the equations above. In Figure 1, the parameter information is ordered, where the FACT intervention group is listed first as Prm2, in comparison to Prm3 representing the control group (reference). Therefore, to indicate the FACT intervention group, we use the first position of the group section, giving ‘group 1 0.’ Using continuous time, time will take the value of 0, resulting in ‘t 0’ as the required syntax. For the interaction term, insert the continuous time value into the first position to achieve ‘group*t 0 0’. The ESTIMATE statements for the intervention group at time 1 and 2 follow the same logic. When checking, we see that the ESTIMATE output in Figure 3 matches the calculations from the formulation above using the parameter estimates in Figure 2.

3.1.2 DIFFERENCES IN SCORES AT EACH TIMEPOINT

Are the outcome measure **scores different by group** at any time (for example, at time 2)?

- Let group = 1, time = 2, so group*time=2

$$= a + \beta_1 1 + \beta_2 2 + \beta_3 1 * 2$$
$$= a + \beta_1 + 2\beta_2 + 2\beta_3$$
$$= 79.5 + 1.2 + 2(0.3) + 2(4.4)$$

- Let group = 0, time = 2, so group*time=0

$$= a + \beta_1 0 + \beta_2 2 + \beta_3 0 * 0$$
$$= a + 2\beta_2$$
$$= 79.5 + 2(0.3)$$

To find the difference between these two groups we subtract:

$$= (a + \beta_1 + 2\beta_2 + 2\beta_3) - (a + 2\beta_2)$$
$$= \beta_1 + 2\beta_3$$
$$= 1.2 + 2(4.4)$$
$$= 10.0$$

$\beta_1 + 2\beta_3$ represents the difference between FACT at time=2 and control at time=2

The necessary ESTIMATE statement is:

```
ESTIMATE "Diff FACT vs. control at t2" group 1 -1 group*t 2 -2;
```

To write the statement for the difference in average CSES between the FACT intervention and control groups at time 2, we include the group, and the group*time interaction term in the ESTIMATE statements. Note that β_1 and β_3 correspond to these terms in the model. The term for the intercept and time are not needed because they drop out of the model. In Figure 1, the parameter information is ordered, where the FACT intervention group is listed first as Prm2, in comparison to Prm3 representing the control group (reference). In SAS, when a 1:1 comparison is made between two groups, the syntax is 1 for the main group and -1 for the comparator group, thus ‘group 1 -1’ is utilized to represent the comparison between the FACT intervention group and control group. Similarly, the interaction term follows the same syntax indicating a comparison, but time is multiplied through the interaction term which gives ‘group*t 2 -2’. When checking, we see that the ESTIMATE output in Figure 3 matches the calculations from the formulation above using the parameter estimates in Figure 2.

Similarly, the ESTIMATE statements for time 0 and time 1 are achieved:

```
ESTIMATE "Diff FACT vs. control at t0" group 1 -1;
ESTIMATE "Diff FACT vs. control at t1" group 1 -1 group*t 1 -1;
```

3.1.3 CHANGE OVER TIME WITHIN GROUP

Since we are treating time as continuous, the model is fitting a line with a single slope for each group.

CONTROL: For the control group, group=0. We want to compare the change from t0 to t1.

- Let group = 0 and time = 1, so group*time=0

$$= a + \beta_1 0 + \beta_2 1 + \beta_3 0 * 1 = a + \beta_2$$
$$= 79.5 + 0.3$$

- Let group = 0 and time = 0, so group*time=0

$$= a + \beta_1 0 + \beta_2 0 + \beta_3 0 * 0$$
$$= a$$
$$= 79.5$$

To find the difference between these two groups we subtract:

$$= (a + \beta_2) - (a) = \beta_2$$
$$= 0.3$$

β_2 represents the difference between control at time=1 and control at time=0. (i.e. slope for control)
The necessary ESTIMATE statement is:

```
ESTIMATE "slope for control" t 1 group*t 0 1;
```

To write the statement for the change over time in average CSES in the control group between time 0 and time 1, we include the time and the group*time interaction term in the ESTIMATE statements. In Figure 1, the parameter information is ordered, where the FACT intervention group is listed first as Prm2, in comparison to Prm3 representing the control group (reference). Using continuous time, time will take the value of 1, resulting in 't 1' as the required syntax since we are comparing time 1 to time 0. For the interaction term, insert the continuous time value into the second position to achieve 'group*t 0 1'. The ESTIMATE statement value for the slope of the control group from time 0 and time 1 will be equivalent to that of the slope of the control group from time 1 to time 2, given the linear nature of the model with a constant rate of change. When checking, we see that the ESTIMATE output in Figure 3 matches the calculations from the formulation above using the parameter estimates in Figure 2.

