Covariance Analysis for the 2010 CCNM Pilot Study on Irritable Bowel Syndrome

By Shintaro Hagiwara and Patrick Boily
Centre for Quantitative Analysis and Decision Support, Carleton University

August 19, 2013

1. Background

Irritable Bowel Syndrome (IBS) is a functional colonic disease with high prevalence. Typical symptoms include "chronic abdominal pain, discomfort, bloating, and alteration of bowel habits" [1]; it has been linked to chronic pain, fatigue, and work absenteeism and is considered to have a severe impact on quality of life [2, 3]. Although there is no known cure for IBS, there are treatments that attempt to relieve symptoms, including dietary adjustments, medication and psychological interventions [1].

In 2010, the Canadian College of Naturopathic Medicine (CCNM) was commissioned to conduct a study to investigate the effect of a probiotic agent on IBS. The study's details and a preliminary data analysis using hierarchical linear models (HLM) can be found in a preliminary report: it's key findings are that a strong placebo/expectation effect is present in the early stages of the study which is not entirely surprising given the nature of the phenomenon under study, and that there is no strong statistical evidence to suspect that the agent itself has much of an effect on mild to moderate IBS [4].

The sponsor has expressed interest in determining whether these findings still hold, when the trial data is examined using analysis of covariance (ANCOVA). The ANCOVA is a general linear model which evaluates whether the population means of a dependent/response variable (in this case, IBS Severity or a measure of Quality of Life (QoL)) are equal across levels of a categorical independent variable (in this case, two treatment effects over time), while statistically controlling for the effects of covariates (in this case, the baseline scores for IBSS and QoL). By comparison with the more traditional analysis of variance (ANOVA), the ANCOVA can be used to increase the likelihood of finding a significant difference between treatment groups (when one exists) by reducing the within-group error variance.

2. Study Parameters

Detailed information regarding the pilot study (start dates, recruitment procedure, outcome measures and safety/adverse reactions) and various other graphs and details can be found in the preliminary report [4].

3. Understanding the Data Structure

3.1 Recruitment

The study recruited 129 participants, 67 of which were assigned to the active treatment group, and 62 to the placebo group. The recruitment procedures included advertisements on the radio, in local newsletters and

newspapers, on the web and in social media. Furthermore, local MDs and NDs were given recruitment posters for their clinic in order to encourage patient referrals.

3.2 Randomization

In order to facilitate a balanced representation in the active treatment group and the placebo group in terms of their demographical characteristics, participants were first categorized by their gender group (M/F) and age group (< or ≧50 years). Within each subgroup, participants were then randomly assigned to the treatment group or the placebo group, in a double-blind fashion (i.e. neither the examiners nor the participants were aware of the groups to which they had been assigned). As the number of treatment/placebo assignments in each group was not intended to be even, this randomization process leads us to (Unbalanced) Randomized Complete Block Design.

3.3 Outcome Measures

The two response variables under considerations are the IBS Severity (IBSS) score and the IBS Quality of Life (QoL) measure. The IBSS scores are collected at the beginning of the study (baseline) and at one-month intervals for three months. The participants are also asked to submit the QoL questionnaire at the start of the study, as well as at the second and the third month of their follow-ups. As a side note, both of these response variables are computed using self-reported data.

3.4 Drop-outs, Missing Observations, and Imputation

Ten participants did not deliver any information after the baseline measure: seven participants from the active group and three from the placebo group. As there was no information regarding the treatment effects for those participants, they were eliminated from the remaining analysis. As shown in Table 1 in [4], drop-outs and the remaining participants do not differ significantly in terms of their baseline IBSS severity and QoL. Furthermore, six participants failed to follow-up after the first or the second month of the study. The following table summarizes the breakdown of those participants.

	Total # of recruited participants	Dropped out after Baseline	Dropped out after Month 1	Dropped out after Month 2	Remaining after Month 3
Placebo	62	3	2	2	55 (88.7%)
Treatment	67	7	1	1	58 (86.7%)
Total	**129**	**10**	**3**	**3**	**113 (87.6%)**

Table 1 – IBSS drop-out data. Only those participants that remain after the first two months are retained.

Since the covariance analysis requires the dataset to be free of missing observations, imputations must be performed before proceeding with the analysis.

It is difficult to study the exact reasons why some participants terminate the follow-up prematurely; however it could be conjectured that participants who complete the study are either more likely to believe in the effect of the active agent or to actually be feeling the effect of the treatment than those who fail to complete the treatment. In fact, taking a look at drop-outs with partial information, it is often the case that these observations do not follow the general downward trend seen in the participants with the complete information. In an attempt to test this conjecture, partial non-respondents should be kept in the analysis.

Therefore, for those participants with recorded observations up to the second follow-up, the Last Observation Carried Forward (LOCF) imputation was favoured over the regression imputation [5], and implemented for the analysis. However, it should be noted that a few participants dropped out of the study after the first follow-up. Due to the observed month-to-month fluctuation in the scores within each patient, it may not be reasonable to

assume that the IBSS scores for these participants is constant over a two month period. For the QoL scores, there is no information regarding the treatment effect for those patients as the QoL questionnaire was not administered at the first follow-up. Therefore, the decision was made to eliminate these participants from further analysis.

To compensate for the fact that the imputation was done prior to the covariance analysis, one degree of freedom is docked for each imputation. Note that only the missing observations at the third month into the study are imputed, as we are interested in comparing the baseline measures and the final measures.

In summary, participants ID25 and ID78 are eliminated from both the IBSS and QoL analyses; ID14, ID34, and ID82 are imputed in both cases; ID92 is eliminated from the IBSS analysis, but is imputed for the QoL analysis

3.5 Outlier Detection

Outlying observations frequently have a dramatic effect on the fitted values of the selected model; should such extreme points be found in the dataset, they need to be studied carefully in order to determine whether they should be retained or removed [6]. If influential observations are identified, remedial measures may need to be applied in order to minimize their undue effects.

