EVALUATION OF NEIMANN-PICK TYPE C1 DISEASE USING MIXED EFFECTS MODELS

by

Michael Backman

B. S. in Statistics, University of Pittsburgh, 2012

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

Michael Backman

It was defended on

November 19, 2015

and approved by

Francis Pike, PhD, Assistant Professor, Department of Critical Care Medicine, School of
Medicine, University of Pittsburgh

Evelyn O. Talbott, PhD, Professor, Department of Epidemiology, Graduate School of Public
Health, University of Pittsburgh

Lisa Weissfeld, PhD, Internship Advisor, Statistics Collaborative, Inc., Washington, D.C.

Thesis Advisor: Vincent Arena, PhD, Associate Professor, Department of Biostatistics,
Graduate School of Public Health, University of Pittsburgh

Vincent C. Arena, PhD

EVALUATION OF NEIMANN-PICK TYPE C1 DISEASE USING MIXED EFFECTS MODELS

Michael Backman, M.S.

University of Pittsburgh, 2015

ABSTRACT

Niemann-Pick Type C1 (NPC1) is an autosomal neurologic pediatric orphan disease, with death occurring by age 20-25. It is estimated that 1 in 150,000 people in the United States (US) and 1 in 100,000 people in the European Union (EU) suffers from NPC1. The disease results from NPC1 gene mutations which cause defective protein activity, leading to harmful levels of cholesterol and sphingolipids in cells.

Using data from a natural history study and a phase 1 study we investigated the efficacy of a potential treatment for NPC1 in patients aged 6 to 26 with multiple follow-up visits. To evaluate the change in severity of neurologic manifestations, we focused on nine different measures: dietary restrictions, diminished lip strength, diminished tongue strength, dysarthria, ability to consume liquids, risk of laryngeal penetration when consuming liquids or solids, ability to consume solids, speech difficulties, and swallowing difficulties. The primary objective of our analysis evaluated the efficacy of the proposed treatment in reducing the aforementioned symptoms of NPC1. For each symptom, a random intercept mixed-effects model was fit incorporating the factors of time, treatment, and their interaction as potential predictors.

Our strategy has certain advantages over a traditional randomized clinical trial (RCT) as it uses patients enrolled from an ongoing natural history study for which we already have some data on the progression of the disease. Furthermore, it provides a pool of patients from which to recruit phase 1 participants. From an analytic point of view, it provides an efficient use of the available data and an alternative design to an RCT that requires greater sample sizes which would be difficult to obtain in the setting of a rare disease. The utilization of our study design can expedite orphan disease research by decreasing the amount of patients necessary to reach meaningful conclusions from efficacy studies. For these reasons, our approach provides a practical alternative in evaluating efficacy of treatment in rare/orphan disease research, a very current public health issue.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0   INTRODUCTION

An orphan (OD) or rare disease is generally understood to be any disease that affects a small proportion of a population. However, there is no universal standard for a disease to be recognized as a rare disease. According to the US Rare Disease Act of 2002, a rare disease is one that affects less than 200,000 people in the United States. In Japan, the definition of a rare disease is one that impacts less than 50,000 people. The European Commission on Public Health gives two criteria for a disease to be classified as a rare disease. The first is that the disease is life-threatening or chronically debilitating. The second criterion is that less than 1 in 2000 people suffers from it. If both of these criteria are met, then it is understood that special action is necessary to combat the disease.

In fact, rare disease and orphan disease are not synonymous terms. Factors such as severity of disease and availability of treatment and resources help determine whether a disease is rare. As a result, especially after the US Orphan Drug Act was passed in 1983, rare diseases are commonly thought of as ODs. In fact, the Orphan Drug Act includes rare and non-rare diseases for which producing drugs is not expected to be profitable.

Despite the overall low incidence of ODs, the prevalence of ODs can vary widely by geographic region and demographic group. Indeed, in 2011 there were about 25 million Americans suffering from ODs. One estimate gives the number of ODs at about 8000 and a great number of them are inherited genetically, affect children at very early stages in life, and have

1

grave physical effects. ODs are by definition rare and while some are relatively well-known (cystic fibrosis and Huntington's disease), others have less than 100 people afflicted with that disease in the United States. Thus, due to the time-consuming, complicated, expensive, and frequently unsuccessful nature of drug development, pharmaceutical companies have little incentive to produce products to treat these conditions. The US Orphan Drug Act incentivized drug companies to produce drugs to treat orphan diseases. As a result, as of 2011, there were about 325 drugs on the market to treat various orphan diseases. Unfortunately, these products only provide treatment for about 5% of known ODs. In addition, most of these 325 drugs specifically treat rare cancers or metabolic diseases, resulting in a dearth of treatment options for diseases in other disease classes. In addition, most of these new treatment options are based on combating symptoms, rather than providing a cure or addressing the person's faulty biological processes that are the starting point for the disease. Finally, the high cost of these products, even by the standards of pharmaceutical drugs, is a significant burden for patients and insurers alike. This is despite the efforts entailed in the Orphan Drug Act to counter those costs.

For this research project, we studied neurologic symptoms of Niemann-Pick Type C1 (NPC1) disease. NPC1 is an orphan disease resulting from mutations in the NPC1 gene. The prevalence of NPC disease is estimated to be 1 in 150,000 people in the United States (US) (Millat, et al., 1999) and 1 in 100,000 people in the European Union (EU) (Vanier, 2010). It is autosomal recessive and mostly occurs in children and teenagers, with premature death by age 20-25.

Niemann-Pick Type C1 (NPC1) disease is a rare, neurodegenerative, inherited, autosomal recessive disorder which primarily manifests in children and teenagers. Niemann-Pick Type C

(NPC1) is characterized by systemic disease (hepatic, splenic, pulmonary) and progressive neurologic manifestations including ataxia, dementia and seizures resulting in early death by the age of 20-25 years of age. The prevalence of NPC disease is estimated to be 1 in 150,000 people in the United States (US) (Millat, et al., 1999) and 1 in 100,000 people in the European Union (EU) (Vanier, 2010). NPC1 disease is a result of mutations in the NPC1 gene whose protein normally functions in neurons and other cells to transport unesterified cholesterol and sphingolipids from the late endosomal/lysosomal (LE/LY) compartment. To control the cellular concentration of cholesterol, sphingolipids and other lipids, cells have a complex trafficking system. Part of this system is the lipid efflux from LE/LY via the action of NPC1, at least in part. In patients lacking or with deficient NPC1, cholesterol, bis-(monoacylglycerol) phosphate, and various sphingolipids accumulate to toxic concentrations in lysosomal storage organelles. Most of the evidence supports that the primary storage metabolite in NPC1 disease, especially in the peripheral tissues, is low-density lipoprotein-derived cholesterol (Rosenbaum & Maxfield, 2011). Cholesterol and sphingolipids are important lipids in mammalian physiology; however, in excessive concentrations both are toxic to cells resulting in cellular dysfunction and death. The neuropathological abnormalities in NPC patients that are caused by this block in cholesterol transport include; brain atrophy, widespread neuronal cytoplasmic vacuolization and neuronal loss, with the Purkinje cells being the most severely affected.

We used data from one natural history study and one phase 1 study. The natural history study consisted of 78 patients with NPC disease and included assessment of disease severity, neuropsychiatric, ophthalmological, audiological, speech, and language pathology. Cholesterol metabolite measurements were taken from blood, urine, and cerebrospinal fluid (CSF). Data was

collected from 2006 to August 2014. Out of these 78 patients, 65 received speech evaluations and form the basis for the results presented here.

We specifically examined NPC1 in natural history patients aged 6 to 26 years at entry to study with multiple visits and all phase 1 patients with multiple visits. In order to evaluate the severity of neurologic manifestations, we used nine variables: dietary restrictions, diminished lip strength, diminished tongue strength, dysarthria, ability to consume liquids, risk of laryngeal penetration when consuming liquids or solids, ability to consume solids, speech difficulties, and swallowing difficulties. A person with dysarthria has abnormally weak speech muscles, resulting in difficulty with clear articulation.

The phase 1 study was a single center and open-label study, meaning that providers and patients were aware of which treatment the patient was receiving. Children with NPC1 disease enrolled in the study received increasingly larger doses of the treatment compound over time via monthly lumbar IT injections, with the goal of arresting the physical and mental deterioration caused by NPC1. This decline is expressed partially through various outcomes that we are studying: difficulty consuming solids and liquids, weakened lip and tongue muscles, dysarthria, risk of laryngeal penetration when ingesting liquids and solids, and speech and swallowing difficulties. It is also generally expressed through other avenues that we are not examining.

There were three primary goals of the phase 1 study, the first of which was to evaluate the safety and tolerability of IT administration of the treatment in patients with NPC disease. The other primary goals were to study the study drug's plasma pharmacokinetics after IT administration and use changes in plasma 24-S hydroxycholesterol level to calculate a biochemically active dose of the treatment. The secondary objectives were to assess the

employment of candidate plasma and CSF biomarkers and to carry out an exploratory appraisal of the utilization of clinical outcome measures in future clinical development.

1.1   NATURAL HISTORY OF ORPHAN DISEASES

We are interested in describing the natural history of orphan diseases. By studying the natural history of an orphan disease, we can glean important information regarding the timeline and severity of the progression of the disease. Particularly in the case of orphan diseases, being well-acquainted with the natural history of the disease is a valuable aid for later studies. In traditional medical trials involving orphan diseases, investigators usually struggle to enroll enough patients to have both treatment and placebo groups of sufficient size. If the natural progression of the disease is well-documented, then knowledge of it could potentially replace the traditional placebo group. Then, researchers do not need to register as many patients to have the same power. If the proposed drug has a large enough effect, a clinical trial with only a single treatment arm can be sufficient.

Analysis of data from natural history studies can provide information for the formation of pertinent research hypotheses for clinical trials and outcomes to be used as endpoints. Finally, patients in natural history studies can be a potential source of participants for future clinical trials. Given the particularly acute struggles to enroll sufficient patients that orphan disease clinical trials face, this last point is crucial. Quick recruitment of willing patients could result in significant reductions in cost and time. This would give pharmaceutical companies additional incentive to attempt to develop treatments for orphan diseases.

