

Statistical Analysis Plan

Eosinophilic Esophagitis (EoE) Intervention Trial-Randomized 1 Food Elimination vs. 4 Food Elimination Diet followed by Swallowed Glucocorticoids

PROTOCOL NUMBER: 7805

STUDY DESIGN: Prospective, non-blinded clinical trial

PRINCIPAL

Marc E. Rothenberg, MD, PhD

INVESTIGATOR:

Cincinnati Children's Hospital Medical Center

Division of Allergy & Immunology

3333 Burnet Avenue, MLC 7028

Cincinnati, OH 45229

Phone: (513) 636-6675

Biostatistician:

Lin Fei, PhD

Division of Biostatistics and Epidemiology

1 INTRODUCTION

Pertaining to the objectives of the protocol, this statistical analysis plan (SAP) specifies, prior to finalization of study database, a comprehensive description of strategies and statistical techniques that are planned to assess clinical efficacy and safety of 1 food elimination (milk elimination alone, 1FED) versus 4 food elimination (milk, egg, wheat, and soy elimination, 4FED) diet intervention of Eosinophilic Esophagitis (EoE), and the following Glucocorticoids treatment.

The primary objective of this protocol is to conduct a prospective, non-blinded randomized trial comparing patient reported outcomes scores in novel empiric elimination dietary therapies in eosinophilic esophagitis (EoE) in order to assess the therapeutic viability of minimally restrictive empiric elimination diets as related to measures of patient's symptoms, quality of life, and pathology. The primary efficacy outcome will be the change in the standardized and validated Pediatric Eosinophilic Esophagitis Symptom Score v2 (PEESS).

Moreover, we also aim to assess the response to topical swallowed steroids in those who were non-responders to empiric dietary therapy regimens used during the initial phases of this study.

A variety of secondary endpoints, including pre- and post-therapy peak eosinophil counts per hpf and percentages of patients that attain histologic remission (<15 eosinophils/hpf), along with detailed histologic and endoscopic scores, novel biomarkers, and patient related outcome (PRO) measures will be analyzed between treatment groups.

In addition, we will assess the possible value of a variety of clinical characteristics and biomarkers in terms of predicting response to therapy. Relationship between PEES response and histologic response will be investigated as well to better understand PRO measures in predicting disease status.

2 STUDY DESIGN

This trial is comprised of 2 phases with each phase lasting 3 months. In the first phase, all participants will be randomized 1:1 to either 1FED or 4FED therapeutic diet. At the end of this phase, an EGD will be performed to assess histologic remission, which represents one of the primary current tools used in evaluating EoE disease activity during clinical trials and daily EoE care. Attainment of remission (esophageal eosinophil counts <15 eosinophils/high powered field) after Phase I will result in study discontinuation and maintenance of the “successful” dietary therapy. Dietary therapy non-responders who were on 4FED during Phase I will receive topical swallowed steroids for 3 months while on an unrestricted diet (Phase 2) followed by EGD with esophageal biopsies. Participants receiving SGC will return to an unrestricted diet (i.e. stop 4FED) prior to initiating SGC therapy. Dietary therapy non-responders who were on 1FED during Phase 1 will proceed to 4FED for 3 months (Phase 2) followed by EGD with esophageal biopsies.

3 STUDY OBJECTIVES

The purpose of this study is to evaluate the following objectives:

1.1 Primary Objective

To perform a prospective, randomized, non-blinded trial that determines patient reported outcomes scores

following 1FED (milk elimination alone) vs. 4FED (milk, egg, wheat, and soy avoidance) and their relative efficacy (Phase I).

- The primary endpoint for this study will be the change in PEES scores from baseline to post therapy assessed after the end of the 3 month Phase I.