Given that we have at most four data points per participant, and due to the large observed within-participant variability over time, it is near impossible to identify within-participant observations which we could deemed to be "extreme". It is, however, significantly easier to identify any abnormal between-participant observations.

Numerous methods can be used to find outliers; none of them are foolproof and good judgement must be used. For this reason, the box-and-whisker plots can help in the search for possible outliers: data points falling below $Q_1 - 1.5 \cdot \text{IQR}$ or above $Q_3 + 1.5 \cdot \text{IQR}$, (where Q_1, Q_3, IQR are the first quartile, the third quartile and the inter-quartile range, respectively) require a more in-depth analysis (see Figure 1).

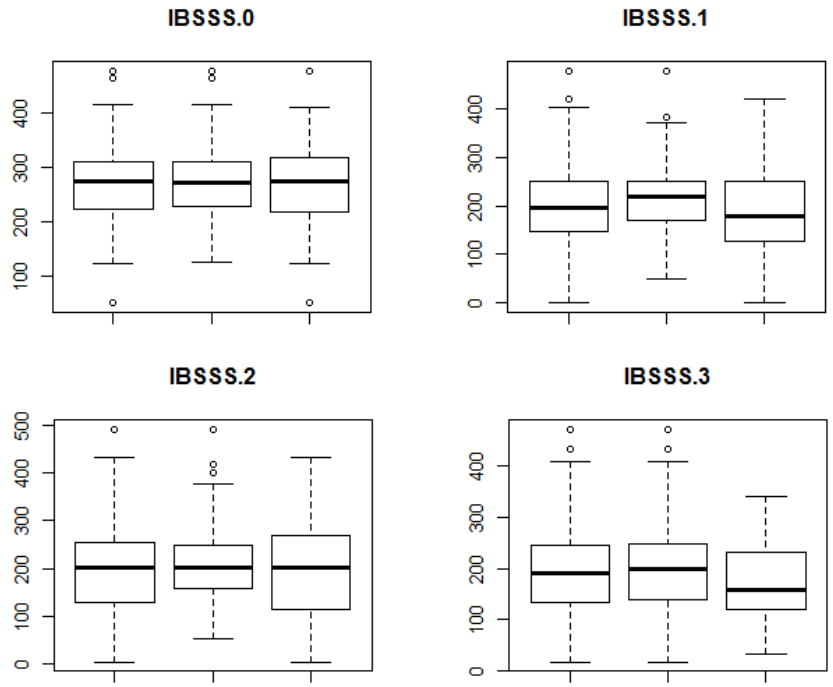

Figure 1 – Box-and-whisker plots for IBS scores at each time point. The three columns represent the Overall, Placebo and Treatment groups, respectively, while circles represent outlying values according to the box-and-whisker test**.**

Tables 2 and 3 summarize the results of the detection of potential outliers in the IBSS (in red) and QoL (in blue) response variables, respectively.

IBSS	Overall	Placebo	Treatment
Baseline	114(480) 74(480) 8(467)* 73(50)	74(480) 8(467)*	114(480) 73(50)
Month 1	8(480)* 45(420)	8(480)* 35(383)	None
Month 2	8(491)*	8(491)* 80(419) 106(400)	None
Month 3	8(472)* 90(453)	8(472)* 90(455)	None

QoL	Overall	Placebo	Treatment
Baseline	None	None	36(91.91)
Month 2	8(86.76)* 39(83.82) 80(83.82) 106(83.09) 123(81.62)	None	39(83.82) 123(81.62) 38(73.53)
Month 3	8(88.24)* 80(83.82)	None	123(72.06)

Tables 2 and **3** – Potential outliers. The IBSS score (on the left, in red) and QoL (on the right, in blue) score is shown in parenthesis with its associated participant ID. The asterisk identifies an observation which is constantly an outlier.

Among the potential outliers, ID8 is found to be anomalous at all observations, except for the baseline measure for the QoL. The probability that this specific participant yields undue influence on the proceedings is high; this will be explored further in the Analysis section.

4. Model Selection

As mentioned in Section 3.2, the participants were stratified according to their gender (M/F) and age group (< or ≥50 years), and then randomized within each block in an effort to promote balanced representation between two treatment groups [4]. From a statistical perspective, blocking is used to isolate controllable variables that are not of the primary interest: since participants were randomized within each block (subgroup), and the number of treatment/placebo assignments in each group was not intended to be even, this randomization process leads us to unbalanced Randomized Complete Block Design (RCBD).

4.1 ANCOVA Models

On top of the treatment and the block effects, ANCOVA models involve the linear effect of a continuous covariate [7]: the models that we use are of the following form:

$$y_{ijk} = \mu + \tau_i + \beta_j + \gamma x_{ijk} + \varepsilon_{ijk},$$

where

- y_{ijk} is the k^{th} **response variable** in the i^{th} treatment group and j^{th} block (the final IBSS or QoL value);
- μ is the **overall mean**;
- τ_i is the i^{th} **treatment effect**;
- β_j is the j^{th} **block effect**;
- γ is the **covariate (or regression) effect**;
- $x_{ijk} = X_{ijk} - \bar{X}$ is the k^{th} **covariate (or concomitant variable)** in the i^{th} treatment group and j^{th} block (the baseline IBSS or QoL value adjusted for the mean), and
- ε_{ijk} is the k^{th} **residual** in the i^{th} treatment group and j^{th} block,

The indices correspond to $i = 1,2,\ j = 1, \dots, 4,\ k = 1, \dots, n_{ij},\ \sum_i \sum_j n_{ij} = N$, where N is the number of participants.

4.2 ANCOVA Assumptions

In order to use an ANCOVA model, four assumptions must be satisfied:

1. *Independence and Normality of Residuals*: the residuals are thought to be independently and identically distributed random variables following a normal distribution with zero mean (i.e. $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_{\varepsilon}^2 \boldsymbol{I})$);
2. *Homogeneity of Residual Variances*: the variance of the residuals must be uniform across treatment groups;
3. *Homogeneity of Regression Slopes*: the regression effect (slope) must be uniform across treatment groups, and
4. *Linearity of Regression*: the regression relationship between the response and the covariate must be linear.