The orphan disease Niemann-Pick Type C1 disease is one member of the family of Niemann-Pick diseases. Niemann-Pick Type C includes disorders related to certain irregularities in intracellular transportation of endocytosed cholesterol. Endocytosis occurs when a cell conveys molecules into the cell by surrounding them through the use of an energy-consuming process. There are two types of Niemann-Pick Type C disease: C1, which is caused by a mutation in the NPC1 gene and C2, which is the result of a NPC2 gene mutation. NPC1 is responsible for about 95% of NPC cases. In both cases, the gene mutation causes a protein to be unable to sufficiently remove endocytosed cholesterol from its current location, as described above.

```
┌─────────────────────────┐
│  Full NH dataset (n= 78) │
└─────────────────────────┘
             │
             ▼
┌─────────────────────────┐
│  Speech dataset (n= 65)  │
└─────────────────────────┘
             │
   ┌─────────┴─────────┐
   ▼      Allocation    ▼
phase 1 study (n= 11)   Natural history comparison (n = 17)
```
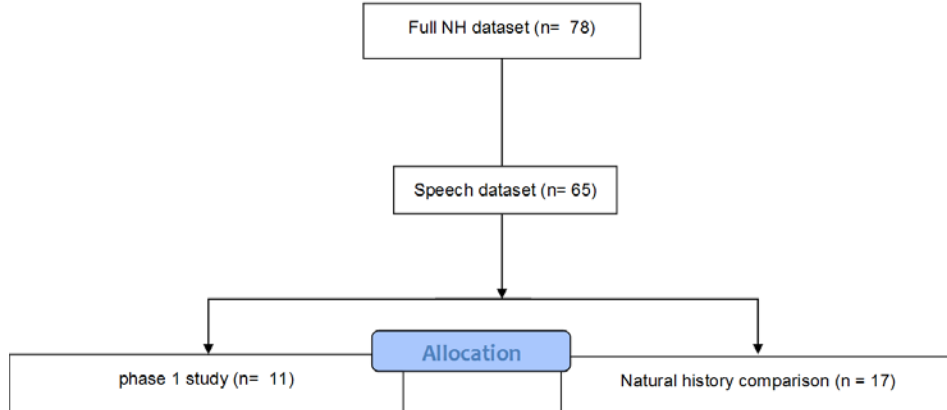
Figure 1: Study Design

All of the phase 1 study and natural history comparison patients originated in natural history study. Of the 78 patients enrolled in the original natural history study, we have data for

65 patients in our dataset, who were then assigned to the phase 1 study or natural history comparison group. Patients in the phase 1 study group, akin to the treatment group in a randomized clinical trial (RCT), received the experimental treatment. The natural history comparison group patients, similar to a control group, received no treatment or placebo.

For our analysis, we used phase 1 study patients with baseline visit information and at least one follow-up visit. To make the groups comparable, we selected patients aged 6-26 with baseline data and at least one follow-up visit for the natural history comparison group. Using this approach, we already have data on all of the patients. This lessoned the difficulty of finding sufficient numbers of treated and non-treated patients, which is especially difficult for orphan diseases.

1.2 HYPOTHESIS OF INTEREST

Our research project involves a potential treatment for speech and swallowing-related neurologic symptoms of Niemann-Pick Type C1 (NPC1) disease. The primary objective is to demonstrate the efficacy of the proposed treatment in reducing speech-related symptoms of NPC1. In order to evaluate efficacy, we will compare values of the outcome variables for the phase 1 and natural history patients. Our null hypothesis is that there is no significant difference between phase 1 and natural history patients for any of the outcome variables. Our alternative hypothesis is that there is a significant difference between the two groups for at least one outcome.

1.3 STATISTICAL APPROACH

We created a generalized linear mixed model to describe the natural history of two different pediatric orphan diseases. Generalized linear mixed models are particularly helpful in situations where we have longitudinal data, but do not have fixed time points or an equal number of data points at each point in time. As a result of this imbalance, multivariate regression is not well-equipped to tackle longitudinal data. Nevertheless, for each patient, outcomes can be well-predicted using linear regression. As a result, we can perform individual-level analysis with linear regression techniques and subsequently utilize multivariate regression functions to compare results for different patients.

A generalized linear mixed model is derived from this combination of models. We start with a linear regression model with an outcome variable, Y. In that model, Y is calculated by adding the error and the product of the values of each predictor variable and its coefficient. We generally assume that the errors are independently and normally distributed with a mean of zero and finite variance.

In the second step, multivariate regression is performed in order to study variability in patients' regression coefficients. . The value of each dependent variable is the sum of a variable to measure random variability and the product of each covariate and its corresponding coefficient. The values of the random variability term are also assumed to be independently and normally distributed with a mean of zero and finite variance, similar to the error in the patient-specific stage.

Generally speaking, estimation of the multivariate regression parameters is the primary goal. The first step is to fit the linear regression model to the observed data in order to calculate

estimates of patient-specific regression coefficients. We then apply the multivariate regression function to those estimates in order to explain variability in various patients' responses.

There are a number of potential problems with this approach. In the first component, the observed data for each patient is summarized separately by an estimate, which results in lost information. The second occurs when we replace the outcome for the multivariate regression model by an estimated value for that outcome. As a result, random variability is added to the model. Finally, the covariance matrix of those estimates is greatly affected by the number of measurements for each patient and when those measurements were taken. These problems are addressed through the use of a general linear mixed model.

With the general linear mixed model, we combine the linear and multivariate regression models into a model with three components. Indeed, "mixed" in the term "mixed models" refers to the mixture of fixed and random effects in the model. The first portion is the product of population-level variables and their regression coefficients, which comes from the multivariable regression model. These population-specific covariates are known as fixed effects because the effect on the individual's outcome is assumed to be identical for all individuals. We will denote the vector containing these variables as X and the vector consisting of the corresponding regression coefficients as β.

The middle term is the product of individual-level parameters and their coefficients, which are the basis for the linear regression model. The patient-specific variables are known as random effects because their values are assumed to vary randomly from patient to patient. The random variability is assumed to be normally and independently distributed with a mean of zero and finite variance. The vector of the patient-specific covariates will be represented by Z and the vector of their coefficients by b.

The final component of our mixed model is the error term vector, ε. We assume that the errors are normally and independently distributed with a mean of zero and finite variance.

Thus, for an arbitrary patient i, a linear mixed-effects model will having the following characteristics:

$Y_i = X_i\beta + Z_ib_i + \varepsilon_i$

b_i is $N(0, D)$

ε_i is $N(0, \sum_i)$

$b_1, , , b_n, \varepsilon_1, , , \varepsilon_n$ are independent,

where Y_i is the vector of outcome values for that patient i. We will assume that Y_i and ε_i have n_i components and that there are N total patients. We will further assume that β, the vector of fixed effects, has p elements and b, the vector of random effects, have q elements. Then, X, the vector of population-level covariates, is a n_i x p matrix and Z, the vector of patient-level covariates, is a n_i x q matrix. Additionally, D is a q x q matrix, where with any (i, j) element in D, $d_{ij} = d_{ji}$ and $\sum_i$ is a n_i x n_i matrix whose parameters are not affected by the value of i and the latter's only effect on $\sum_i$ is its dimension.

Within the world of mixed models, there is the random slope model and the random intercept model. The random slope model is one in which we believe that variability in individuals' slopes is caused by treatment differences and from variability in the patients themselves. If we presume that all the variability in patients' slopes is due to differences in treatment of those patients, then we can omit the random slope component from the model. This mixed-effects model without a random slope component is known as the random intercept model. .

We treated time as a fixed effect, due to limited sample size. Our models also included the intercept as a random effect, as we expect some variability in their effects, even if their magnitudes are held constant.

We utilized the mixed effects model to understand changes over time in the outcomes of interest and to understand the role of the proposed treatment in modifying the course of the disease. To best understand the treatment effect, we fit three mixed models for each outcome variable. The predictor variables for the models consisted of: time (in years) only, time and a treatment indicator, and both main effects and their interaction. The results for each group of models are summarized in Table 2, Table A-1, and Table 3, respectively. We used a significance level of 0.05 for all models and treated a p-value between 0.05 and 0.20 as exhibiting a trend toward significance.

We then estimated the individual slopes representing the rate of change over time using an estimate statement and tested for a difference in these slopes by using the significance level associated with the interaction term in the model. Due to the limited sample size, significance levels are less likely to reach a level of 0.05 or smaller, so overall trends and consistency of the results across endpoints are also examined.

2.0   RESULTS

We utilized data for speech and swallowing-related neurological symptoms of NPC1. We employed scales related to dietary restriction, diminished lip strength, diminished tongue strength, dysarthria, liquid, risk of laryngeal penetration, solid food, speech, and swallowing.

2.1   BASELINE INFORMATION

Our first step was to obtain baseline descriptive information for these outcome variables. We restricted our analysis to natural history patients aged 6-26 at baseline who had multiple visits and all phase 1 patients with multiple visits. We calculated means and standard deviations for all patients, the natural history group, and the phase 1 study group for each predictor. The results are given in Table 1.

Table 1: Descriptive Information for Outcome Variables for Niemann-Pick Type C1 Speech and Swallowing-related Neurological Symptoms and Age

Variable	All Patients Mean (SD)	Natural History (N = 17) Mean (SD)	Phase 1 Study (N = 11) Mean (SD)
AGE (years)	15.14 (5.14)	14.35 (5.48)	16.36 (4.54)
DIET	0.22 (0.64)	0.25 (0.77)	0.18 (0.40)
LIP	2.30 (0.91)	2.13 (1.02)	2.55 (0.69)
TONGUE	2.32 (0.94)	2.18 (1.07)	2.55 (0.69)
DYSARTHRIA	2.43 (0.88)	2.35 (1.00)	2.55 (0.69)
LIQUID	3.65 (0.68)	3.72 (0.77)	3.55 (0.52)
PENETRATION	4.30 (1.27)	4.29 (1.45)	4.30 (0.95)
SOLID FOOD	3.70 (0.54)	3.75 (0.58)	3.64 (0.50)
SPEECH	1.82 (0.39)	1.76 (0.44)	1.91 (0.30)
SWALLOWING	6.21 (1.62)	6.12 (2.00)	6.36 (0.81)
STUDY TIME (years)	1.66 (1.12)	2.19 (1.14)	0.85 (0.29)

All of these measurements were taken at baseline, with the exception time spent enrolled in study. The scale to measure dietary restriction is 0-3, with a higher score indicating greater dietary restrictions. Specifically, it is a composite measure of the patient's diet status with regards to consumption of solids and liquids. The lip and tongue weakness variables use the same scale, which ranges from 1-4, with larger values indicating greater levels of weakness. The speech and dysarthria outcomes are linked. Patients were first scored on their speech fluency. Those who received a grade of 1 (normal), were given a value of 1 (normal articulation) for the dysarthria variable. Patients with a grade of 2 for the speech variable, implying impaired speech, were then also given a value for the dysarthria variable, ranging from 2-4. The penetration variable is a measure of the risk of inhaling liquids or solids, with the risk of inhalation increasing as the value of the variable decreases. Swallowing was measured on a 1-7 scale, with higher values corresponding to greater proficiency with swallowing. Study time was defined as the amount of time that a patient spent in whichever study there were assigned to. Thus, even

though all patients began in the natural history study, for phase 1 patients, we only included their time spent in the phase 1 study.

With the exception of study time, the means for each of the outcome variables and age do not vary significantly by patient group at baseline. The average length of time in the study was markedly longer for natural history patients than phase 1 study patients. In addition, the natural history group has larger SDs with the difference clearly non-significant in some cases, and possibly so in others. In the case of significant differences, that would indicate markedly greater variability in the scores of the natural history patients, compared to the phase 1 study patients, which would not necessarily entail positive or negative consequences.