1.2 Secondary Objectives

- To extend Phase I of this study with a prospective non-blinded trial that determines the rate of remission after treatment with 4FED (in participants failing 1FED in Phase I) or SGC (in participants failing 4FED in Phase I) (Phase II).
- To evaluate the effect of each therapy on histological remission (defined as a post therapy eosinophil count of <15 eosinophils/HPF) and by a variety of changes in eosinophils, including (a) pre- and post-therapy peak eosinophil counts; (b) partial remission (2-14 peak eosinophils/hpf); and (c) complete histological remission (≤ 1 peak eosinophils/hpf).
- To evaluate the effect of each therapy by utilizing the histology scoring system (HSS) created to express the severity and extent of other abnormalities in the gastrointestinal (GI) tract that often accompany eosinophilic inflammation.
- To determine the impact of each therapeutic intervention on biomarkers using the EoE Diagnostic Panel (EDP).
- To evaluate the clinical and psychosocial effect of each therapy using the Peds Quality of Life Inventory Eosinophilic Esophagitis Module (PedsQL EoE) to assess EoE symptoms and problems/feelings related to eating.
- To determine if any clinical parameters predict response to therapeutic intervention.
- To determine if any biomarkers including component-resolved diagnostics (CRD) [serum IgE component testing] predict response to therapeutic intervention.
- To determine if skin testing, in the form of prick and patch testing, predicts response to therapeutic intervention.
- Also to assess relationship between primary PEES score changes and histological scores (HSS).

4 STUDY PARTICIPANTS

4.1 Number of Participants

A total of 292 participants between the ages of 6-17 years were to be enrolled. Because of unexpectedly slow

enrollment, the actual number of participants will be around 50 total. The DMCC at the University of South Florida Health Informatics Institute will serve as the central data management center, with CCHMC serving as the primary site and the central IRB. Additional sites include: Children's Hospital of Colorado, Riley Hospital for Children, Children's Hospital of Philadelphia, University of California at San Diego/Rady Children's Hospital, Northwestern University/Lurie Children's Hospital, Tufts Medical Center, Mt. Sinai Kravis Children's Hospital, Arkansas Children's Hospital, and the University of North Carolina.

5 ANALYSIS VARIABLES

5.1 Primary Efficacy Endpoint:

- Pediatric EoE Symptom Score (PEESS, version 2). The PEES total score change from pre-treatment to post-treatment is the primary efficacy endpoint. The Parent Report for Children and Teens PEES will be used as the primary.

The PEES total score comprises of two types: frequency and severity. A total of 11 frequency questions and 9 severity questions are there respectively. The raw scores of 0 to 4 will be converted to 0 to 100 by multiplying each score by 25. The total metric score is calculated as the mean of non-missing items from the 20 total possible. If more than 50% (10 items) are missing for a particular patient on a visit, the score will not be calculated, rather set as missing.

There are four domains: Dysphagia, GERD, Nausea/vomiting, and Pain. Domain specific scores are calculated similarly.

Dysphagia:

How often does your child have trouble swallowing?

How bad is your child's trouble swallowing?

How often does your child feel like food gets stuck in his/her throat or chest?

How bad is it when your child gets food stuck in his/her throat or chest?

How often does your child need to drink a lot to help swallow food?

How bad is it when your child needs to drink a lot to help swallow food?

How often does your child eat less than others?

How often does your child need more time to eat than others?

GERD:

How often does your child have heartburn (burning in the chest, mouth, or throat)?

How bad is your child's heartburn (burning in the chest, mouth, or throat)?

How often does your child have food come back up in his/her throat when eating?

How bad is it when food comes back up in your child's throat?

Nausea/vomiting:

How often does your child vomit (throw up)?

How bad is your child's vomiting (throwing up)?

How often does your child feel nauseous (feel like throwing up but does not)?

How bad is your child's nausea (feeling like throwing up, but does not)?

Pain:

How often does your child have chest pain, ache, or hurt?

How bad is your child's chest pain, ache, or hurt?

How often does your child have stomach aches or belly aches?

How bad are your child's stomach aches or belly aches?