The first of these assumptions can be tested with the help of a **QQ-plot** and a scatter plot of **residual vs. fitted values**, while the second may use the **Bartlett** or the **Levene** test. The final assumption is not as crucial as the other three assumptions. Various remedial methods can be applied should any of these assumptions fail [6].

The third assumption, however, is critical to the ANCOVA model. It can be tested with the **equal slope test**: we run an ANCOVA regression on the model of Section 4.1 with an additional interaction term $x \times \tau$. If the interaction is not significant, the third assumption is satisfied. In the event that the interaction term is statistically significant, a different approach (e.g. moderated regression analysis, mediation analyses) is required as using the original ANCOVA model is not prescribed [8]. ANCOVA assumptions will be verified for both IBSS and QoL response variables in sections 5 and 6 respectively.

5. IBSS Analysis

A total of 129 participants were recruited for the study, ten of which dropped out after their baseline assessments. A further three drop-outs were removed (see Section 3.4), leaving a total of N = 116 participants for the IBSS analysis. In order to accommodate the three imputations (again, see Section 3.4), three degrees of freedom are docked from the residual source in the ANCOVA analysis.

5.1 Full Dataset

The ANOVA table for the **Full IBSS ANCOVA Model** is found in Table 4. At first glance, as the *p*-value for the treatment effect is about 0.095, we conclude that there is not enough evidence to suggest that the treatment has an effect at the 0.05 significance level (but there appears to be a significant effect of the treatment on IBSS at the 0.10 significance level).

Source	df	Type III SS	MS	*F*	*p*-value
τ (Treatment)	1	20324	20324	2.838	0.09498
β (Block)	3	14090	4696.667	0.65588	0.58108
γ (Covariate)	1	110609	110609	15.4451	0.00015
ε (Residual)	110 – 3 = 107	766275	7161.449		

Table 4 – ANOVA table for the Full IBSS Model with degrees of freedom modified to accommodate imputation.

The ANCOVA assumptions are verified as follows. The assumptions of normality and independence of the residuals is satisfied based on the visual assessment of diagnostic plots in Figure 2.

The Bartlett statistic against homogeneous variances of the residuals in the treatment group vs. those in the placebo group is $X^2 = 0.5437$, which yields a p-value of 0.4450. There is thus insufficient evidence to conclude that the variances are non-homogeneous across treatment groups. A plot of the variances corroborates the assertion that the second assumption is met (see Figure 3).

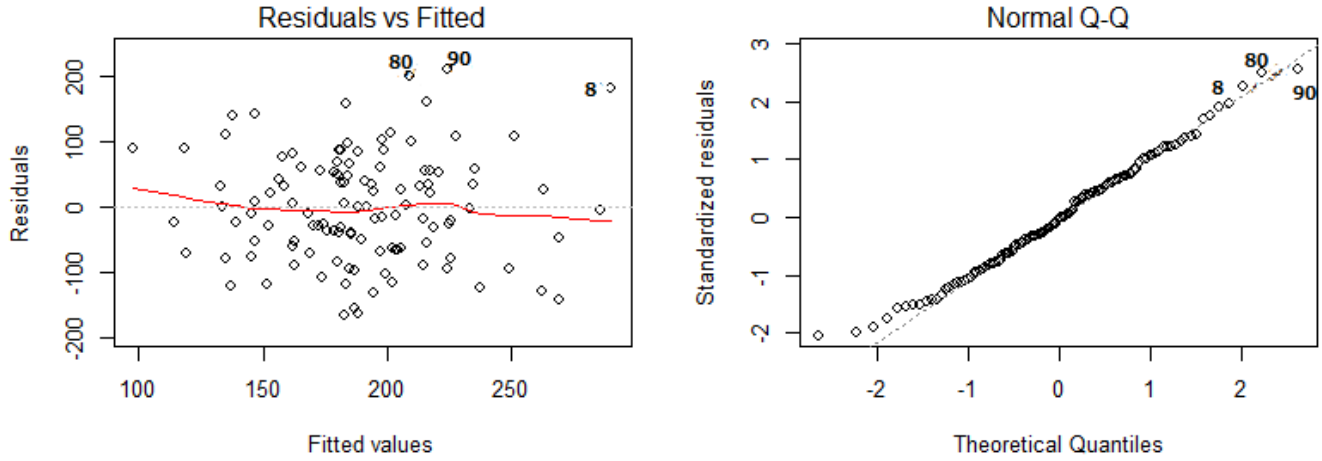

Figure 2 – Normality and independence of the Full IBSS Model residuals.

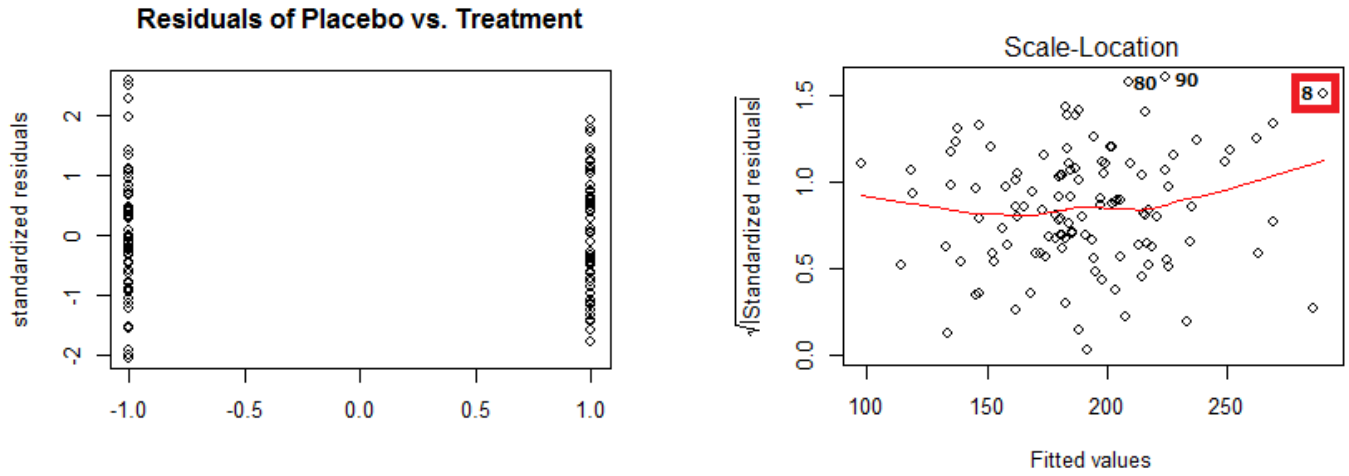

Figure 3 – Homogeneity of variance across treatment groups in the Full IBSS Model (on the left); possible outliers/influential observations (on the right).