2.2   SPAGHETTI PLOTS

As a further descriptive measure, we created spaghetti plots to visually display how each patient's values change over time for each outcome of interest. For each outcome variable, we created separate spaghetti plots for the natural history and phase 1 studies groups. Each spaghetti plot incorporated a separate line for each patient. We used years enrolled in the study as the x-axis variable. The spaghetti plots for level of dietary restrictions over time for phase 1 and natural history follow.
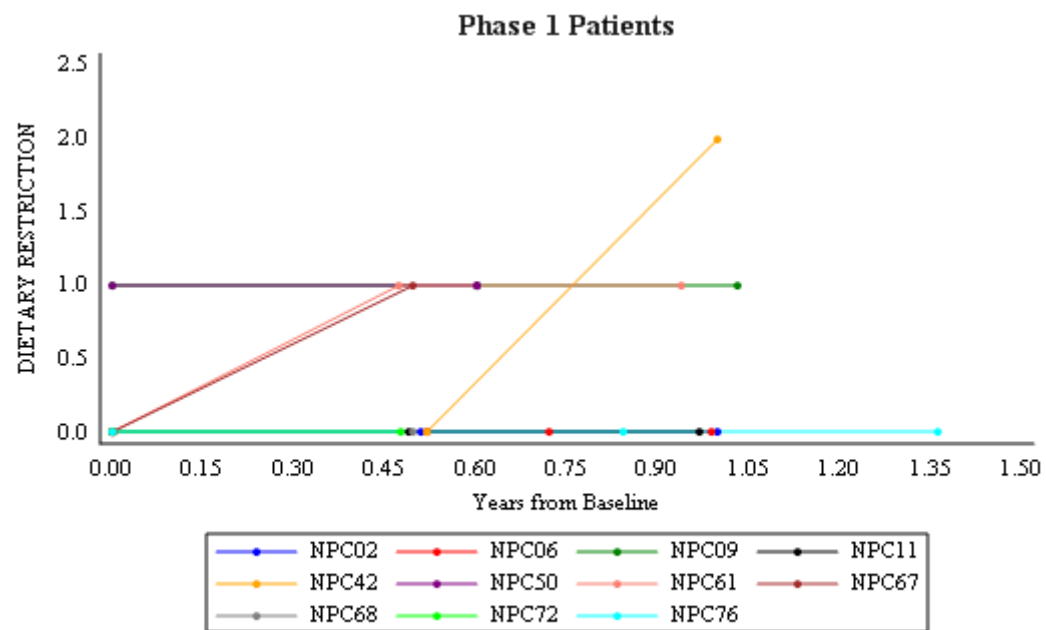
Figure 2. Restrictions on Solid and Liquid Consumption
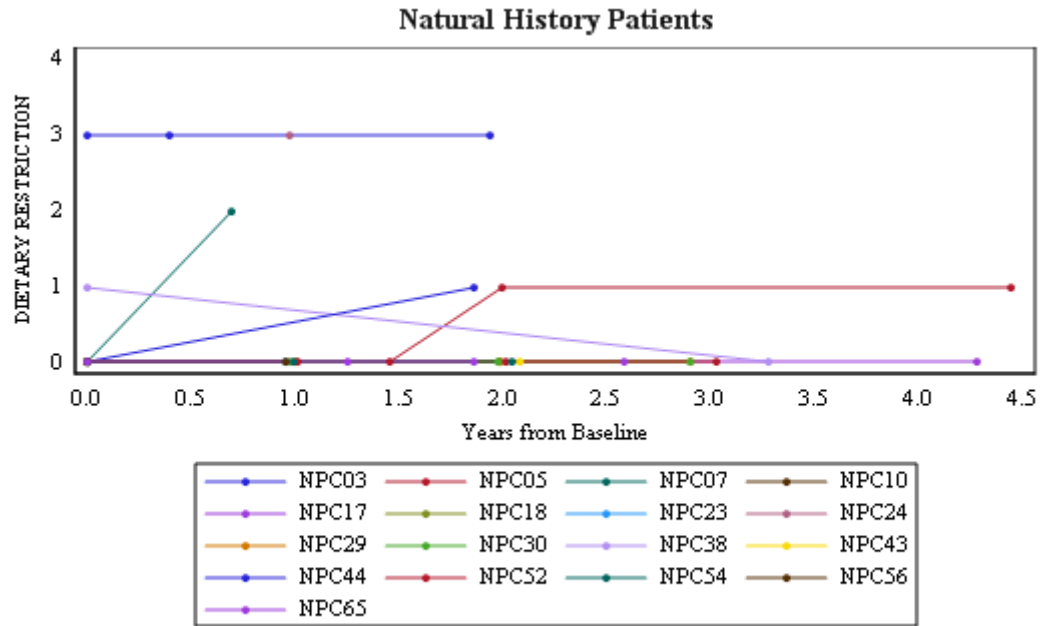
Figure 3: Restrictions on Solid and Liquid Consumption

There are some noticeable differences between the two groups, starting with the difference in time spent in the study. The average amount of time spent in the study for the phase 1 patients was 0.85 years, with a maximum of about 1.36 years. The natural history group mean was 2.19 years, with a maximum of about 4.44.

The dietary restriction variable is a composite measure of the patient's restrictions on consumptions of solids and liquids with a higher score indicating greater dietary restrictions. In the phase 1 group, two patients' scores increased from zero (no dietary restrictions) to one (diet is one level below regular diet status in solid or liquid consistency). These levels are the various

values for the solid food and liquid variables, which are shown in Figure A-1 in the appendix. Another patient's score increased from zero to two (two levels below regular diet status with liquids and solids combined or in either). All of the other phase 1 patients remained constant with a value of zero or one for the duration of their time in the study.

In the natural history group, two patients maintained a score of three (diet is at least two levels below regular diet status with both liquids and solids) for the entirety of their participation in the study. Also unlike the phase 1 study group where there were no improvements, one patient was able to downgrade their evaluation from one to zero. The rest of the natural history group was similar to the phase 1 study group, with two patients worsening from zero to one, one patient deteriorating from zero to two, and the rest remaining constant. Thus, from this descriptive work, it appears that the treatment does not significantly affect the neurological symptoms of Niemann-Pick Type C1 disease. In other words, the natural history and phase 1 patients are experiencing similar severity for the NPC1 neurologic symptoms of interest.

2.3   MIXED MODELS

In order to formally test the significance of treatment and time, for each outcome variable, we created three mixed models: time as the only predictor, time and an indicator for involvement in the phase 1 study, and both main effects and their interaction. The results are summarized in Tables 2, A-1, and 3, respectively.

With regards to the time-only models, diminished tongue strength ($p = 0.033$), solid food ($p = 0.011$), and swallowing ($p = 0.018$) were significant. The coefficients for those three models were 0.080, -0.156, -0.163, respectively, demonstrating changes over time in the full population.

With the solid food scale, a lower score indicates a decreased ability to consume solid foods. Thus, the negative coefficient for the solid food variable points to further restrictions in solid ingestion over time. On the swallowing scale, a lower value also corresponds to decreased physical ability. As a result, the negative coefficient for the swallowing outcome variable similarly implies a decreased capability to swallow. By contrast, with the diminished tongue strength scale, a higher score refers to a worse physical state. Thus, the positive coefficient indicates greater weakness of the tongue muscles over time. Hence, all three variables demonstrate that NPC patients decrease in physical ability as time goes on.

In addition, diminished lip strength and liquid showed a trend toward significance, with p-values of 0.127 and 0.119 and coefficients of 0.047 and -0.090, respectively. Akin to the diminished tongue strength scale, a higher value for the diminished lip strength variable indicates increased muscle weakness. Therefore, the positive coefficient for this variable demonstrates further decrease in physical ability over time. The liquid variable scale, similar to the solid food scale, assigns lower values to patients with greater physical impairment. As a result, the liquid variable's negative coefficient also implies reduced physical abilities. Thus, all five of the significant time-only models point to reduced physical and mental capabilities for NPC1 patients over time. As a result, time appears to have a significant role to play in the severity of NPC1 symptoms when not considering other factors.

Time was a significant predictor of diminished tongue strength, solid food restrictions, and swallowing difficulty in the models with both main effects, with coefficients of 0.082, -

18

0.162, and -0.163 and p-values equal to 0.028, 0.009, and 0.019, respectively. As addressed in the previous paragraph, the signs of all three coefficients indicate a worsening in physical capability over time. Time exhibited a trend toward significance in the diminished lip strength (p = 0.109) and liquid restrictions (p = 0.084) models, with corresponding coefficients of 0.050 and -0.101. Once again, all of the models in which time is a significant or close to significant factor paint a picture of decreased physical ability over time.

The indicator for involvement in the phase 1 study was not a significant factor for any of the outcome variables in this group of models, although there was a trend toward significance in the diminished lip strength (p = 0.172), liquid restrictions (p = 0.143), and speech (p = 0.099) models. Those three models' coefficients were 0.479, -0.409, and 0.209. With the diminished lip strength and speech scales, larger values indicate increased weakness and difficulty, respectively. The liquid restrictions scale works in the opposite direction, with smaller values corresponding to more restrictions on consumption of liquids. The indicator for phase 1 study involvement equals one for a phase 1 patient and zero for a natural history patient. Therefore, the signs of these coefficients point to lower physical abilities for these three outcomes in the phase 1 study group, relative to the natural history group.

In the full models, time was a significant predictor of diminished tongue strength (p = 0.033) and solid food restrictions (p = 0.023) and coefficients equal to 0.083 and -0.147, correspondingly. Again, these results imply increased physical disability as time goes on. Time was also a borderline significant predictor of swallowing difficulties (p = 0.057) and had a trend toward significance in the diminished lip strength model, with a p-value of 0.136. The corresponding coefficients of -0.134 and 0.049 reinforce the notion of physical deterioration as time goes on.

The indicator component was not significant for any of the models, although it trended toward significance in the diminished lip strength ($p = 0.184$) and speech difficulties ($p = 0.179$) models. In those two models, the coefficients for the indicators were 0.472 and 0.179, respectively. Similar to the main effects-only models, this would point to lower physical abilities in the phase 1 patients compared to the natural history patients.

The interaction between time and the treatment indicator was not statistically significant in any of these models, although there was a trend toward significance in the dietary restrictions, liquid restrictions, and swallowing difficulties models, with coefficients of 0.334, -0.324, and -0.373 and p-values of 0.094, 0.134, and 0.138, respectively. The signs of the coefficients all imply that phase 1 patients worsen more quickly over time than natural history patients.

Time was a significant predictor of diminished tongue strength and solid food in all three groups of models. In addition, it was a statistically significant covariate for swallowing in the time-only and both main effects models and borderline statistically significant in the full model. Time also displayed a trend toward significance for diminished lip strength in all three groups of models. In all cases, the signs of the coefficients indicated that physical outcomes worsen as time goes on.

The indicator was never a statistically significant factor, although it exhibited a trend toward significance for diminished lip strength and speech difficulties in both sets of models. The coefficients for the indicator in those cases indicated a worse prognosis for phase 1 patients, compared to natural history patients. The lack of statistical significance, with only a trend toward significance present, makes that conclusion tenuous. Similarly, the interaction displays a trend toward significance for the dietary restrictions, liquid restrictions, and swallowing difficulties

variables, but no statistical significance. As a result, the conclusion that phase 1 patients physically deteriorated over time faster than natural history patients is similarly tenuous.