INTERVENTION: For the FACT intervention group, group=1, and we want to compare the change from t0 to t1.

- Let group = 1, time = 1, so group*time=1

$$= a + \beta_1 1 + \beta_2 1 + \beta_3 1 * 1$$
$$= a + \beta_1 + \beta_2 + \beta_3$$
$$= 79.5 + 1.2 + 0.3 + 4.4$$

- Let group = 1, time = 0, so group*time=0

$$= a + \beta_1 1 + \beta_2 0 + \beta_3 0 * 0$$
$$= a + \beta_1$$
$$= 79.5 + 1.22$$

To find the difference between these two groups we subtract:

$$= (a + \beta_1 + \beta_2 + \beta_3) - (a + \beta_1)$$
$$= \beta_2 + \beta_3$$
$$= 0.3 + 4.4$$
$$= 4.7$$

$\beta_2 + \beta_3$ represents the difference between FACT at time=1 and FACT at time=0. (i.e. slope for FACT)
The necessary ESTIMATE statement is:

```
ESTIMATE "slope for FACT" t 1 group*t 1 0;
```

To write the statement for the change over time in average CSES in the FACT intervention group between time 0 and time 1, we include the time and the group*time interaction term in the ESTIMATE statements. This is formed in the same way as that of the control group, except the interaction term differs. For the interaction term, insert the continuous time value into the first position to achieve ‘group*t 1 0’. When checking, we see that the ESTIMATE output in Figure 3 matches the calculations from the formulation above using the parameter estimates in Figure 2.

3.1.4 CHANGE DIFFERENTLY BY GROUP

To find the difference in slopes between the FACT intervention and control groups, consider the above slopes for the control and intervention groups separately and subtract them. The difference in slopes between the FACT intervention and control groups can be derived:

$$= (\beta_2 + \beta_3) - \beta_2$$

$$= \beta_3$$

$$= 4.4$$

The necessary ESTIMATE statement is:

```
ESTIMATE "diff in slope" group * t 1 -1;
```

To write the statement for the difference in slopes between the FACT intervention and control groups, the interaction term is involved. When comparing two groups in SAS ESTIMATE statements 1 and -1 are used. When checking, we see that the ESTIMATE output in Figure 3 matches the calculations from the formulation above using the parameter estimates in Figure 2.

3.2 SUMMARY OF RESULTS OUTPUT

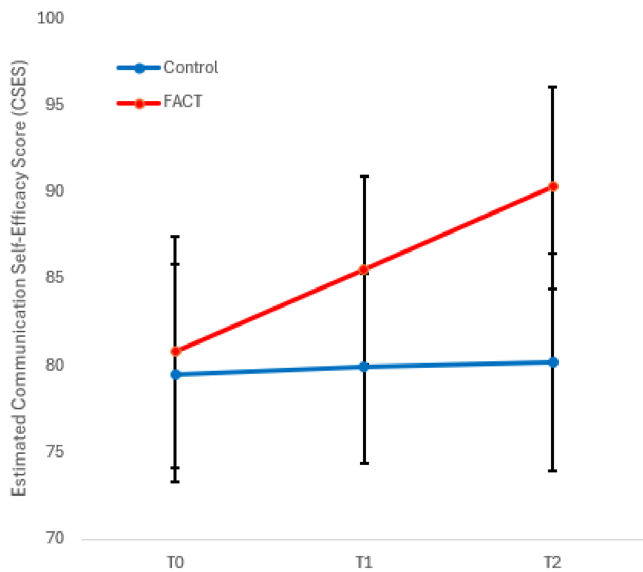
Figure 3. ESTIMATE statement results.