Furthermore, with a p-value of 0.0015 for the covariate effect, it seems reasonable to assume that the relationship between the response and the covariate is indeed linear.

Finally, the test for equal slopes compares the original model $y \sim \tau + \beta + \gamma x$ to the modified interaction model

$$y \sim \tau + \beta + \gamma x + \rho(x \times \tau).$$

The lack of significance of the interaction term is interpreted as favourable to the third assumption.

The appropriate ANOVA table is shown in Table 5; while the corresponding *p*-value shows a lack of significance at the 0.05 significance level, it also indicates borderline significance at the 0.10 significance level.

Model	df_ε	RSS	df_{diff}	SS	*F*	*p*-value
Original	107	766275				
Interaction	106	748021	1	18254	2.5867	0.1107

Table 4 – Homogeneity of regression slopes across treatment groups for the Full IBSS Model with degrees of freedom modified to accommodate imputation.

While the ANCOVA assumptions are met at the 0.05 significance level, thereby providing evidence that the results of Table 4 are statistically valid, the assumptions are not satisfied at the 0.10 significance level. Furthermore, there remains some uncertainty in the critical third assumption. This combination provides an impetus to study the effect of possible influential observations. A scatter-plot of the fitted values against the square root of the standardized residuals (see Figure 3 above) reveals yet again a potential outlier in ID8.

Given the consistent abnormal behaviour of ID8, we shall re-run the ANCOVA without the possibly influential ID8 data point.

5.2 Reduced IBSS Dataset

The ANOVA table for the **Reduced IBSS ANCOVA Model** is shown in Table 6 below. The removal of ID8 has the dramatic effect of changing our conclusions to the point that there is no longer enough evidence to suggest that the treatment has an effect even at the 0.10 significance level.

Source	df	Type III SS	MS	*F*	*p*-value
τ (Treatment)	1	15485	15485	2.2479	0.1368
β (Block)	3	11985	3995	0.5799	0.6295
γ (Covariate)	1	76421	76421	11.0937	0.0010
ε (Residual)	106	730204	6888.717		

Table 6 – ANOVA table for the Reduced IBSS Model with degrees of freedom modified to accommodate imputation.

The normality and the independence of the residuals can be assessed visually – the appropriate plots (akin to Figures 2 and 3) were essentially the same and are therefore not shown. The Bartlett's test statistic for the Reduced Model is $X^2 = 0.3798$, corresponding to a large *p*-value of 0.7594. Furthermore, with a *p*-value of 0.0010, the linearity of the regression between the response and the covariate is highly significant.

It remains only to verify the assumption of the equal regression slopes. With a *p*-value of 0.288 (see Table 7), the implication is that the third assumption is indeed valid at reasonable significance levels.

Model	df_ε	RSS	df_{diff}	SS	*F*	*p*-value
Original	106	730204				
Interaction	105	722362	1	7842.5	1.1400	0.2881

Table 7 – Homogeneity of regression slopes across treatment groups for the Reduced IBSS Model with degrees of freedom modified to accommodate imputation.

There is thus ample evidence to suggest that ID8 is indeed an influential observation and should be removed from the analysis, and that the results of Table 6 are statistically valid.

In that case, the ANCOVA coefficients are found to be as in Table 8.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 177.7275   20.4987   8.670 4.53e-14 ***
x             0.3573    0.1058   3.378  0.00101 **
tauT        -23.2885   15.3177  -1.520  0.13131
betaOM       36.0756   33.2704   1.084  0.28062
betaYF       25.6123   21.1943   1.208  0.22949
betaYM       18.1062   27.9178   0.649  0.51799
---
```

Table 8 – ANCOVA coefficients for the Reduced IBSS Model. The placebo effect $\boldsymbol{\tau_1}$ and the females-over-50 block effect $\boldsymbol{\beta_1}$ are both set to 0. The only significant coefficients (at reasonable significance levels) are the intercept (the overall mean $\boldsymbol{\mu}$), and x (the covariate effect $\boldsymbol{\gamma}$). Individually, the others cannot be differentiated from 0 at the 0.13 significance level (in the best case scenario).

6. QoL Analysis

As before, a total of 129 participants were recruited for the study, ten of which dropped out after the baseline assessment. This time however, only two drop-outs were removed (see Section 3.4), leaving a total of $N = 117$ participants for the QoL analysis. In order to accommodate the four imputations (again, see Section 3.4), four degrees of freedom are docked from the residual source in the ANCOVA analysis.

6.1 Full Dataset

The ANOVA table for the **Full QoL ANCOVA Model** is found in Table 9. At first glance, there is still not enough evidence to suggest that the treatment has an effect at the 0.05 significance level, but unlike the IBSS case, the threshold is very nearly reached (treatment p-value=0.0556).

Source	df	Type III SS	MS	F	p-value
τ (Treatment)	1	998	998	3.7453	0.05560
β (Block)	3	398.7	136.7	0.5130	0.6742
γ (Covariate)	1	13949.9	13949.9	52.35163	<0.0001
ε (Residual)	111 – 4=107	28511.8	266.4654		

Table 9 – ANOVA table for the Full QoL Model with degrees of freedom modified to accommodate imputation.

With the aids of the normal Q-Q plot and the scatter plot of the residuals against the fitted values, there is no strong evidence to suspect the validity of the normality and the independence of the residuals (the two plots are essentially the same as Figures 2 and 3, and are not shown here).

The Levene's test statistic for the Full QoL Model is $W = 1.3327$, with an associated p-value of 0.2508 for equal variances in residuals across two treatment groups. Furthermore, with the covariate p-value of 0.0010, the linearity of the regression between the response and the covariate seems highly significant.

Finally, with a *p*-value around 0.37, Table 10 implies that there is no strong evidence to suspect the validity of the most critical ANCOVA assumption: the assumption of the equal slopes.

Model	df_ε	RSS	df_{diff}	SS	*F*	*p*-value
Original	107	28512				
Interaction	106	28295	1	216.58	0.8113617	0.3698

Table 10 – Homogeneity of regression slopes across treatment groups for the Full QoL Model (IBSS) with degrees of freedom modified to accommodate imputation.

As all of the assumptions for the ANCOVA are verified, the parameter estimates are found to be as in Table 11.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 30.39160    4.00702   7.585 1.10e-11 ***
betaOM       7.58904    6.50888   1.166   0.2461
betaYF       2.80135    4.13083   0.678   0.4991
betaYM       0.89675    5.47816   0.164   0.8703
x            0.56871    0.07717   7.369 3.25e-11 ***
tauT        -5.86381    2.97492  -1.971   0.0512 .
---
```