Table 2: Time-only Mixed Models Results for Outcome Variables for Niemann-Pick Type C1 Speech and Swallowing-related Neurological Symptoms

Outcome	Time	p-value
Diet	0.066	0.221
Lip	0.047	0.127
Tongue	0.080	0.033*
Dysarthria	0.043	0.276
Liquid	-0.090	0.119
Penetration	-0.085	0.356
Solid food	-0.156	0.011*
Speech	0.002	0.935
Swallowing	-0.163	0.018*

*: p-value < .05

Table 3: Full Mixed Models Results for Outcome Variables for Niemann-Pick Type C1 Speech and Swallowing-related Neurological Symptoms

Outcome	Time	p-value	Treatment group	p-value	Interaction	p-value
Diet	0.039	0.481	-0.257	0.429	0.334	0.094
Lip	0.049	0.136	0.472	0.184	0.014	0.904
Tongue	0.083	0.033*	0.423	0.261	-0.018	0.896
Dysarthria	0.043	0.301	0.241	0.474	0.030	0.841
Liquid	-0.077	0.201	-0.245	0.408	-0.324	0.134
Penetration	-0.064	0.507	-0.038	0.941	-0.387	0.283
Solid food	-0.147	0.023	-0.107	0.730	-0.201	0.376
Speech	0.005	0.873	0.179	0.191	0.059	0.589
Swallowing	-0.134	0.057	0.253	0.709	-0.373	0.138

*: p-value < .05

2.4 DIFFERENCE IN SLOPES

In order to test the significance of the interaction between time and the indicator for involvement in the phase 1 study, we used estimate statements to calculate the slopes for each study group in the full group and determined the probability of a statistically significant difference between the slopes. The results are show in Table 4. There is a trend toward significance in the dietary restrictions ($p = 0.094$), liquid restrictions ($p = 0.134$), and swallowing difficulties ($p = 0.138$). In all three cases, the magnitude of the slope for the phase 1 group is greater than for the natural history group. In the case of the dietary restriction outcome, both slopes are positive, whereas for the liquid and swallowing variables, both coefficients are negative. With the dietary restriction variable, larger values indicate greater restrictions on diet.

By contrast, the liquid and swallowing scales work in the opposite manner, with smaller values corresponding to increased limitations. Therefore, these results, however tenuous significance-wise, would seem to imply greater severity of symptoms among phase 1 patients than natural history patients, calling into question the efficacy of the experimental treatment.

Table 4: Differences in Niemann-Pick C1 Disease Swallowing and Speech-related Neurologic Symptom Outcomes by Study Group

Outcome	Phase 1 (SE)	Natural History Slope (SE)	p-value for difference
Diet	0.372 (0.187)	0.039 (0.054)	0.094
Lip	0.063 (0.109)	0.049 (0.032)	0.904
Tongue	0.065 (0.132)	0.083 (0.038)	0.896
Dysarthria	0.073 (0.143)	0.043 (0.041)	0.841
Liquid	-0.401 (0.204)	-0.077 (0.059)	0.134
Penetration	-0.451 (0.343)	-0.064 (0.095)	0.283
Solid food	-0.347 (0.216)	-0.147 (0.062)	0.376
Speech	0.064 (0.105)	0.005 (0.030)	0.589
Swallowing	-0.507 (0.237)	-0.134 (0.069)	0.138

2.5   DIAGNOSTICS

To determine the best model for each outcome variable, we compared the AIC and BIC among each variable's three models. The results are summarized in Table 5.

Table 5: AIC and BIC Values for Mixed Models for Outcome Variables for Niemann-Pick Type C1 Speech and Swallowing-related Neurological Symptoms

	AIC			BIC		
Variable	Time only	Both main effects	Full model	Time only	Both main effects	Full model
Diet	148.9	149.4	147.9	151.6	152.0	150.6
Lip	103.8	102.2	104.7	106.5	104.8	107.3
Tongue	124.2	123.1	125.2	126.8	125.7	127.8
Dysarthria	124.5	124.3	126.3	127.2	127.0	128.9
Liquid	149.5	148.0	147.0	152.1	150.7	149.6
Penetration	220.8	220.2	219.3	223.5	222.9	221.9
Solid food	154.0	154.2	154.5	156.7	156.8	157.2
Speech	46.7	46.3	48.6	49.4	48.9	51.2
Swallowing	211.0	210.0	208.7	213.7	212.7	211.4

AIC and BIC are measures of the quality of the model and can be used to compare the relative validity of nested models. Using these criteria, the full model was the best choice for the dietary restriction, liquid restrictions, risk of laryngeal penetration, and swallowing difficulties variables.   The model with just the main effects was the best choice of the three for the diminished lip strength, diminished tongue strength, dysarthria, and speech difficulties models. The solid food outcome variable was the only dependent variable for which the time-only model had the lowest AIC and BIC values. Thus, it would seem that the model with both main effects might be the best choice because it is simpler than the full model.

We performed diagnostics on the mixed models to assess the assumptions. Due to the presence of a random intercept in our mixed models, we utilized the conditional residuals. We used categorical outcomes, so patterns in the residuals do not indicate a defect in our model. We examined histograms of the conditional residuals and quantile-residual plots in order to determine if the residuals were normally distributed. Most of the residuals did not appear to be clearly non-normally distributed. Some of the residuals did seem to be non-normally distributed, which is probably in part due to our small sample size, a common problem in orphan disease studies.

We used Cook's distance to determine if there were any overly influential points. Using a value of one as a threshold value for influence, we did not encounter any unduly influential points.

3.0   CONCLUSION

We wanted to evaluate the efficacy of a treatment for Niemann-Pick Type C1 disease. We used nine outcome variables to test for a difference: dietary restriction, diminished lip strength, diminished tongue strength, dysarthria, liquid restrictions, risk of laryngeal penetration, solid food restrictions, speech difficulties, and swallowing difficulties. As an exploratory step, we used proc means in SAS and patient-level spaghetti plots to look for differences by treatment group. These descriptive steps did not indicate a significant difference in severity of neurological symptoms of NPC 1 between phase 1 study patients and natural history study patients.

We utilized mixed-effects models to determine which predictor variables significantly affected or had a trend toward significantly affecting each outcome variable. Time is likely a significant factor in tongue strength and difficulty with solid food, due to its statistical significance in all three groups of models for those two variables. Additionally, it appears to also play a significant role in swallowing difficulties, because it was a statistically significant covariate for swallowing in the time-only and both main effects models and borderline statistically significant in the full model. Time is also possibly a factor to consider in lip strength, due to the exhibition of a trend toward significance in all three sets of models.

The indicator was never a statistically significant factor, although it exhibited a trend toward significance for diminished lip strength and speech difficulties in both sets of models that it was involved in, indicating the possibility of a relationship.

The interaction between time and the treatment indicator was not statistically significant in any of these models, although there was a trend toward significance in the dietary restrictions, liquid restrictions, and swallowing difficulties models, so it may be a factor to consider for those symptoms.

For each outcome variable, we compared the AIC and BIC values among the three groups of models. The full model and model with only the main effects were the best choice for four outcome variables each, indicating that each of them is a superior choice to the time-only, but not clarifying the supremacy of either model.

We employed histograms of the conditional residuals and quantile-residual plots to test the assumption of normality of residuals. The conditional residuals for most of the models did not appear to be significantly non-normally distributed. Some plots did show some non-normal behavior, in all likelihood partially due to our small sample size, which is a weak point of our study. Small sample sizes are a common problem in orphan disease studies, and hopefully a future study could be conducted with a larger sample size.

We studied the efficacy of a potential treatment for Niemann-Pick Type C1 disease, which has a lengthy period of decreasing physical ability associated with neurological symptoms, before culminating in early death. Due to the severe toll of the disease on the patient and all people involved in the patient's care, efforts to minimize the burden of NPC1 are certainly relevant to public health.

The limited sample sizes, due to the rarity of NPC1, are a hindrance in evaluating treatment efficacy. Our study design can partially mitigate some of the problems associated with the small sample sizes. From an existing natural history study, we created phase 1 study and natural history groups, similar to the treatment and placebo groups in a more traditional parallel

group randomized clinical trial (RCT). This strategy has certain advantages over a traditional RCT. It uses patients enrolled in an ongoing natural history study for which we already have some data on progression of the disease. Furthermore, it provides a pool of patients from which to recruit phase 1 participants. Additionally, from an analytic point of view, it provides an efficient use of the available data and an alternative design to an RCT that requires greater sample sizes which would be difficult to obtain in our setting of a rare disease. The utilization of our study design can expedite orphan disease research by decreasing the amount of patients necessary to reach meaningful conclusions from efficacy studies. For these reasons, our approach provides a practical alternative in evaluating efficacy of treatment in rare/orphan disease research, a very current public health issue.

Table 6: Both Main Effects Mixed Models Results for Outcome Variables for Niemann-Pick Type C1 Speech and Swallowing-related Neurological Symptoms

Outcome	Time	p-value	Indicator	p-value
Diet	0.065	0.235	-0.089	0.773
Lip	0.050	0.109	0.479	0.172
Tongue	0.082	0.028	0.414	0.262
Dysarthria	0.045	0.252	0.256	0.437
Liquid	-0.101	0.084	-0.409	0.143
Penetration	-0.091	0.326	-0.239	0.617
Solid food	-0.162	0.009	-0.209	0.473
Speech	0.009	0.748	0.209	0.099
Swallowing	-0.163	0.019	0.068	0.919

APPENDIX B: Outcome scales

Table 7: Scale for Swallowing Difficulties among Niemann-Pick Type C1 Patients

Leve1	Description
1	Individual is not able to swallow anything by mouth. All nutrition and hydration are received through non-oral means.
2	Individual is not able to swallow safely by mouth for nutrition and hydration, but may take some consistency with consistent maximal cues in therapy only. Alternative method of feeding is required.
3	Alternative method of feeding is required as individual takes less than 50% of nutrition and hydration by mouth and/or swallowing is safe with consistent use of moderate cues to use compensatory strategies and/or maximal diet restriction.
4	Swallowing is safe, but usually requires moderate cues to use compensatory strategies and/or the individual has moderate diet restrictions and/or still requires tube feeding and/or oral supplements.
5	Swallowing is with minimal diet restriction and/or occasionally requires minimal cueing to use compensatory strategies. The individual may occasionally self-cue. All nutrition and hydration needs are met by mouth at mealtime.
6	Swallowing is safe and the individual eats and drinks independently and may rarely require minimal cueing. The individual usually self-cues when difficulty occurs. May need to avoid specific food items (e.g. popcorn, nuts) or require additional time (due to dysphagia).
7	The individual's ability to eat independently is not limited by swallow functions. Swallowing would be safe and efficient for all consistencies. Compensatory strategies are effectively used when needed.