Contrast Estimate Results

Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
model est Control at t0	79.5292	73.2743	85.7841	79.5292	3.1913	0.05	73.2743	85.7841	621.02	<.0001
model est Control at t1	79.8653	74.3966	85.3341	79.8653	2.7902	0.05	74.3966	85.3341	819.29	<.0001
model est Control at t2	80.2015	73.8875	86.5155	80.2015	3.2215	0.05	73.8875	86.5155	619.80	<.0001
model est FACT at t0	80.7569	74.1389	87.3750	80.7569	3.3766	0.05	74.1389	87.3750	572.00	<.0001
model est FACT at t1	85.5088	80.1358	90.8819	85.5088	2.7414	0.05	80.1358	90.8819	972.91	<.0001
model est FACT at t2	90.2607	84.4096	96.1119	90.2607	2.9854	0.05	84.4096	96.1119	914.12	<.0001
Diff FACT vs. control at t0	1.2277	-7.8784	10.3339	1.2277	4.6461	0.05	-7.8784	10.3339	0.07	0.7916
Diff FACT vs. control at t1	5.6435	-2.0231	13.3101	5.6435	3.9116	0.05	-2.0231	13.3101	2.08	0.1491
Diff FACT vs. control at t2	10.0593	1.4509	18.6676	10.0593	4.3921	0.05	1.4509	18.6676	5.25	0.0220
slope for Control	0.3362	-2.7604	3.4327	0.3362	1.5799	0.05	-2.7604	3.4327	0.05	0.8315
slope for FACT	4.7519	1.5664	7.9374	4.7519	1.6253	0.05	1.5664	7.9374	8.55	0.0035
diff in slope	4.4158	-0.0268	8.8583	4.4158	2.2666	0.05	-0.0268	8.8583	3.80	0.0514

Using the model-based estimates at each timepoint, along with the confidence intervals, a summary graphic can be made to show summary changes over time (Figure 4).

Figure 4. ESTIMATE statement results for mean CSES over time in cancer patients receiving FACT intervention vs. the control.



4 CATEGORICAL TIME

When treating time as a categorical variable, the model can be written as:

$$g(\mu_{it}) = a + \beta_1 group + \beta_2 time_1 + \beta_3 time_2 + \beta_4 group_0 * time_1 + \beta_5 group_0 * time_2 + \beta_6 group_1 * time_0 + \beta_7 group_1 * time_1 + \beta_8 group_1 * time_2$$

Where g(.) is the specified link function, μ_{it} is the mean CSES at time t for subject i. The following SAS code represents the model, where reference level coding is utilized:

```
proc genmod data=dataset;
            class group(ref="Control (waitlist)") t(ref="0") index;
            model outcome = group t group*t/ dist=normal link=identity;
            repeated subject=index / type=exch;
run;
```

When running the code above, the following output is produced detailing the parameter information and the parameter estimates (Figure 5 and Figure 6). The only item that differs in the PROC GENMOD statement from treating time as continuous is adding the time variable to the class statement. Using the equation derived in the next section, the parameter estimates can be used to check the ESTIMATE statement results.

Figure 5. Parameter information for model with categorical time.

Parameter Information			
Parameter	**Effect**	**group**	**t**
Prm1	Intercept		
Prm2	group	Treatment (FACT)	
Prm3	group	Control (waitlist)	
Prm4	t		1
Prm5	t		2
Prm6	t		0
Prm7	group*t	Treatment (FACT)	1
Prm8	group*t	Treatment (FACT)	2
Prm9	group*t	Treatment (FACT)	0
Prm10	group*t	Control (waitlist)	1
Prm11	group*t	Control (waitlist)	2
Prm12	group*t	Control (waitlist)	0

Figure 6. Parameter estimates after running the PROC GENMOD procedure with categorical time.

Analysis Of GEE Parameter Estimates Empirical Standard Error Estimates					95% Confidence Limits			
Parameter			**Estimate**	**Standard Error**			**Z**	**Pr > |Z|**
Intercept			79.1187	3.3201	72.6114	85.6260	23.83	<.0001
group	Treatment (FACT)		1.3119	4.8867	-8.2658	10.8895	0.27	0.7883
group	Control (waitlist)		0.0000	0.0000	0.0000	0.0000	.	.
t	1		1.8928	1.9770	-1.9821	5.7676	0.96	0.3384
t	2		0.5292	3.1727	-5.6892	6.7477	0.17	0.8675
t	0		0.0000	0.0000	0.0000	0.0000	.	.
group*t	Treatment (FACT)	1	3.9663	3.5357	-2.9634	10.8961	1.12	0.2619
group*t	Treatment (FACT)	2	8.9695	4.5409	0.0694	17.8696	1.98	0.0482
group*t	Treatment (FACT)	0	0.0000	0.0000	0.0000	0.0000	.	.
group*t	Control (waitlist)	1	0.0000	0.0000	0.0000	0.0000	.	.
group*t	Control (waitlist)	2	0.0000	0.0000	0.0000	0.0000	.	.
group*t	Control (waitlist)	0	0.0000	0.0000	0.0000	0.0000	.	.

4.1 COMPARISONS OF INTEREST

We can address the same four main research questions for this scenario via ESTIMATE statements: 1) What are the model estimated outcome measure scores at each time point for each group?; 2) Are the outcome measure scores different by group at any timepoint?; 3) Is there a significant change over time in outcome measure within group (separately for control and FACT intervention)?; and 4) Does the outcome measure change differently over time by group? (i.e. is there a difference in slopes between the FACT intervention and control groups?).