Table 11 – ANCOVA coefficients for the Full QoL Model. The placebo effect $\boldsymbol{\tau_1}$ and the females-over-50 block effect $\boldsymbol{\beta_1}$ are both set to 0. The only significant coefficients (at reasonable significance levels) are the intercept (the overall mean $\boldsymbol{\mu}$), and *x* (the covariate effect $\boldsymbol{\gamma}$). Individually, the other block coefficients cannot be differentiated from 0 at the 0.25 significance level (in the best case scenario). The treatment effect $\boldsymbol{\tau_2}$, while associated with a relatively small *p*-value, still cannot be differentiated from 0 at the 0.05 level.

6.2 Reduced Dataset

Using standard outlier detection techniques, four observations can be identified as possible outliers in the Full QoL Model: ID39, ID53, ID68, ID 106 (see Figure 4).

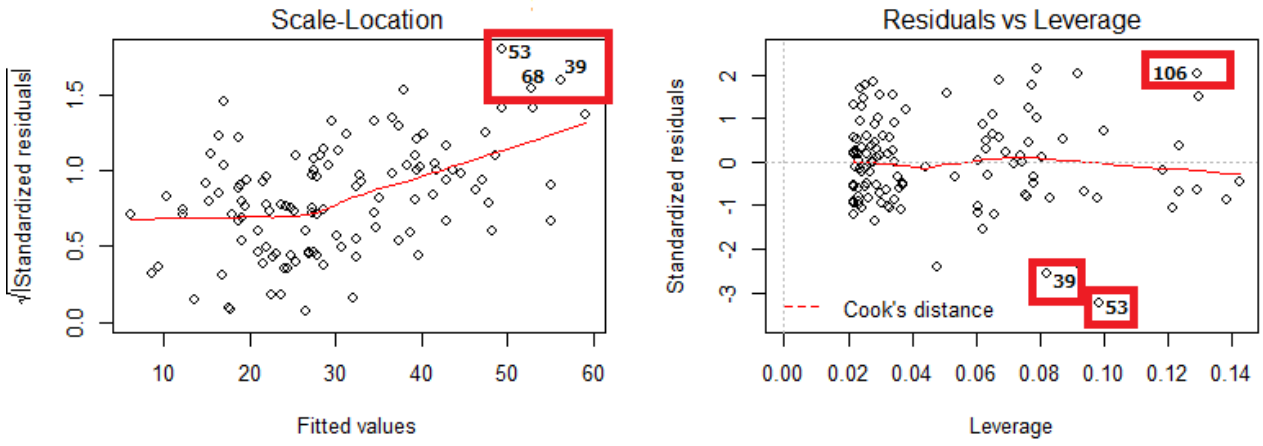

Figure 4 – Possible outliers/influential observations in the Full QoL Model.

It should be noted that there are thirteen patients (seven from the treatment group, six from the placebo group) who have improved their QoL scores by more than 35 points over three months (ranging from 35.29 to 75). At the same time, five patients have increased their scores by more than 15 points (i.e., they felt worse) over three months period (ranging from 15.44 to 33.82).

Furthermore, when we compare patients' QoL scores between the second and third follow-ups, we notice that four participants (three from the treatment group, one from the placebo group) have improved their scores by more than 30 points (ranging from 30.88 to 66.91 points), and five patients (three from the treatment group, two from the placebo group) have worsened by more than 15 points (ranging from 16.18 to 35.3).

This implies that the QoL scores have large variability within each participant. The above outliers may have shown particularly large improvements, but there seems to be enough within-participant variance in general to mark them as "typical", in some sense. As such, we will not build a Reduced QoL Model as eliminating these outliers may result in a significant loss of information.

7. Conclusions and Recommendations

We end the report with key findings of our analysis, as well as some recommendations for future investigations.

7.1 Blocking and Balanced Designs

In this report, we have found that blocking (or subgrouping) the participants according to their gender and age does not play an important role in the ANCOVA. In future studies involving this probiotic agent, blocking should only be used if there are compelling reasons to suspect that treatment effects are different for at least one subgroup, as blocking results in fewer degrees of freedom.

Special care should also be taken to have a balanced design (i.e., equal number of replicates for each subgroup), especially if subgroup analyses are of interest: in the 2010 IBS Study, for instance, the overwhelming number of female participants and small number of male participants make any conclusions about male subgroups statistically unsound.

7.2 Recruitment Process

In the 2010 IBS Study, participants needed to come forward to be selected. The recruitment process used advertisements on the radio, in local newsletters and newspapers, on the web and social media, as well as posters with which local MDs and NDs could encourage patient referrals.