Table 8: Scale for Scale for Solid Food Restrictions among Niemann-Pick Type C1 Patients

Solid Food Restrictions	Description	Code
Regular	No restrictions	4
Reduced one level (soft)	Meats are cooked until soft with no tough or stringy foods. Might include meats like meat loaf, baked fish, and soft chicken. Vegetables are cooked soft.	3
Reduced two levels (mechanical soft)	Meats are chopped or ground. Vegetables are of one consistency	2
Reduced three levels (puree)	Meats and vegetables are pureed	1

Table 9: Scale for Liquid Restrictions among Niemann-Pick Type C1 Patients

Liquid Restrictions	Description	Code
Regular	Thin liquids, no restrictions	4
Reduced one level (soft)	Mildly thick liquids (e.g. nectar syrup)	3
Reduced two levels (mechanical soft)	Moderately thick liquids (e.g. honey)	2
Reduced three levels (puree)	Extra thick liquids (e.g. pudding)	1

Table 10: Scale for Dietary Restrictions among Niemann-Pick Type C1 Patients

Level of Restriction	Description	Code
Maximum	Diet is two or more levels below a regular diet status in both solid and liquid consistency	3
Moderate	Diet is two or more levels below a regular diet status in solid or liquid consistency (but not both) or diet is one level below in both solid and liquid consistency	2
Minimum	Diet is one levels below a regular diet status in both solid and liquid consistency	1

Table 11: Scale for Lip/Tongue Weakness among Niemann-Pick Type C1 Patients

Description	Code
Normal (no weakness)	1
Mild	2
Moderate	3
Severe	4

Table 12: Scale for Speech Fluency among Niemann-Pick Type C1 Patients

Description	Code
Normal	1
Impaired	2

Table 13: Scale for Risk of Laryngeal Penetration among Niemann-Pick Type C1 Patients

Description	Code
aspiration 50-90%	1
aspiration 10-50%	2
Contrast enters the airway consistently on one or more textures	3
Risk minimal: intermittent laryngeal penetration with retrograde excursion on one or more textures	4
No obvious risk: contrast does not enter the airway	5

Table 14: Scale for Dysarthria among Niemann-Pick Type C1 Patients

Description	Code
Normal (no dysarthria)	1
Mild	2
Moderate	3
Severe	4

I. Baseline.sas

```
proc format;
   value F_PH1F
      0 = "NATURAL HISTORY PATIENT"
      1 = "PHASE 1 PATIENT";
   value F_RIGHT_AGEF
      0 = "NOT BETWEEN 6 AND 26 YEARS OLD"
      1 = "BETWEEN 6 AND 26 YEARS OLD";
   value F_PH1_VISITF
      0 = "NATURAL HISTORY VISIT"
      1 = "PHASE 1 VISIT";
   value F_1ST_PH1_VISITF
      0 = "NOT 1ST PHASE 1 VISIT"
      1 = "1ST PHASE 1 VISIT";
run;


* Definition of program-specific macros;

*create macro to do descriptive analysis for each outcome variable;
%macro DESCRIPTIVE_SPEECH (DATA = , BYVAR = );
proc sort data = &DATA;
   by &BYVAR;
run;

ods rtf file = "E:\Proj\Cydan\Programs\MB
Programs\Out\BASELINE_DESCRIPTIVE\&DATA._BY_STUDY.rtf";
title "&DATA. OUTCOME VARIABLE BASELINE INFORMATION by &BYVAR.";
proc means data = &DATA mean std maxdec = 2;
   class &BYVAR;
   var AGE_NEW DIETARY_RESTRICTION DIMINISHED_LIP_STRENGTH
DIMINISHED_TONGUE_STRENGTH DYSARTHRIA LIQUID PENETRATION_ASP_CORRECTED
SOLID_FOOD SPEECH SWALLOWING_OUTCOME;
run;
```

```sas
proc freq data = &DATA;
   by &BYVAR;
   table DIETARY_RESTRICTION DIMINISHED_LIP_STRENGTH
DIMINISHED_TONGUE_STRENGTH DYSARTHRIA LIQUID PENETRATION_ASP_CORRECTED
SOLID_FOOD SPEECH SWALLOWING_OUTCOME / missing;
run;

proc gchart data = &DATA;
   by &BYVAR;
   vbar DIETARY_RESTRICTION DIMINISHED_LIP_STRENGTH
DIMINISHED_TONGUE_STRENGTH DYSARTHRIA LIQUID PENETRATION_ASP_CORRECTED
SOLID_FOOD SPEECH SWALLOWING_OUTCOME / discrete;
run; quit;

proc univariate data = &DATA;
   by &BYVAR;
   var DIETARY_RESTRICTION DIMINISHED_LIP_STRENGTH DIMINISHED_TONGUE_STRENGTH
DYSARTHRIA LIQUID PENETRATION_ASP_CORRECTED SOLID_FOOD SPEECH
SWALLOWING_OUTCOME;
   histogram DIETARY_RESTRICTION DIMINISHED_LIP_STRENGTH
DIMINISHED_TONGUE_STRENGTH DYSARTHRIA LIQUID PENETRATION_ASP_CORRECTED
SOLID_FOOD SPEECH SWALLOWING_OUTCOME / normal;
run;
ods rtf close;

ods rtf file = "E:\Proj\Cydan\Programs\MB
Programs\Out\BASELINE_DESCRIPTIVE\&DATA._ALL_PATIENTS.rtf";
title "&DATA. OUTCOME VARIABLE BASELINE INFORMATION";
proc means data = &DATA mean std maxdec = 2;
   var AGE_NEW DIETARY_RESTRICTION DIMINISHED_LIP_STRENGTH
DIMINISHED_TONGUE_STRENGTH DYSARTHRIA LIQUID PENETRATION_ASP_CORRECTED
SOLID_FOOD SPEECH SWALLOWING_OUTCOME;
run;

proc freq data = &DATA;
   table DIETARY_RESTRICTION DIMINISHED_LIP_STRENGTH
DIMINISHED_TONGUE_STRENGTH DYSARTHRIA LIQUID PENETRATION_ASP_CORRECTED
SOLID_FOOD SPEECH SWALLOWING_OUTCOME / missing;
run;

proc gchart data = &DATA;
   vbar DIETARY_RESTRICTION DIMINISHED_LIP_STRENGTH
DIMINISHED_TONGUE_STRENGTH DYSARTHRIA LIQUID PENETRATION_ASP_CORRECTED
SOLID_FOOD SPEECH SWALLOWING_OUTCOME / discrete;
run; quit;
```

```sas
proc univariate data = &DATA;
   var DIETARY_RESTRICTION DIMINISHED_LIP_STRENGTH DIMINISHED_TONGUE_STRENGTH
DYSARTHRIA LIQUID PENETRATION_ASP_CORRECTED SOLID_FOOD SPEECH
SWALLOWING_OUTCOME;
   histogram DIETARY_RESTRICTION DIMINISHED_LIP_STRENGTH
DIMINISHED_TONGUE_STRENGTH DYSARTHRIA LIQUID PENETRATION_ASP_CORRECTED
SOLID_FOOD SPEECH SWALLOWING_OUTCOME / normal;
run;
ods rtf close;
%mend DESCRIPTIVE_SPEECH;




*====================================== MAIN PROGRAM
======================================>;



**************************************************************************
*******SPEECH_FINAL
CODE*********************************************************************
**********;
data SPEECH_FINAL;
   set ANALYSIS.SPEECH_FINAL (rename = (DIMINISHED_TONGUE_STENGTH =
DIMINISHED_TONGUE_STRENGTH));
   by SUBJID_NPC;
   if(first.SUBJID_NPC and ASTDY = 0) then ASTDY_YRS = 0;
   if (first.SUBJID_NPC and last.SUBJID_NPC) then delete; *only interested in
patients with at least one follow-up visit;
   if (COHORT ge 1 and COHORT le 4) then F_PH1 = 1;
   else F_PH1=0;
   format F_PH1 F_PH1F.;
   if F_PH1 = 0 then
      if (first.SUBJID_NPC and AGE_NEW ge 6 and AGE_NEW le 26) then /*if NH
patient is in desired age range at baseline*/
      do;
         F_RIGHT_AGE = 1;
         retain F_RIGHT_AGE;
      end;
    else if (first.SUBJID_NPC and (AGE_NEW lt 6 or AGE_NEW gt 26)) then
      do;
         F_RIGHT_AGE = 0;
         retain F_RIGHT_AGE;
      end;
```

36

```
    if (F_PH1 = 0 and F_RIGHT_AGE = 0) then delete; *get rid of NH patients
that aren't between 6 and 26 years old at baseline;
    label F_RIGHT_AGE  = "flag for being between 6 and 26 years old";
    format F_RIGHT_AGE F_RIGHT_AGEF.;
run;

*need to get rid of time spent on natural history study for phase 1 patients;
data SPEECH_NH_PH1_POST2013; *doesn't include natural history records for
phase 1 patients;
  set SPEECH_FINAL;
  by SUBJID_NPC;
  informat DOE mmddyy10.; *get SAS date value for visit dates;
  BASE_SAS_DATE_VALUE = mdy(1,1,2013); *get SAS date value for Jan 1, 2013;

  *create indicator for phase 1 visit;
   if(F_PH1 = 1 and yrdif(BASE_SAS_DATE_VALUE, DOE, 'ACT/ACT') > 0) then
F_PH1_VISIT = 1;
   else F_PH1_VISIT = 0;
   label F_PH1_VISIT = "PHASE 1 VISIT INDICATOR";
   format F_PH1_VISIT F_PH1_VISITF.;

   *determine first phase 1 visit;
   if F_PH1_VISIT = 0 then NUM_PH1_VISITS = 0;
   else NUM_PH1_VISITS = NUM_PH1_VISITS + 1;
   retain NUM_PH1_VISITS;
   label NUM_PH1_VISITS = "NUMBER OF PHASE 1 VISITS";

   if NUM_PH1_VISITS = 1 then F_1ST_PH1_VISIT = 1;
   else F_1ST_PH1_VISIT = 0;
   label F_1ST_PH1_VISIT = "1st PHASE 1 VISIT INDICATOR";
   format F_1ST_PH1_VISIT F_1ST_PH1_VISITF.;

   *calculate time on phase 1 study;
   if F_1ST_PH1_VISIT = 1 then
      do;
         DOE_1ST_PH1_VISIT = DOE;
         retain DOE_1ST_PH1_VISIT;
      end;
    if F_PH1_VISIT = 0 then DOE_1ST_PH1_VISIT = .; *make missing for all non-
phase 1 visits;
   if (F_PH1 = 1 and F_PH1_VISIT = 0) then PH1_YRS = 0;
   else if F_PH1 = 1 then PH1_YRS = yrdif(DOE_1ST_PH1_VISIT, DOE, 'ACT/ACT');
   else PH1_YRS = 0;
   label DOE_1ST_PH1_VISIT = "SAS DATE VALUE FOR 1ST PH1 VISIT";
   label PH1_YRS = "NUMBER OF YEARS ON PHASE 1 STUDY";
```

```sas
    format PH1_YRS 5.2; *5 spaces for number in all: 2 decimal places, 1 for
decimal, which leaves 2 places left of decimal;

  *calculate time that phase 1 patients spent on natural history study;
    if F_PH1 = 1  then NH_YRS = ASTDY_YRS - PH1_YRS;
    else NH_YRS = ASTDY_YRS;
    label NH_YRS = "NUMBER OF YEARS IN NH STUDY";
    format NH_YRS 5.2; *5 spaces for number in all: 2 decimal places, 1 for
decimal, which leaves 2 places left of decimal;


  if (F_PH1 = 1 and F_PH1_VISIT = 1) then ASTDY_YRS = PH1_YRS;
  else if (F_PH1 = 1 and F_PH1_VISIT = 0) then ASTDY_YRS = 0;
  else if F_PH1 = 0 then ASTDY_YRS = NH_YRS;
  *drop PH1_YRS NH_YRS;

  if (F_PH1_VISIT = 0 and F_PH1 = 1) then delete; *delete non-phase 1
records for phase 1 patients;
run;

data TEST;
   set SPEECH_NH_PH1_POST2013;
   keep SUBJID_NPC F_PH1 ASTDY_YRS F_PH1_VISIT PH1_YRS NH_YRS;
run;

data SPEECH_FINAL_BL;
   set SPEECH_NH_PH1_POST2013;
   by SUBJID_NPC;
   if first.SUBJID_NPC;
run;

*get mean and SDs for time on study by study and overall;
data SPEECH_FINAL_FINAL;
   set SPEECH_NH_PH1_POST2013;
   by SUBJID_NPC;
   if last.SUBJID_NPC;
run;

proc sort data = SPEECH_FINAL_FINAL;
   by F_PH1;
run;

ods rtf file = "E:\Proj\Cydan\Programs\MB
Programs\Out\DESCRIPTIVE\YEARS.rtf";
title "TIME ON STUDY by F_PH1";
proc means data = SPEECH_NH_PH1_POST2013 mean max std maxdec = 2;
```

```sas
    class F_PH1;
    var ASTDY_YRS;
run;

title "TIME ON STUDY";
proc means data = SPEECH_FINAL_FINAL mean std maxdec = 2;
    var ASTDY_YRS;
run;
ods rtf close;

%DESCRIPTIVE_SPEECH (DATA = SPEECH_FINAL_BL, BYVAR = F_PH1);




/*proc contents data = SPEECH_FINAL;
run;*/




proc means data = SPEECH_FINAL_PH1;
    var ASTDY_YRS;
run;


proc means data = SPEECH_FINAL_NH;
    var ASTDY_YRS;
run;


/*proc freq data = ANALYSIS.SPEECH_FINAL;
    table SWALLOWING_OUTCOME;
run;*/

data NH_PATIENTS;
    set ANALYSIS.SPEECH_FINAL;
    by SUBJID_NPC;
    if (COHORT ge 1 and COHORT le 4) then delete;
    if last.SUBJID_NPC;
run;

proc means data = NH_PATIENTS;
    var ASTDY_YRS;
        run;
```