4.1.1 ESTIMATED SCORES AT EACH TIMEPOINT

CONTROL: Model estimates for the mean CSES score for the control group at each time point
Estimate for control at time=0; Let group = 0, time = 0

$$g(\mu_{it}) = a + \beta_1(0) + \beta_2(0) + \beta_3(0) + \beta_4(0)(0) + \beta_5(0)(0) + \beta_6(0)(0) + \beta_7(0)(0) + \beta_8(0)(0)$$
$$= a$$
$$= 79.1$$

The ESTIMATE statement is:

```
ESTIMATE "model est control at t0" int 1  group 0 1  t 0 0 1
  group*t 0 0 0 0 0 1;
```

To write the statement for the control group at time 0, we include the intercept, time, group, and the group*time interaction term in the ESTIMATE statements. The intercept is included as 'int 1', derived directly from the linear combinations in the equations above. In Figure 5, the parameter information is ordered, where the FACT intervention group is listed first as Prm2, in comparison to Prm3 representing the control group (reference). Therefore, to indicate the control group, we use the second position of the group section, giving 'group 0 1.' Similarly, the parameter order for the time variable is important for formulating the ESTIMATE statement such that Prm4 represents time 1, Prm5 represents time 2, and Prm6 represents t0, thus implying 't 0 0 1'. The interaction term, group*t, has 6 positions to fill in the ESTIMATE statement. This is achieved by the 2 study groups (FACT intervention vs. control) multiplied by 3 time points (t0, t1, t2). Based on Figure 5, Prm12, the last of the 6 interaction parameters, represents the interaction for the control group*t0, thus 'group*t 0 0 0 0 0 1'. When checking, we see that the ESTIMATE output in Figure 8 matches the calculations from the formulation above using the parameter estimates in Figure 6.

The following ESTIMATE statements extend the example to obtain model estimates for the control group at times 1 and 2:

```
ESTIMATE "model est control at t1" int 1  group 0 1  t 1 0 0
  group*t 0 0 0 1 0 0;
ESTIMATE "model est control at t2" int 1  group 0 1  t 0 1 0
  group*t 0 0 0 0 1 0;
```

INTERVENTION: Model estimates for the mean CSES score for the FACT intervention group at each time point:

Estimate for FACT intervention group at time=1; Let group = 1, time = 1

$$g(\mu_{it}) = a + \beta_1(1) + \beta_2(1) + \beta_3(0) + \beta_4(0) * (1) + \beta_5(0)(0) + \beta_6(1)(0) + \beta_7(1)(1) + \beta_8(1)(0)$$
$$= a + \beta_1 + \beta_2 + \beta_7$$
$$= 79.1 + 1.3 + 1.9 + 4.0$$
$$= 86.3$$

The ESTIMATE statement is:

```
ESTIMATE "model est FACT at t1" int 1  group 1 0  t 1 0 0
  group*t 1 0 0 0 0 0;
```

To write the statement for the FACT intervention at time 1, we include the intercept, time, group, and the group*time interaction term in the ESTIMATE statements. The intercept is included as 'int 1', derived directly from the linear combinations in the equations above. In Figure 5, the parameter information is ordered, where the FACT intervention group is listed first as Prm2, in comparison to Prm3 representing the control group (reference). Therefore, to indicate the FACT intervention group, we use the first position of the group section, giving 'group 1 0.' Similarly, the parameter order for the time variable is important for formulating the ESTIMATE statement such that Prm4 represents time 1, Prm5 represents time 2, and Prm6 represents t0, thus implying 't 1 0 0'. The interaction term, group*t, has 6 positions to fill in the ESTIMATE statement. Based on Figure 5, Prm7, is the relevant interaction parameter representing the interaction for the FACT intervention group*t1, thus 'group*t 1 0 0 0 0 0'. When checking, we see that the ESTIMATE output in Figure 8 matches the calculations from the formulation above using the parameter estimates in Figure 6.

A helpful tool is the /e tool on the ESTIMATE statement to check coefficients for parameters. The following code produces the output below (Figure 7):

```
ESTIMATE "model est FACT at t1" int 1  group 1 0  t 1 0 0
  group*t 1 0 0 0 0 0/e;
```

Figure 7. Coefficients for ESTIMATE statements.