The elephant in the room is that this type of recruitment process leads to self-selection biases: the participants in the 2010 IBS Study may not constitute a representative sample of IBS sufferers, which makes it difficult to generalize the result of the analyses beyond the collected sample, even when there is a significant impact.

This is a problem that plagues numerous clinical studies – unfortunately, it is quite difficult to counter this situation.

7.3 Practical Significance of Results

Our interpretation of the ANCOVA results is similar to those of the preliminary report: with the caveat brought up in section 7.2, there is simply not enough evidence to conclude that the agent is effective against IBS [4].

It is true that the effect of the treatment on the (self-reported) QoL score is nearly statistically significant at the 0.05 significance level. The corresponding estimated treatment effect is -6.4454, which means that on average, participants in the treatment group seem to have lost an extra 6.4 QoL points over the course of three months, compared to those in the placebo group. However, given the amount of variability in individuals from month to

month, we are reluctant to conclude that the agent under investigation provides a practically significant improvement in the average participant's quality of life.

Further investigation may shed some light on the situation and will help us determine if the relationship between the agent and QoL is causal or spurious.

7.4 Publication of Results

Even though this study did not find any statistically significant improvement for IBSS, it should be published in order to counter publication bias.

References

[1] "Irritable Bowel Syndrome," [Online]. Available: http://en.wikipedia.org/wiki/Irritable_bowel_syndrome. [Accessed 5 May 2013].

[2] P. Paré, J. Gray, S. Lam, R. Balshaw, S. Khorasheh, M. Barbeau, S. Kelly and C. R. McBurney, "Health-related quality of life, work productivity, and health care resource utilization of subjects with irritable bowel syndrome: baseline results from LOGIC (Longitudinal Outcomes Study of Gastrointestinal Symptoms in Canada), a naturalistic study," *Clinical Therapeutics,* vol. 28, no. 10, pp. 1726-35, 2006.

[3] S. Maxion-Bergemann, F. Thielecke, F. Abel and R. Bergemann, "Costs of irritable bowel syndrome in the UK and US," *PharmacoEconomics,* vol. 24, no. 1, pp. 21-37, 2006.

[4] P. Herman, C. Kooley and D. Seely, "Double-blind placebo-controlled pilot study to investigate the effects of an investigational Probiotic on Irritable Bowel Syndrome," 2011.

[5] S. Hagiwara, "Nonresponse Error in Survey Sampling: Comparison of Different Imputation Methods," Ottawa, 2012.

[6] M. H. Kutner, C. J. Nachtsheim, J. Neter and W. Li, Applied Linear Statistical Models, 5th ed., New York: McGraw-Hill/Irwin, 2004.

[7] P. W. John, Statistical Design and Analysis of Experiments, New York: Macmillan, 1971.

[8] S. Green and N. Salkind, Using SPSS for Windows and Macintosh: Analyzing and Understanding Data, 6th ed., Upper Saddle River, NJ: Prentice Hall, 2011.

CQADS only endorses the main part of this report. However, the results of other analyses are also presented in the following addenda.

Addendum 1 – IBS Severe Participants Only

The sponsor has expressed interest in determining whether the treatment effect was significant among the severe group of IBS sufferers (self-reported baseline IBSS between 300 and 500). The results are shown below. There were 16 severe IBS sufferers in the placebo group and 19 in the treatment group. Participant ID82 (placebo group) was imputed using the LOCF method described in the main report; ID92 (placebo group) was eliminated as the participant had no information after the first month.

The difference in the treatment/placebo effect is significant (see Table 12, on page 12) with a p-value of 0.049. The inclusion of a covariate seems unnecessary (which makes sense as γ is there to account for the baseline score differences among individuals; in this case, we are restricting the variation of the initial IBSS score from 300 to 500). As was the case previously, the block effects are not significant.

Source	df	Type III SS	MS	*F*	*p*-value
τ (Treatment)	1	45934	45934	4.219	0.049
β (Block)	3	5279	1759.7	0.162	0.92
γ (Covariate)	1	191	191	0.0175	0.90
ε (Residual)	28	304873	10888.3		

Table 12 – ANOVA table for the Reduced IBSS Model with degrees of freedom modified to accommodate imputation (severe IBSS).

Note, however, that with a multiple R^2 coefficient of 0.15, the model is **not** statistically significant, that is, it does not really explain the variation among changes in participants' scores and the treatment effect having borderline significance does not really mean anything. The small sample size and small study length are likely factors explaining this phenomenon.

There are also a number of diagnostic issues – the normality assumption does not seem adequate, as can be seen by the Normal Q-Q scatter plot in Figure 5 (though Shapiro-Wilk test for normality yields *p*-value of 0.18 so we cannot reject the assumption of normality); furthermore, Levene's test statistic is W=3.115 with a *p*-value of 0.087 indicates that there is some evidence to suspect that the assumption of equal variances between treatment groups is inadequate. The assumption of equal slopes is met as the *p*-value for the test is 0.695.

Since the covariate should not be included in the model as it does not add enough information, running an ANOVA considering only group and treatment effects (see Table 13, page 13) yields a *p*-value of 0.41 for the entire model and a multiple R^2 coefficient of 0.13. Again, such a small R^2 implies that the model fails to capture the cause for the changes in patients' IBSS scores.

As has been discussed previously, there is a fair amount of month-to-month variability within each patient, which makes any meaningful analysis (on quite a small sample) difficult. The participants' self-selection bias might also come into play.

Another consideration is that it is not clear whether the self-reported IBS severity category is sufficiently robust as a stratification mechanism; a non-negligible proportion of severe IBS sufferers in the placebo group drop to a lower intensity category after one month (40%, with an average drop of 172 points), casting doubt as to the appropriateness of that classification in the first place. An objective method to rate IBS scores might improve the situation.