II. F_spaghetti_MB.sas

```sas
proc format;
   value F_PH1F
      0 = "NATURAL HISTORY PATIENT"
      1 = "PHASE 1 PATIENT";
   value F_RIGHT_AGEF
      0 = "NOT BETWEEN 6 AND 26 YEARS OLD"
      1 = "BETWEEN 6 AND 26 YEARS OLD";
   value F_PH1_VISITF
      0 = "NATURAL HISTORY VISIT"
      1 = "PHASE 1 VISIT";
   value F_1ST_PH1_VISITF
      0 = "NOT 1ST PHASE 1 VISIT"
      1 = "1ST PHASE 1 VISIT";
run;


data SPEECH_FINAL;
   set ANALYSIS.SPEECH_FINAL;
   by SUBJID_NPC;
   DIMINISHED_TONGUE_STRENGTH = DIMINISHED_TONGUE_STENGTH;
   drop DIMINISHED_TONGUE_STENGTH;
   label DIMINISHED_TONGUE_STRENGTH = "DIMINISHED TONGUE STRENGTH";
   if (first.SUBJID_NPC and last.SUBJID_NPC) then delete;
   if (ASTDY_YRS = . and first.SUBJID_NPC and ASTDY = 0) then ASTDY_YRS = 0;
   if (COHORT ge 1 and COHORT le 4) then F_PH1 = 1;
   else F_PH1 = 0;
   label F_PH1 = "INDICATOR FOR PHASE 1 STUDY";
   format F_PH1 F_PH1F.;

   label F_RIGHT_AGE  = "flag for being between 6 and 26 years old";
   format F_RIGHT_AGE F_RIGHT_AGEF.;
   if (first.SUBJID_NPC and AGE_NEW ge 6 and AGE_NEW le 26) then
      do;
         F_RIGHT_AGE = 1;
         retain F_RIGHT_AGE;
      end;
    else if (first.SUBJID_NPC and (AGE_NEW lt 6 or AGE_NEW gt 26)) then
      do;
         F_RIGHT_AGE = 0;
         retain F_RIGHT_AGE;
      end;
run;
```

```
data SPEECH_NH_PH1_POST2013; *doesn't include natural history records for
phase 1 patients;
   set SPEECH_FINAL;
   by SUBJID_NPC;
   informat DOE mmddyy10.; *get SAS date value for visit dates;
   BASE_SAS_DATE_VALUE = mdy(1,1,2013); *get SAS date value for Jan 1, 2013;

   *create indicator for phase 1 visit;
    if(F_PH1 = 1 and yrdif(BASE_SAS_DATE_VALUE, DOE, 'ACT/ACT') > 0) then
F_PH1_VISIT = 1;
    else F_PH1_VISIT = 0;
    label F_PH1_VISIT = "PHASE 1 VISIT INDICATOR";
    format F_PH1_VISIT F_PH1_VISITF.;

    *determine first phase 1 visit;
    if F_PH1_VISIT = 0 then NUM_PH1_VISITS = 0;
    else NUM_PH1_VISITS = NUM_PH1_VISITS + 1;
    retain NUM_PH1_VISITS;
    label NUM_PH1_VISITS = "NUMBER OF PHASE 1 VISITS";

    if NUM_PH1_VISITS = 1 then F_1ST_PH1_VISIT = 1;
    else F_1ST_PH1_VISIT = 0;
    label F_1ST_PH1_VISIT = "1st PHASE 1 VISIT INDICATOR";
    format F_1ST_PH1_VISIT F_1ST_PH1_VISITF.;

    *calculate time on phase 1 study;
    if F_1ST_PH1_VISIT = 1 then
       do;
          DOE_1ST_PH1_VISIT = DOE;
          retain DOE_1ST_PH1_VISIT;
       end;
    if F_PH1_VISIT = 0 then DOE_1ST_PH1_VISIT = .; *make missing for all non-
phase 1 visits;
    if (F_PH1 = 1 and F_PH1_VISIT = 0) then PH1_YRS = 0;
    else if F_PH1 = 1 then PH1_YRS = yrdif(DOE_1ST_PH1_VISIT, DOE, 'ACT/ACT');
    else PH1_YRS = 0;
    label DOE_1ST_PH1_VISIT = "SAS DATE VALUE FOR 1ST PH1 VISIT";
    label PH1_YRS = "NUMBER OF YEARS ON PHASE 1 STUDY";
    format PH1_YRS 5.2; *5 spaces for number in all: 2 decimal places, 1 for
decimal, which leaves 2 places left of decimal;

   *calculate time that phase 1 patients spent on natural history study;
    if F_PH1 = 1  then NH_YRS = ASTDY_YRS - PH1_YRS;
    else NH_YRS = ASTDY_YRS;
    label NH_YRS = "NUMBER OF YEARS IN NH STUDY";
```

```sas
    format NH_YRS 5.2; *5 spaces for number in all: 2 decimal places, 1 for
decimal, which leaves 2 places left of decimal;


  if (F_PH1 = 1 and F_PH1_VISIT = 1) then ASTDY_YRS = PH1_YRS;
  else if (F_PH1 = 1 and F_PH1_VISIT = 0) then ASTDY_YRS = 0;
  else if F_PH1 = 0 then ASTDY_YRS = NH_YRS;

  if (F_PH1_VISIT = 0 and F_PH1 = 1) then delete; *delete non-phase 1
records for phase 1 patients;
  if (F_PH1 = 0 and F_RIGHT_AGE = 0) then delete; *delete records for NH
patients who aren't aged 6-26 at baseline;
  *drop PH1_YRS NH_YRS;
run;

/*data TEST;
   set SPEECH_NH_PH1_POST2013;
   keep SUBJID_NPC F_PH1 ASTDY_YRS F_PH1_VISIT PH1_YRS NH_YRS;
run;*/

*keep all natural history patients aged 6-26 years at baseline with multiple
visits;
data SPEECH_FINAL_NH;
   set SPEECH_NH_PH1_POST2013  end=EOF;
   by SUBJID_NPC;
   if (COHORT ge 1 and COHORT le 4) then delete;
run;


*keep all phase 1 study patients with multiple visits;
data SPEECH_FINAL_PH1;
   set SPEECH_NH_PH1_POST2013  end=EOF;
   by SUBJID_NPC;
   if (COHORT ge 1 and COHORT le 4);
run;

/*proc freq data = SPEECH_FINAL_PH1; table AGE_NEW; run;*/

*SCI spaghetti plot macro;


*ods path show;

/*proc means data = SPEECH_FINAL_PH1;
   var ASTDY_YRS;
run;*/
```

```
/*proc means data = SPEECH_FINAL_NH;
   var ASTDY_YRS;
run;*/



/*options nomgen nosymbolgen;
ods path show;
ods path work.sciplot(read);*/

*PH1 PLOTS;
ods rtf file = "&PROGFOLD\figures\&PROGNAME\PH1\OUTCOME_VARS.rtf" style =
sciplot;
%MAKE_PLOT_PH1(DIETARY_RESTRICTION, %str(DIETARY RESTRICTION), %str(mean
DIETARY RESTRICTION total (max = 6)));
%MAKE_PLOT_PH1(DIMINISHED_LIP_STRENGTH, %str(DIMINISHED LIP STRENGTH),
%str(mean DIMINISHED LIP STRENGTH total (max = 8)));
%MAKE_PLOT_PH1(DIMINISHED_TONGUE_STRENGTH, %str(DIMINISHED TONGUE STRENGTH),
%str(mean DIMINISHED TONGUE STRENGTH total (max = 8)));
%MAKE_PLOT_PH1(DYSARTHRIA, %str(DYSARTHRIA), %str(mean DYSARTHRIA total (max
= 8)));
%MAKE_PLOT_PH1(LIQUID, %str(LIQUID), %str(mean LIQUIDtotal (max = 8)));
%MAKE_PLOT_PH1(PENETRATION_ASP_CORRECTED, %str(PENETRATION_ASP_CORRECTED),
%str(mean PENETRATION_ASP_CORRECTED total (max = 10)));
%MAKE_PLOT_PH1(SOLID_FOOD, %str(SOLID FOOD), %str(mean SOLID FOOD total (max
= 8)));
%MAKE_PLOT_PH1(SPEECH, %str(SPEECH), %str(mean SPEECH total (max = 4)));
%MAKE_PLOT_PH1(SWALLOWING_OUTCOME, %str(SWALLOWING_OUTCOME), %str(mean
SWALLOWING_OUTCOME total (max = 14)));
ods rtf close;



%readonly(FIGURES);

proc freq data = SPEECH_FINAL_NH;
   table DIETARY_RESTRICTION;
   run;
```