Coefficients for Contrast model est Control at t0												
Label	Prm1	Prm2	Prm3	Prm4	Prm5	Prm6	Prm7	Prm8	Prm9	Prm10	Prm11	Prm12
model est Control at t0	1	0	1	0	0	1	0	0	0	0	0	1

The following ESTIMATE statements extend the example to obtain model estimates for the FACT intervention group at each time 0 and 2:

```
ESTIMATE "model est FACT at t0" int 1  group 1 0  t 0 0 1
  group*t 0 0 1 0 0 0;
ESTIMATE "model est FACT at t2" int 1  group 1 0  t 0 1 0
  group*t 0 1 0 0 0 0;
```

4.1.2 DIFFERENCES IN SCORES AT EACH TIMEPOINT

Are the outcome measure **scores different by group** at any time (for example, at time 2)?

- Let group = 1, time = 2

$$g(\mu_{it}) = a + \beta_1(1) + \beta_2(0) + \beta_3(1) + \beta_4(0)(0) + \beta_5(0)(1) + \beta_6(1)(0) + \beta_7(1)(0) + \beta_8(1)(1)$$
$$= a + \beta_1 + \beta_3 + \beta_8$$
$$= 79.1 + 1.3 + 0.5 + 9.0$$

- Let group = 0, time = 2

$$g(\mu_{it}) = a + \beta_1(0) + \beta_2(0) + \beta_3(1) + \beta_4(1)(0) + \beta_5(1)(1) + \beta_6(0)(0) + \beta_7(0)(0) + \beta_8(0)(1)$$
$$= a + \beta_3 + \beta_5$$
$$= 79.1 + 0.5 + 0$$

To find the difference between these two groups we subtract:

$$= (a + \beta_1 + \beta_3 + \beta_8) - (a + \beta_3 + \beta_5)$$
$$= \beta_1 + \beta_8 - \beta_5$$
$$= 1.3 + 9.0 + 0$$
$$= 10.3$$

$\beta_1 + \beta_8 - \beta_5$ represents the difference between FACT at time=2 and control at time=2

The necessary ESTIMATE statement is:

```
ESTIMATE "Diff FACT v Control at t2" group 1 -1 group*t 0 1 0 0 -1 0;
```

To write the statement for the difference in average CSES between the FACT intervention and control groups at time 2, we include the group, and the group*time interaction term in the ESTIMATE statements. Note that $\beta_1 + \beta_8 - \beta_5$ correspond to these terms in the model. In Figure 5, the parameter information is ordered, where the FACT intervention group is listed first as Prm2, in comparison to Prm3 representing the control group (reference). In SAS, when a 1:1 comparison is made between two groups, the syntax is 1 for the main group and -1 for the comparator group, thus 'group 1 -1' is utilized to represent the comparison between the FACT intervention group and control group. Similarly, the interaction term follows the same syntax indicating a comparison, Based on Figure 5, Prm8 and Prm11 are the relevant interaction parameters, thus 'group*t 0 1 0 0 -1 0'. When checking, we see that the ESTIMATE output in Figure 8 matches the calculations from the formulation above using the parameter estimates in Figure 6.

While these differences can also be achieved with the LSMEANS statement, `lsmeans group*t / cl diff;`, the comparisons are produced for all combinations of group and time, thus cluttering the output with comparisons that are irrelevant to compare in clinical practice. It is useful to be able to logically write the estimate statements to simplify the output produced by the GEMOD procedure.

The following ESTIMATE statements extend the example to obtain model estimates for the difference in mean CSES estimate for the FACT intervention group vs. the control group at time 0 and time 1.

```
ESTIMATE "Diff FACT vs. Control at t0" group 1 -1 group*t 0 0 1 0 0 -1;
ESTIMATE "Diff FACT vs. Control at t1" group 1 -1 group*t 1 0 0 -1 0 0;
```

4.1.3 CHANGE OVER TIME (BETWEEN CERTAIN TIMEPOINTS) WITHIN GROUP

CONTROL: For the control group, group=0. We want to compare the change from t0 to t1. From the model above, we can write the following:

- Let group = 0 and time = 1

$$= a + \beta_2$$
$$= 79.1 + 1.9$$

- Let group = 0 and time = 0

$$= a$$
$$= 79.1$$

To find the difference between these two groups we subtract:

$$= (a + \beta_2) - (a) = \beta_2$$
$$= 1.9$$

β_2 represents the difference between control at time=1 and control at time=0. (i.e. slope for control)
The necessary ESTIMATE statement is:

```
ESTIMATE 'slope for Control t0-t1' t 1 0 -1 group*t 0 0 0 1 0 -1;
```

To write the statement for the change over time in average CSES in the control group between time 0 and time 1, we include the time and the group*time interaction term in the ESTIMATE statements. In Figure 5, the parameter information is ordered, where the FACT intervention group is listed first as Prm2, in comparison to Prm3 representing the control group (reference). Similarly, the parameter order for the time variable is Prm4(time 1), Prm5(time 2), and Prm6(time 0). Since a comparison is being made between time 0 and time 1, the appropriate syntax is 't 1 0 -1'. For the interaction terms, insert the relevant parameters are Prm10 and Prm12, to achieve 'group*t 0 0 0 1 0 -1'. When checking, we see that the ESTIMATE output in Figure 8 matches the calculations from the formulation above using the parameter estimates in Figure 6.