5.2 Secondary Efficacy Endpoints:

- Patient Related Outcome (PRO) metrics pre- and post-treatment
 - PedsQL EoE Module
 - Generic Health Related Quality of Life measures
 - PROMIS General Health Questionnaire and PedsQL
 - Individual domain PEES scores
- Histology and endoscopy measures
 - Percent of participants who achieve histologic remission (<15 peak eosinophils/HPF) post therapy assessed after the end of the 3 months
 - Division of patients into complete (all peak eosinophils ≤1/HPF) or partial (the worse of 3 scores is from 2 to 14) remission, and non-remission (at least one > 14/HPF)
 - Histology Scoring system
 - Endoscopic scoring system (EREFS)
- Biomarkers pre- and post-treatment
 - Assessment of Eosinophil Diagnostic Panel EoE Score
- Clinical parameters and biomarkers to predict responsiveness to diet and SGC therapy

- Correlation of clinical characteristics to response to therapy
- Assessment of predictive value of skin prick and patch allergy skin testing
- Correlation between presence of positive milk component allergy testing to response to 1FED
- Predictive biomarker-analysis of EoE transcriptome from EDP analysis on baseline biopsies in correlation to response to treatment.

5.2.1.1 Secondary End-Point Analysis: Biomarkers

In this sub-aim, we hypothesize that molecular biomarkers will change (improve) with each treatment phase, at least in part. Furthermore, we hypothesize that biomarkers may reveal (a) novel insight into therapeutic resistance and/or relapse; (b) reveal patient sub-groups not identified by histology; and (c) ultimately prove to be readily usable biomarkers in clinical settings and possess personalized medicine value. We will focus on molecular profiling of esophageal genes with the Eosinophil Diagnostic Panel (EDP), a set of 94 mRNA transcripts that have been established to differentiate EoE from controls (e.g. GERD and normal individuals), and to identify exposure and transcriptional signaling to SGCs²⁸. Essentially, the associated EDP algorithm renders the raw Ct values of each embedded gene after real-time PCR in a way that the upregulated genes and down-regulated genes are summed up individually. A quantitative “EoE score” is derived from the ΔCt (normalized to GAPDH) summation to reflect disease severity and for statistical analysis. Utilizing different gene sets, EDP is also able to predict steroid exposure and remission status based on the same algorithm²⁸. In this case, the EoE score will be calculated at baseline endoscopy and in the post-treatment biopsies to assess change with treatment.

The change in “EoE score” will be analyzed between responders and non-responders based on histology assessment (<15 eosinophils/HPF) at the end of Phase I via an ANOVA model, after proper transformation of the score, e.g. logarithmic if needed. If normality assumption is seriously violated, a non-parametric method of Wilcoxon test will be applied.

5.2.2 AIM 2: Clinical Parameters & Biomarkers That Predict Therapeutic Response

5.2.2.1 Predictive Biomarker-Analysis Of EoE Transcriptome

In this sub-aim, we aim to identify biomarkers that may predict responsiveness to any or all of the treatment

interventions associated with this trial. As such, in this sub-aim, we will perform the EDP analysis on baseline biopsies, focused on identifying transcript levels that differentiate patients that respond to each of the interventions.

Baseline esophageal gene expression was compared between responders and non-responder patients by statistically screening two cohorts. A list of genes with $p < 0.05$ (or $FDR < 0.05$), fold change > 2.0 will be identified.

5.3 End of Therapy (EOT) And Follow-Up

The primary efficacy endpoint (e.g. PEES v2 scoring) will be determined at the end of Phase I & II (weeks 12 & 25) at which time all participants will undergo EGD with biopsy. Histologic scoring of biopsy samples obtained during the end of Phase I endoscopy, will be used to determine the treatment pathway to be followed. Participants who achieve histologic remission at the end of Phase I will have completed the study and will resume follow-up care with their primary GI and/or Allergy physician(s). Those whose disease remains active in terms of histologic disease activity (eosinophil count ≥ 15 per HPF) will go on to Phase II (twelve week treatment with 4FED for Phase I 1FED non-responders **OR** SGC for Phase I 4FED non-responders).

6 STATISTICAL METHODS

6.1 Statistical Analyses: Aim 1

Demographic and patient characteristics will be summarized for the two treatment groups using mean \pm standard deviation or median [interquartile range] for continuous variables and frequency and percentages for categorical variables. Treatment comparisons will be performed using the two-sample t-test or Wilcoxon Rank sum test for continuous variables and the Fisher's Exact test for categorical variables. All tests will be conducted at $\alpha=0.05$, including the primary and secondary analyses.