III. F_spaghetti_MB_NH.sas

```sas
proc format;
   value F_PH1F
      0 = "NATURAL HISTORY PATIENT"
      1 = "PHASE 1 PATIENT";
   value F_RIGHT_AGEF
      0 = "NOT BETWEEN 6 AND 26 YEARS OLD"
      1 = "BETWEEN 6 AND 26 YEARS OLD";
   value F_PH1_VISITF
      0 = "NATURAL HISTORY VISIT"
      1 = "PHASE 1 VISIT";
   value F_1ST_PH1_VISITF
      0 = "NOT 1ST PHASE 1 VISIT"
      1 = "1ST PHASE 1 VISIT";
run;

data SPEECH_FINAL;
   set ANALYSIS.SPEECH_FINAL;
   by SUBJID_NPC;
   DIMINISHED_TONGUE_STRENGTH = DIMINISHED_TONGUE_STENGTH;
   drop DIMINISHED_TONGUE_STENGTH;
   label DIMINISHED_TONGUE_STRENGTH = "DIMINISHED TONGUE STRENGTH";
   if (first.SUBJID_NPC and last.SUBJID_NPC) then delete;
   if (ASTDY_YRS = . and first.SUBJID_NPC and ASTDY = 0) then ASTDY_YRS = 0;
   if (COHORT ge 1 and COHORT le 4) then F_PH1 = 1;
   else F_PH1 = 0;
   label F_PH1 = "INDICATOR FOR PHASE 1 STUDY";
   format F_PH1 F_PH1F.;

   label F_RIGHT_AGE  = "flag for being between 6 and 26 years old";
   format F_RIGHT_AGE F_RIGHT_AGEF.;
   if (first.SUBJID_NPC and AGE_NEW ge 6 and AGE_NEW le 26) then
      do;
         F_RIGHT_AGE = 1;
         retain F_RIGHT_AGE;
      end;
    else if (first.SUBJID_NPC and (AGE_NEW lt 6 or AGE_NEW gt 26)) then
      do;
         F_RIGHT_AGE = 0;
         retain F_RIGHT_AGE;
      end;
run;

data SPEECH_NH_PH1_POST2013; *doesn't include natural history records for
phase 1 patients;
```

```sas
    set SPEECH_FINAL;
    by SUBJID_NPC;
    informat DOE mmddyy10.; *get SAS date value for visit dates;
    BASE_SAS_DATE_VALUE = mdy(1,1,2013); *get SAS date value for Jan 1, 2013;

    *create indicator for phase 1 visit;
    if(F_PH1 = 1 and yrdif(BASE_SAS_DATE_VALUE, DOE, 'ACT/ACT') > 0) then
F_PH1_VISIT = 1;
    else F_PH1_VISIT = 0;
    label F_PH1_VISIT = "PHASE 1 VISIT INDICATOR";
    format F_PH1_VISIT F_PH1_VISITF.;

    *determine first phase 1 visit;
    if F_PH1_VISIT = 0 then NUM_PH1_VISITS = 0;
    else NUM_PH1_VISITS = NUM_PH1_VISITS + 1;
    retain NUM_PH1_VISITS;
    label NUM_PH1_VISITS = "NUMBER OF PHASE 1 VISITS";

    if NUM_PH1_VISITS = 1 then F_1ST_PH1_VISIT = 1;
    else F_1ST_PH1_VISIT = 0;
    label F_1ST_PH1_VISIT = "1st PHASE 1 VISIT INDICATOR";
    format F_1ST_PH1_VISIT F_1ST_PH1_VISITF.;

    *calculate time on phase 1 study;
    if F_1ST_PH1_VISIT = 1 then
        do;
            DOE_1ST_PH1_VISIT = DOE;
            retain DOE_1ST_PH1_VISIT;
        end;
    if F_PH1_VISIT = 0 then DOE_1ST_PH1_VISIT = .; *make missing for all non-
phase 1 visits;
    if (F_PH1 = 1 and F_PH1_VISIT = 0) then PH1_YRS = 0;
    else if F_PH1 = 1 then PH1_YRS = yrdif(DOE_1ST_PH1_VISIT, DOE, 'ACT/ACT');
    else PH1_YRS = 0;
    label DOE_1ST_PH1_VISIT = "SAS DATE VALUE FOR 1ST PH1 VISIT";
    label PH1_YRS = "NUMBER OF YEARS ON PHASE 1 STUDY";
    format PH1_YRS 5.2; *5 spaces for number in all: 2 decimal places, 1 for
decimal, which leaves 2 places left of decimal;

    *calculate time that phase 1 patients spent on natural history study;
    if F_PH1 = 1  then NH_YRS = ASTDY_YRS - PH1_YRS;
    else NH_YRS = ASTDY_YRS;
    label NH_YRS = "NUMBER OF YEARS IN NH STUDY";
    format NH_YRS 5.2; *5 spaces for number in all: 2 decimal places, 1 for
decimal, which leaves 2 places left of decimal;
```

```sas
   if (F_PH1 = 1 and F_PH1_VISIT = 1) then ASTDY_YRS = PH1_YRS;
   else if (F_PH1 = 1 and F_PH1_VISIT = 0) then ASTDY_YRS = 0;
   else if F_PH1 = 0 then ASTDY_YRS = NH_YRS;

   if (F_PH1_VISIT = 0 and F_PH1 = 1) then delete; *delete non-phase 1
records for phase 1 patients;
   if (F_PH1 = 0 and F_RIGHT_AGE = 0) then delete; *delete records for NH
patients who aren't aged 6-26 at baseline;
  *drop PH1_YRS NH_YRS;
run;

/*data TEST;
   set SPEECH_NH_PH1_POST2013;
   keep SUBJID_NPC F_PH1 ASTDY_YRS F_PH1_VISIT PH1_YRS NH_YRS;
run;*/

*keep all natural history patients aged 6-26 years at baseline with multiple
visits;
data SPEECH_FINAL_NH;
   set SPEECH_NH_PH1_POST2013  end=EOF;
   by SUBJID_NPC;
   if (COHORT ge 1 and COHORT le 4) then delete;
run;


*keep all phase 1 study patients with multiple visits;
data SPEECH_FINAL_PH1;
   set SPEECH_NH_PH1_POST2013  end=EOF;
   by SUBJID_NPC;
   if (COHORT ge 1 and COHORT le 4);
run;

/*proc freq data = SPEECH_FINAL_PH1; table AGE_NEW; run;*/

*SCI spaghetti plot macro;

ods rtf file = "&PROGFOLD\figures\&PROGNAME\NH\DIETARY_RESTRICTION.rtf" style
= sciplot;
title 'Natural History Patients';
%MAKE_PLOT_NH(DIETARY_RESTRICTION, %str(DIETARY RESTRICTION), %str(mean
PENETRATION_ASP_CORRECTED total (max = 10)));
ods rtf close;

ods rtf file = "&PROGFOLD\figures\&PROGNAME\NH\DIMINISHED_LIP_STRENGTH.rtf"
style = sciplot;
```

```sas
title 'Natural History Patients';
%MAKE_PLOT_NH(DIMINISHED_LIP_STRENGTH, %str(LIP STRENGTH), %str(mean
DIMINISHED LIP STRENGTH total (max = 10)));
ods rtf close;

ods rtf file =
"&PROGFOLD\figures\&PROGNAME\NH\DIMINISHED_TONGUE_STRENGTH.rtf" style =
sciplot;
title 'Natural History Patients';
%MAKE_PLOT_NH(DIMINISHED_TONGUE_STRENGTH, %str(TONGUE STRENGTH), %str(mean
DIMINISHED TONGUE STRENGTH total (max = 10)));
ods rtf close;

ods rtf file = "&PROGFOLD\figures\&PROGNAME\NH\DYSARTHRIA.rtf" style =
sciplot;
title 'Natural History Patients';
%MAKE_PLOT_NH(DYSARTHRIA, %str(DYSARTHRIA), %str(mean DYSARTHRIA total (max =
10)));
ods rtf close;

ods rtf file = "&PROGFOLD\figures\&PROGNAME\NH\LIQUID.rtf" style = sciplot;
title 'Natural History Patients';
%MAKE_PLOT_NH(LIQUID, %str(LIQUID), %str(mean LIQUID total (max = 10)));
ods rtf close;

ods rtf file = "&PROGFOLD\figures\&PROGNAME\NH\PENETRATION_ASP_CORRECTED.rtf"
style = sciplot;
title 'Natural History Patients';
%MAKE_PLOT_NH(PENETRATION_ASP_CORRECTED, %str(PENETRATION ASP), %str(mean
PENETRATION_ASP_CORRECTED total (max = 10)));
ods rtf close;

ods rtf file = "&PROGFOLD\figures\&PROGNAME\NH\SOLID_FOOD.rtf" style =
sciplot;
title 'Natural History Patients';
%MAKE_PLOT_NH(SOLID_FOOD, %str(SOLID FOOD), %str(mean SOLID FOOD total (max =
10)));
ods rtf close;

ods rtf file = "&PROGFOLD\figures\&PROGNAME\NH\SPEECH.rtf" style = sciplot;
title 'Natural History Patients';
%MAKE_PLOT_NH(SPEECH, %str(SPEECH), %str(mean SPEECH total (max = 10)));
ods rtf close;

ods rtf file = "&PROGFOLD\figures\&PROGNAME\NH\SWALLOWING_OUTCOME.rtf" style
= sciplot;
```

```
title 'Natural History Patients';
%MAKE_PLOT_NH(SWALLOWING_OUTCOME, %str(SWALLOWING OUTCOME), %str(SWALLOWING
OUTCOME total (max = 10)));
ods rtf close;
%readonly(FIGURES);
    IV. Mixed_Models.sas


* Definition of program-specific macros;
*macro for mixed models;
%macro MIXED (OUTCOME = , DATA = );

ods rtf file = "E:\Proj\Cydan\Programs\MB programs\OUT\MIXED
MODELS\&OUTCOME..rtf";
ods exclude ModelInfo ClassLevels Dimensions NObs IterHistory CovParms;
ods output SOLUTION_F = solutionF(keep = EFFECT ESTIMATE PROBT);
*TIME-ONLY MODEL;
title "&OUTCOME";
title2 "TIME ONLY";
proc mixed data=&DATA method=reml covtest asycov asycorr ic;
    class SUBJID_NPC;
    id SUBJID_NPC;
    model &OUTCOME =ASTDY_YRS/solution cl influence residual;
    random INTERCEPT /patient=SUBJID_NPC type=un gcorr;
run;

data SOLUTION_F;
    set SOLUTION_F;
    format ESTIMATE 5.3
           PROBT    5.3;
run;




/**PHASE 1 INDICATOR-ONLY MODEL;
title2 "PHASE 1 INDICATOR ONLY";
proc mixed data=&DATA method=reml covtest asycov asycorr ic;
    class SUBJID_NPC;
    id SUBJID_NPC;
    model &OUTCOME =F_PH1 /solution cl influence residual;
    random INTERCEPT /patient=SUBJID_NPC type=un gcorr;
run;*/

*BOTH MAIN EFFECTS MODEL;
title2 "BOTH MAIN EFFECTS";
proc mixed data=&DATA method=reml covtest asycov asycorr ic;
    class SUBJID_NPC;
```

```
   id SUBJID_NPC;
   model &OUTCOME =F_PH1 ASTDY_YRS /solution cl influence residual;
   random INTERCEPT /patient=SUBJID_NPC type=un gcorr;
run;

*BOTH MAIN EFFECTS AND INTERACTION MODEL;
title2 "BOTH MAIN EFFECTS AND INTERACTION";
proc mixed data=&DATA method=reml covtest asycov asycorr ic;
   class SUBJID_NPC;
   id SUBJID_NPC;
   model &OUTCOME =F_PH1 ASTDY_YRS F_PH1 * ASTDY_YRS/solution cl influence
residual;
   random INTERCEPT /patient=SUBJID_NPC type=un gcorr;
   estimate "natural history" INTERCEPT 0 ASTDY_YRS 1 F_PH1 0 F_PH1 *
ASTDY_YRS 0/cl;
   estimate "phase 1" INTERCEPT 0 ASTDY_YRS 1 F_PH1 0 F_PH1 * ASTDY_YRS 1/cl;
run;
ods rtf close;
%mend MIXED;


*==================================== MAIN PROGRAM
=====================================>;
proc format;
   value F_PH1F
      0 = 'NOT IN PHASE 1 STUDY'
      1 = 'IN PHASE 1 STUDY';
   value F_RIGHT_AGEF
      0 = "NOT BETWEEN 6 AND 26 YEARS OLD"
      1 = "BETWEEN 6 AND 26 YEARS OLD";
   value F_PH1_VISITF
      0 = "NATURAL HISTORY VISIT"
      1 = "PHASE 1 VISIT";
   value F_1ST_PH1_VISITF
      0 = "NOT 1ST PHASE 1 VISIT"
      1 = "1ST PHASE 1 VISIT";
run;


* Data steps;
data SPEECH_FINAL;
   set ANALYSIS.SPEECH_FINAL (rename=(DIMINISHED_TONGUE_STENGTH =
DIMINISHED_TONGUE_STRENGTH));
   label DIMINISHED_TONGUE_STRENGTH = "DIMINISHED TONGUE STRENGTH";
run;
```

```sas
*we want all phase 1 patients with multiple visits and NH patients aged 6-26
at baseline with multiple visits;
data SPEECH_FINAL_SUBSET;
   set SPEECH_FINAL;
   by SUBJID_NPC;
   if (first.SUBJID_NPC and last.SUBJID_NPC) then delete;
   if(first.SUBJID_NPC and ASTDY = 0) then ASTDY_YRS = 0;
   if (COHORT ge 1 and COHORT le 4) then F_PH1 = 1;
   else F_PH1=0;
   format F_PH1 F_PH1F.;
   if F_PH1 = 0 then
      if (first.SUBJID_NPC and AGE_NEW ge 6 and AGE_NEW le 26) then /*if NH
patient is in desired age range at baseline*/
      do;
         F_RIGHT_AGE = 1;
         retain F_RIGHT_AGE;
      end;
    else if (first.SUBJID_NPC and (AGE_NEW lt 6 or AGE_NEW gt 26)) then
      do;
         F_RIGHT_AGE = 0;
         retain F_RIGHT_AGE;
      end;
   if (F_PH1 = 0 and F_RIGHT_AGE = 0) then delete; *get rid of NH patients
that aren't between 6 and 26 years old at baseline;
   label F_RIGHT_AGE  = "flag for being between 6 and 26 years old";
   format F_RIGHT_AGE F_RIGHT_AGEF.;

run;

*need to get rid of time spent on natural history study for phase 1 patients;
data SPEECH_NH_PH1_POST2013; *doesn't include natural history records for
phase 1 patients;
  set SPEECH_FINAL_SUBSET;
  by SUBJID_NPC;
  informat DOE mmddyy10.; *get SAS date value for visit dates;
  BASE_SAS_DATE_VALUE = mdy(1,1,2013); *get SAS date value for Jan 1, 2013;

  *create indicator for phase 1 visit;
   if(F_PH1 = 1 and yrdif(BASE_SAS_DATE_VALUE, DOE, 'ACT/ACT') > 0) then
F_PH1_VISIT = 1;
   else F_PH1_VISIT = 0;
   label F_PH1_VISIT = "PHASE 1 VISIT INDICATOR";
   format F_PH1_VISIT F_PH1_VISITF.;

   *determine first phase 1 visit;
   if F_PH1_VISIT = 0 then NUM_PH1_VISITS = 0;
```

```sas
   else NUM_PH1_VISITS = NUM_PH1_VISITS + 1;
   retain NUM_PH1_VISITS;
   label NUM_PH1_VISITS = "NUMBER OF PHASE 1 VISITS";

   if NUM_PH1_VISITS = 1 then F_1ST_PH1_VISIT = 1;
   else F_1ST_PH1_VISIT = 0;
   label F_1ST_PH1_VISIT = "1st PHASE 1 VISIT INDICATOR";
   format F_1ST_PH1_VISIT F_1ST_PH1_VISITF.;

   *calculate time on phase 1 study;
   if F_1ST_PH1_VISIT = 1 then
      do;
         DOE_1ST_PH1_VISIT = DOE;
         retain DOE_1ST_PH1_VISIT;
      end;
   if F_PH1_VISIT = 0 then DOE_1ST_PH1_VISIT = .; *make missing for all non-
phase 1 visits;
   if (F_PH1 = 1 and F_PH1_VISIT = 0) then PH1_YRS = 0;
   else if F_PH1 = 1 then PH1_YRS = yrdif(DOE_1ST_PH1_VISIT, DOE, 'ACT/ACT');
   else PH1_YRS = 0;
   label DOE_1ST_PH1_VISIT = "SAS DATE VALUE FOR 1ST PH1 VISIT";
   label PH1_YRS = "NUMBER OF YEARS ON PHASE 1 STUDY";
   format PH1_YRS 5.2; *5 spaces for number in all: 2 decimal places, 1 for
decimal, which leaves 2 places left of decimal;

  *calculate time that phase 1 patients spent on natural history study;
   if F_PH1 = 1  then NH_YRS = ASTDY_YRS - PH1_YRS;
   else NH_YRS = ASTDY_YRS;
   label NH_YRS = "NUMBER OF YEARS IN NH STUDY";
   format NH_YRS 5.2; *5 spaces for number in all: 2 decimal places, 1 for
decimal, which leaves 2 places left of decimal;



  if (F_PH1 = 1 and F_PH1_VISIT = 1) then ASTDY_YRS = PH1_YRS;
  else if (F_PH1 = 1 and F_PH1_VISIT = 0) then ASTDY_YRS = 0;
  else if F_PH1 = 0 then ASTDY_YRS = NH_YRS;

   if (F_PH1_VISIT = 0 and F_PH1 = 1) then delete; *delete non-phase 1
records for phase 1 patients;
  *drop PH1_YRS NH_YRS;
run;

data TEST;
   set SPEECH_NH_PH1_POST2013;
   keep SUBJID_NPC F_PH1 ASTDY_YRS F_PH1_VISIT PH1_YRS NH_YRS;
run;
```

```
* Proc code;
*run mixed models macro for each outcome of interest;
%MIXED(OUTCOME = DIETARY_RESTRICTION, DATA = SPEECH_NH_PH1_POST2013);
%MIXED(OUTCOME = DIMINISHED_LIP_STRENGTH, DATA = SPEECH_NH_PH1_POST2013);
%MIXED(OUTCOME = DIMINISHED_TONGUE_STRENGTH, DATA = SPEECH_NH_PH1_POST2013);
%MIXED(OUTCOME = DYSARTHRIA, DATA = SPEECH_NH_PH1_POST2013);
%MIXED(OUTCOME = LIQUID, DATA = SPEECH_NH_PH1_POST2013);
%MIXED(OUTCOME = PENETRATION_ASP_CORRECTED, DATA = SPEECH_NH_PH1_POST2013);
%MIXED(OUTCOME = SOLID_FOOD, DATA = SPEECH_NH_PH1_POST2013);
%MIXED(OUTCOME = SPEECH, DATA = SPEECH_NH_PH1_POST2013);
%MIXED(OUTCOME = SWALLOWING_OUTCOME, DATA = SPEECH_NH_PH1_POST2013);




*********************************************END OF MIXED MODELS FOR
SPEECH_FINAL
DATASET**********************************************************************
*********;




proc sort data = ANALYSIS.SPEECH_FINAL;
   by SUBJID_NPC ASTDY_YRS;
run;

title 'SPEECH_FINAL';
proc print data = ANALYSIS.SPEECH_FINAL (obs = 20);
   by SUBJID_NPC;
   var SUBJID_NPC SUBJID_CDA ASTDY_YRS DIETARY_RESTRICTION
DIMINISHED_LIP_STRENGTH DIMINISHED_TONGUE_STENGTH DYSARTHRIA LIQUID
PENETRATION_ASP_CORRECTED SOLID_FOOD SPEECH SWALLOWING_OUTCOME;
run;
ods rtf close;


/*proc freq data = SPEECH_FINAL;
   tables SUBJID_CDA*COHORT;
run;*/
```

```
/*proc freq data = SPEECH_FINAL;
   tables F_PH1 ASTDY_YRS;
run;*/

/* END OF PROGRAM */
```