Similarly, for the difference between control at time=2 and time=1

```
ESTIMATE 'slope for Control t1-t2' t -1 1 0 group*t 0 0 0 -1 1 0;
```

INTERVENTION: For the FACT intervention group, group=1, and we want to compare the change from t0 to t1.

- Let group = 1, time = 1

$$= a + \beta_1 + \beta_2 + \beta_7$$
$$= 79.1 + 1.3 + 1.9 + 4.0$$

- Let group = 1, time = 0

$$= a + \beta_1 + \beta_6$$
$$= 79.1 + 1.3 + 0$$

To find the difference between these two groups we subtract:

$$= (a + \beta_1 + \beta_2 + \beta_7) - (a + \beta_1 + \beta_6)$$
$$= \beta_2 + \beta_7 - \beta_6$$
$$= 1.9 + 4.0 - 0$$
$$= 5.9$$

$\beta_2 + \beta_7 - \beta_6$ represents the difference between FACT at time=1 and FACT at time=0. (i.e. slope for FACT)

The necessary ESTIMATE statement is:

```
ESTIMATE 'slope for FACT t0-t1' t 1 0 -1 group*t 1 0 -1 0 0 0;
```

To write the statement for the change over time in average CSES in the FACT intervention group between time 0 and time 1, we include the time and the group*time interaction term in the ESTIMATE statements. This is formed in the same way as that of the control group, except the interaction term differs. For the interaction term, insert the appropriate interaction parameters (Prm7 and Prm9) to achieve 'group*t 1 0 -1 0 0 0'. When checking, we see that the ESTIMATE output in Figure 8 matches the calculations from the formulation above using the parameter estimates in Figure 6.

Similarly, for the difference between the FACT intervention at time 2 and time 1:

```
ESTIMATE 'slope for FACT t1-t2' t -1 1 0 group*t -1 1 0 0 0 0;
```

4.1.4 CHANGE DIFFERENTLY BY GROUP

To find the difference in slopes between the FACT intervention and control groups, consider the above slopes for the control and intervention groups separately at time 0 and time 1 and subtract them. The difference in slopes between the FACT intervention and control groups can be derived.

$$(\beta_2 + \beta_7 - \beta_6) - \beta_2$$
$$= \beta_7 - \beta_6$$
$$= 4.0$$

The necessary ESTIMATE statement is:

```
ESTIMATE "diff in slope t0-t1" group*t 1 0 -1 -1 0 1;
```

To write the statement for the difference in slopes between the FACT intervention and control groups, the interaction term is involved. When comparing two groups in SAS ESTIMATE statements 1 and -1 are used. Since we are making two inherent comparisons and finding the difference of slopes, the interaction parameters for both the comparison between outcome scores for control and the comparison between outcome scores for the FACT intervention from time 0 to time 1 need to be included as comparisons in the ESTIMATE statement. This results in utilizing Prm7, Prm 9, Prm10, and Prm12 from Figure 5 to give group*t '1 0 -1 -1 0 1'. When checking, we see that the ESTIMATE output in Figure 8 matches the calculations from the formulation above using the parameter estimates in Figure 6.

Similarly, the difference in slopes between the FACT intervention and control groups at time 1 and time 2 is as follows:

```
ESTIMATE "diff in slope t1-t2" group*t -1 1 0 1 -1 0;
```

4.2 SUMMARY OF RESULTS OUTPUT

For all examples above involving categorical time, the ESTIMATE output in Figure 8 matches the calculations from the formulas above using the parameter estimates in Figure 6. Note that all model estimates are similar to estimates in section 3 when treating time as continuous.