There was going to be two sets of analysis populations in terms of efficacy assessment, the intent-to-treat and per protocol. However, since the total enrollment has been reduced to about 50, all patients randomized to Phase I of the study who have at least one clinical observation post randomization will be considered in intent-to-treat set. Only this set of population will be included in the statistical analyses. Originally, we plan to define patients who are at least 80% compliant to dietary assignment during Phase I, and who do not have any major protocol violations as in per protocol set. Now all analyses will be based on intent-to-treat population. Patients who do drop out of the trial before the end of Phase I will have their PEES done at the time of dropout. The primary endpoint is overall PEES score change from pre-treatment to the end of Phase I. If normality assumption is approximately valid, a linear mixed effect model will be applied to the primary efficacy variable,

account for site random effect, and other potentially important covariates, e.g. gender, race, and PPI use etc. Because of non-normal nature of the scores in previous research, a Wilcoxon rank sum test may be applied to compare between two treatment groups when normality is violated.

As a secondary endpoint analysis, remission rates (as defined by the percentage of participants achieving histologic remission (eosinophils less than 15/HPF) at the end of Phase I will be analyzed in the following 3 steps.

- (1) Remission rates will be computed along with their 95% CIs for each diet group separately;
- (2) Remission rates will be compared between the two treatment groups using the generalized linear mixed effects model with the logit link (for binary outcome) that accounts for the clustering within sites. Important covariates will be analyzed in this model including gender, race, PPI use, atopy, age, and anthropomorphic features. Treatment group comparisons for secondary endpoints will be performed using the generalized linear mixed effects model using the link that is appropriate for the endpoint (i.e. logit for binary, log for Poisson count, identity for continuous). These models will account for the clustering within sites and will include the same covariates as for the primary analysis.
- (3) The remission rate for non-responders who proceed to Phase II will be summarized separately depending on the treatment patients received during Phase I and Phase II. The 95% confidence interval for the rates will also be provided.

6.2 Statistical Analyses: Aim 2

To evaluate the ability of biomarkers and phenotypic characteristics to predict response to therapeutic intervention, multivariable logistic regression analysis may be conducted with histological evaluation of remission/no remission as the dependent variable and biomarker levels as the independent variables. Multicollinearity of the independent variables will be assessed. The area under the receiver operating characteristic curve (AUROC) will be estimated along with 95% confidence interval. This analysis will be conducted based on the remission status at the end of Phase I. It may also be repeated at the end of Phase II by including those patients who also responded to SGC.

Additionally, a correlation analysis between PEES scores and biomarkers will be performed. Spearman's correlation coefficient and its confidence intervals will be computed to further assess predictive utility for the biomarkers of interest.

Missing data will be handled in the following manner. For the primary efficacy measure of PEES, each patient is expected to complete the questionnaire; even if the patient terminates participation early (dropout) for any reason. The total PEES score is the average of 20 item scores, as long as there are at least 10 item response non-missing, as described previously. No imputation of missing total PEES will be performed. The early dropout reasons will be recorded, and rates will be assessed between treatment groups for potential patterns.

For all other secondary efficacy and demographic variables, missing data will not be imputed. However, if a participant's biomarker panel is partially missing, multiple imputation may be applied for missing markers in predictive modeling analyses.

Heterogeneity of treatment effect is referred to as the variability in the direction and magnitude of individual or subgroup treatment effects that are non-random. Descriptive heterogeneity of treatment effect analyses will first be carried out for subgroups by gender, by study site, and by age of subject. These analyses are included in the analysis plan above for primary and secondary efficacy analyses, where proper estimates of treatment effect will be extracted. Because of the small sample size in this study, any hypothesis testing of treatment heterogeneity is considered exploratory. P-values generated from multiple comparisons are not adjusted. The goal is to identify possible heterogeneity among subgroups for further confirmation. The interaction of demographic and baseline characteristics with treatment group effects may be tested in the generalized linear mixed effects models to assess heterogeneity.