BIBLIOGRAPHY

1. Zhang M., Zhu C., et al. (2011). The Orphan Disease Networks. The American Journal of Human Genetics. 88,755-766

2. Sardana D., Zhu C., et al. (2011). Drug repositioning for orphan diseases. Briefings in Bioinformatics. VOL 12. NO 4. 346-356

3. Verbeke and Molenberghs (2000), Linear Mixed Models for Longitudinal Data. Springer

4. Temple, RJ. 2013 The Regulatory Pathway for Rare Diseases Lessons Learned from Examples of Clinical Study Designs for Small Populations Presentation presented at: Symposium: Best Practices in Clinical Study Design for Rare Diseases; Washington, D.C.
http://www.fda.gov/downloads/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/UCM352759.pdf

5. Mayo Clinic:dysarthria. C1998-2015. Mayo Foundation for Medical Education and Research; [accessed 2015 Nov 8]. (http://www.mayoclinic.org/diseases-conditions/dysarthria/basics/definition/con-20035008

6. Marsh, M.; McMahon, HT (July 1999). "The structural era of endocytosis". Science 285 (5425): 215–20. doi:10.1126/science.285.5425.215. PMID 10398591. Retrieved 2009-06-19.

7. Yanjanin NM, Vélez JI, Gropman A, et al. Linear Clinical Progression, Independent of Age of Onset, in Niemann-Pick Disease, type C. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics. 2010;153B(1):132-140. doi:10.1002/ajmg.b.30969.

8. Vanier, Niemann-Pick disease type C Orphanet Journal of Rare Disease 2010, 5:16 doi: 10.1186/1750-1172-5-16