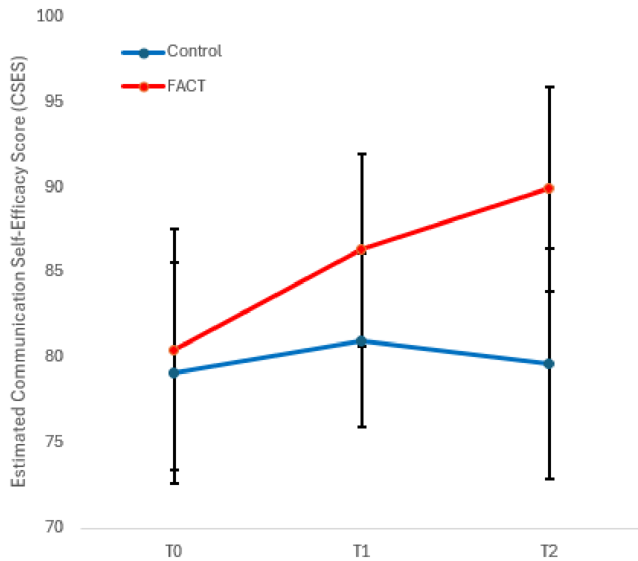
Figure 8. ESTIMATE statement results for categorical time

Contrast Estimate Results

Label	Mean Estimate	Mean Confidence Limits		L'Beta Estimate	Standard Error	Alpha	L'Beta Confidence Limits		Chi-Square	Pr > ChiSq
model est Control at t0	79.1187	72.6114	85.6260	79.1187	3.3201	0.05	72.6114	85.6260	567.87	<.0001
model est Control at t1	81.0115	75.8794	86.1435	81.0115	2.6184	0.05	75.8794	86.1435	957.20	<.0001
model est Control at t2	79.6480	72.9239	86.3720	79.6480	3.4307	0.05	72.9239	86.3720	538.99	<.0001
model est FACT at t0	80.4306	73.4030	87.4581	80.4306	3.5856	0.05	73.4030	87.4581	503.19	<.0001
model est FACT at t1	86.2896	80.6607	91.9185	86.2896	2.8719	0.05	80.6607	91.9185	902.74	<.0001
model est FACT at t2	89.9293	83.9252	95.9335	89.9293	3.0634	0.05	83.9252	95.9335	861.78	<.0001
Diff FACT vs. Control at t0	1.3119	-8.2658	10.8895	1.3119	4.8867	0.05	-8.2658	10.8895	0.07	0.7883
Diff FACT vs. Control at t1	5.2782	-2.3391	12.8954	5.2782	3.8864	0.05	-2.3391	12.8954	1.84	0.1744
Diff FACT vs. Control at t2	10.2814	1.2668	19.2959	10.2814	4.5994	0.05	1.2668	19.2959	5.00	0.0254
slope for Control t0-t1	1.8928	-1.9821	5.7676	1.8928	1.9770	0.05	-1.9821	5.7676	0.92	0.3384
slope for FACT t0-t1	5.8591	0.1139	11.6043	5.8591	2.9313	0.05	0.1139	11.6043	4.00	0.0456
slope for Control t1-t2	-1.3635	-5.7673	3.0403	-1.3635	2.2469	0.05	-5.7673	3.0403	0.37	0.5440
slope for FACT t1-t2	3.6397	-0.8976	8.1770	3.6397	2.3150	0.05	-0.8976	8.1770	2.47	0.1159
diff in slope t0-t1	3.9663	-2.9634	10.8961	3.9663	3.5357	0.05	-2.9634	10.8961	1.26	0.2619
diff in slope t1-t2	5.0032	-1.3198	11.3262	5.0032	3.2261	0.05	-1.3198	11.3262	2.41	0.1209

Using the model-based estimates at each timepoint, along with the confidence intervals, a summary graphic can be made to show summary changes over time (Figure 9).

Figure 9. ESTIMATE statement results for mean CSES over time in cancer patients receiving FACT intervention vs. the control (categorical time).



5 CONCLUSION

Often in cancer research, continuous outcomes, such as survey instrument scores, are measured repeatedly over time. A common method for appropriately handling correlated outcome data collected over time is the usage of marginal (linear) models, with generalized estimating equation (GEE) methods. This is a powerful tool; however, the ESTIMATE statements utilized to discern model estimated outcome scores are not easily understood. By using the underlying model structure and equations to stencil the model parameters and double check estimates produced, the ESTIMATE statements are demystified and more clearly illustrated. This allows for more useful and straightforward output.