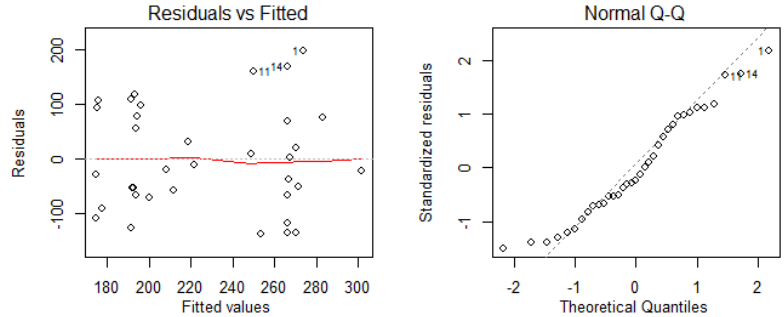

Figure 5 – Diagnostic plots for ANCOVA model restricted to severe IBS sufferers. Departures from the model assumptions are easy to spot.

Model	df	Type III SS	MS	*F*	*p*-value
Full	4	38371	9592.75	1.03	0.41

Table 13 – ANOVA model (without covariate)

Addendum 2 – IBS Severe Participants with Complete Responses Only

Eliminating participant ID82 from the sample given in Addendum 1, it was found that the inclusion of a covariate was unnecessary (as the covariance analysis yielded a *p*-value of 0.94 for γ).

The results of running an ANOVA considering only group and treatment effects are summarized in Table 14. From the ANOVA, the treatment effect seems to be statistically significant with a p-value of 0.025; however, a *p*-value of 0.21 for the whole model, a multiple R^2 coefficient of 0.18, a small sample size and exclusion of non-respondent(s) make it difficult for this model to capture the cause for the changes in patients' IBSS scores.

The normal Q-Q plot (Figure 6) shows a moderate departure of the normality assumption for the ANOVA; however, given a small sample size, this is within an acceptable deviation.

Source	df	Type III SS	MS	*F*	*p*-value
Model	4	61921	15480.25	1.55	0.21
τ (Treatment)	1	55768	55768	5.59	0.025
β (Block)	3	6153	2051	0.21	0.89
ε (Residual)	29	289335	9977		

Table 14 – ANOVA table for the Full IBSS Model (Severe IBSS, Complete Responses).

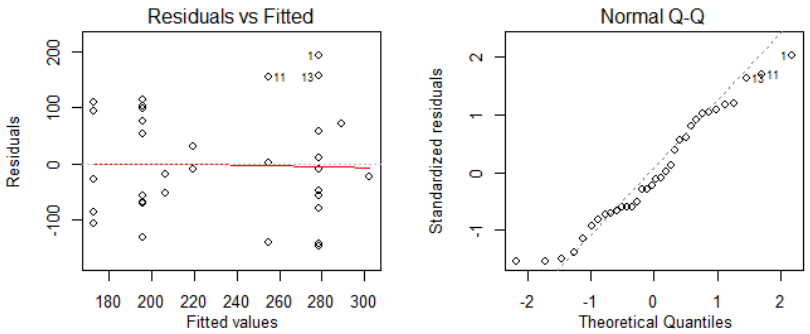

Figure 6 – Diagnostic plots for ANOVA model restricted to severe IBS sufferers without imputation.

Addendum 3 – IBS Complete Respondents Only

From 129 subjects recruited for this study, 113 persons (88%) completed the third follow-up. The results of the ANCOVA for complete respondents are summarized in Table 15.

The treatment effect is not found to be statistically significant (p-value of 0.112). The assumptions of ANCOVA are verified as follow: the p-value of the equal slope test is 0.124; the Bartlett's test of equal variance has p-value of 0.60; and the normal Q-Q plot shows well behaved middle points (Figure 7). It should be noted that all of the outliers found in the diagnostic check are within 3 standard deviations in magnitude; therefore all outliers are retained in the analysis.

Source	df	Type III SS	MS	F	p-value
τ (Treatment)	1	17584	17584	2.56	0.112
β (Block)	3	13657	4552.3	0.66	0.576
γ (Covariate)	1	110558	110558	16.12	0.0001
ε (Residual)	107	733825	6858.2		

Table 15 – ANOVA table for the Full IBSS Model with degrees of freedom modified to accommodate imputation (Complete Responses).

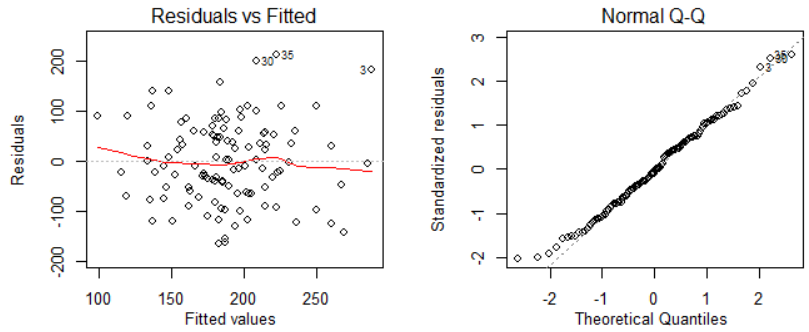

Figure 7 – Diagnostic plots for ANCOVA model restricted to Complete Respondents.

Addendum 4 – QoL Severe Participants Only

The sample for the analysis on the QoL scores of severe IBS sufferers consists of 34 participants, of which 17 subjects are in the placebo group and 19 in the treatment group. Two participants from the placebo group, ID 82 and 92 are imputed using the LOCF method described in the main report.

The main results of the covariance analysis are summarized in Table 16. The analysis yields that the treatment effect is statistically significant, with corresponding p-value of 0.0072.

The assumptions of ANCOVA are verified as the p-value of the equal slope test was 0.98, and the Levene's test for equal variance had p-value of 0.53. The normal Q-Q plot (Figure 8) shows a moderate departure from the normality assumption; however, given the small sample size, this is within an acceptable deviation. It should also be noted that all of the outliers found in the diagnostic check are within 3 standard deviations in magnitude; therefore all outliers are retained in the analysis.

Even though the agent is found to be statistically significant, it is difficult to generalize the effectiveness of the agent seen in this analysis. The analysis was based on a sample size of 36, of which only 6 were men. Therefore, without further studies incorporating a larger sample size, it is statistically unsound to generalize the results of this analysis to severe IBS sufferers. Furthermore, it should be noted that sampling bias and self-reporting bias may also be present

Source	df	Type III SS	MS	*F*	*p*-value
τ (Treatment)	1	2497.5	2497.5	8.421	0.0072
β (Block)	3	1534.7	511.6	1.72	0.185
γ (Covariate)	1	4211.7	4211.7	14.2	0.0008
ε (Residual)	30-2=28	8304.8	296.5893		

Table 16 – ANOVA table for the Full QoL Model with degrees of freedom modified to accommodate imputation.

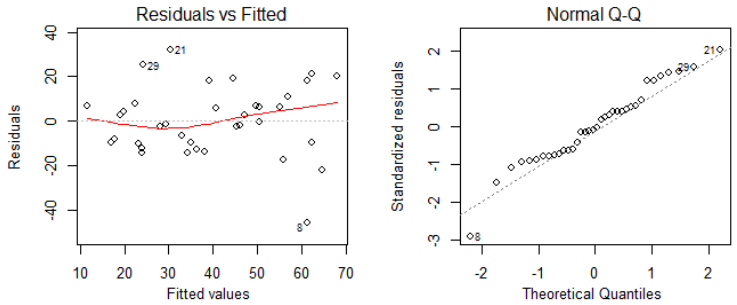

Figure 8 – Diagnostic plots for ANCOVA model restricted to QoL scores of severe IBS sufferers.

Addendum 5 – QoL Severe Participants with Complete Responses Only

In this section, ID 82 and 92 are eliminated from the sample given in Addendum 4.

The results of the covariance analysis are summarized in Table 17. The analysis yields that the treatment effect is statistically significant, with corresponding *p*-value of 0.0060.

However, due to its small sample size, elimination of partial respondents, and potential sampling and self-report bias, further studies are necessary to investigate the effectiveness of the agent found in this analysis.

The assumptions of ANCOVA are verified as the p-value of the equal slope test was 0.83, and the Levene's test for equal variance had *p*-value of 0.67. The normal Q-Q plot (Figure 8) shows a moderate departure from the normality assumption; however, given its small sample size, this is within an acceptable deviation. It should also be noted that all of the outliers found in the diagnostic check are within 3 standard deviations in magnitude; therefore all outliers are retained in the analysis.

Source	df	Type III SS	MS	*F*	*p*-value
τ (Treatment)	1	2511.7	2551.7	8.84	0.0060
β (Block)	3	1444.6	481.5	1.70	0.19
γ (Covariate)	1	3363.8	3363.8	11.84	0.0018
ε (Residual)	28	7953.9	284.1		

Table 17 – ANOVA table for the Full QoL Model with degrees of freedom modified to accommodate imputation (Severe IBS, Complete Responses).

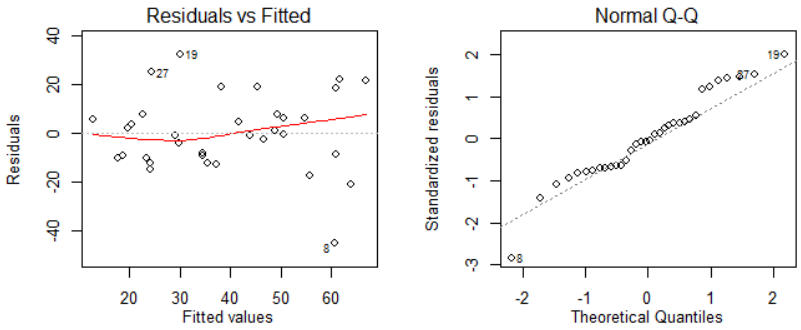

Figure 9 – Diagnostic plots for ANCOVA model restricted to QoL scores of severe IBS sufferers.

Addendum 6 – QoL Complete Respondents Only

From 129 subjects recruited for this study, 113 persons (88%) completed the third follow-up. The results of the ANCOVA for complete respondents are summarized in Table 17.

The treatment effect is found to be statistically significant at 95% level (p-value of 0.048). The diagnostic checks for the assumptions of ANCOVA are given as follows: the equal slope test yields a p-value of 0.358; the Levene's test of equal variance has p-value of 0.157; the residuals do not show major deviation from the normality assumption. It should be noted that participant ID 53 has standard residual of -3.21; however, elimination of ID 53 would result in the violation of the assumption of equal slopes (with corresponding p-value of 0.09). Therefore, no further analysis was conducted.

Source	df	Type III SS	MS	F	p-value
τ (Treatment)	1	1017.0	1017.0	4.01	0.048
β (Block)	3	366.3	122.1	0.48	0.70
γ (Covariate)	1	12413.3	12413.3	48.96	<0.0001
ε (Residual)	107	27128.1	253.5		

Table 18 – ANOVA table for the Full QoL Model with degrees of freedom modified to accommodate imputation.

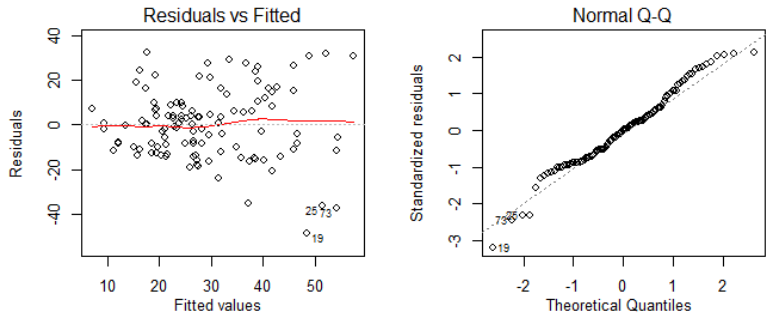

Figure 10 – Diagnostic plots for ANCOVA model restricted to QoL scores of complete respondents.