6 REFERENCES

1. Nakamura, Z.M., Deal, A.M., Yopp, J.M., Wardell, A.C., Manning, M., Pak, P., Cassidy, A., Hanson, L.C., Jung, A., Song, M.-K., Valle, C.G., Walker, C., Won, H., Park, E.M. and Rosenstein, D.L. (2025), A Pilot Randomized Controlled Trial of Families Addressing Cancer Together for Parents With Cancer: Feasibility, Acceptability, and Preliminary Effects. Psycho-Oncology, 34: e70072. https://doi.org/10.1002/pon.70072

2. Zeger SL, Liang KY. Longitudinal data analysis using generalized linear models. Biometrica. 1986;73:13-22.

3. SAS Institute Inc. 2016. SAS/STAT® 14.2 User's Guide. Cary, NC: SAS Institute Inc.

7 CONTACT INFORMATION

Your comments and questions are valued and encouraged. We have a dataset and code that we can provide to be used to replicate the output in this paper.

Contact the authors at:

Alexis Wardell
Biostatistician
UNC Lineberger Comprehensive Cancer Center
ally_wardell@med.unc.edu

Allison Deal
Senior Biostatistician
UNC Lineberger Comprehensive Cancer Center
Allison_Deal@med.unc.edu

8 APPENDICES

8.1 GENMOD FOR CONTINUOUS WITH ALL ESTIMATE STATEMENTS

```
proc genmod data=dataset;
          class group(ref="Control (waitlist)") index;
          model outcome = group t group*t/ dist=normal link=identity;
          repeated  subject=index / type=exch;

          /*1*/
          ESTIMATE "model est control at t0"  int 1  group 0 1  t 0
                group*t 0 0;
          ESTIMATE "model est control at t1"  int 1  group 0 1  t 1
                group*t 0 1;
          ESTIMATE "model est control at t2"  int 1  group 0 1  t 2
                group*t 0 2;
          ESTIMATE "model est FACT at t0"     int 1  group 1 0  t 0
                group*t 0 0;
          ESTIMATE "model est FACT at t1"     int 1  group 1 0  t 1
                group*t 1 0;
          ESTIMATE "model est FACT at t2"     int 1  group 1 0  t 2
                group*t 2 0;

          /*2*/
          ESTIMATE "Diff FACT vs. control at t0" group 1 -1 ;
          ESTIMATE "Diff FACT vs. control at t1" group 1 -1 group*t 1 -1;
          ESTIMATE "Diff FACT vs. control at t2" group 1 -1 group*t 2 -2;

          /*3*/
          ESTIMATE "slope for Control"  t 1 group*t  0 1 ;
          ESTIMATE "slope for FACT"     t 1 group*t  1 0 ;

          /*4*/
          ESTIMATE "diff in slope" group*t  1 -1 ;
    run;
```

8.2 GENMOD FOR CATEGORICAL

```
proc genmod data=dataset;
        class group(ref="Control (waitlist)") t(ref="0") index;
        model outcome = group t group*t/ dist=normal link=identity;
        repeated subject=index / type=exch;

        /*1*/
        ESTIMATE "model est Control at t0"  int 1  group 0 1  t 0 0 1
             group*t 0 0 0 0 0 1;
        ESTIMATE "model est Control at t1"  int 1  group 0 1  t 1 0 0
             group*t 0 0 0 1 0 0;
        ESTIMATE "model est Control at t2"  int 1  group 0 1  t 0 1 0
             group*t 0 0 0 0 1 0;
        ESTIMATE "model est FACT at t0"     int 1  group 1 0  t 0 0 1
             group*t 0 0 1 0 0 0;
        ESTIMATE "model est FACT at t1"     int 1  group 1 0  t 1 0 0
             group*t 1 0 0 0 0 0;
        ESTIMATE "model est FACT at t2"     int 1  group 1 0  t 0 1 0
             group*t 0 1 0 0 0 0;

        /*2*/
        ESTIMATE "Diff FACT vs. Control at t0" group 1 -1
             group*t 0 0 1 0 0 -1 ;
        ESTIMATE "Diff FACT vs. Control at t1" group 1 -1
             group*t 1 0 0 -1 0 0;
        ESTIMATE "Diff FACT vs. Control at t2" group 1 -1
             group*t 0 1 0 0 -1 0;

        /*3*/
        ESTIMATE 'slope for Control t0-t1' t 1 0 -1 group*t 0 0 0 1 0 -1;
        ESTIMATE 'slope for FACT t0-t1'    t 1 0 -1 group*t 1 0 -1 0 0 0;
        ESTIMATE 'slope for Control t1-t2' t -1 1 0 group*t 0 0 0 -1 1 0;
        ESTIMATE 'slope for FACT t1-t2'    t -1 1 0 group*t -1 1 0 0 0 0;

        /*4*/
        ESTIMATE "diff in slope t0-t1" group*t  1 0 -1 -1 0 1;
        ESTIMATE "diff in slope t1-t2" group*t  -1 1 0 1 -1 0;
    run;
```