# CIDA Data Analysis Plan

## Cover Page

## Introduction

In order to provide context for the analysis both to the investigator and other statisticians, briefly describe the project itself. As a start, consider looking at the early text in the research protocol submitted to COMIRB.

This section need not be extensive. A paragraph or two is usually sufficient.

Example:

*The data for this project come from the Childhood Asthma Management Program (CAMP), a clinical trial carried out in children with asthma. The trial was designed to determine the long-term effects of 3 treatments (budesonide, nedocromil, or placebo) on pulmonary function as measured by normalized FEV1 over a 5-6.5 year period. The design of CAMP was a multicenter, masked, placebo-controlled, randomized trial. A total of 695 children (210 in the budesonide group, 210 in the nedocromil group and 275 in the placebo group) aged 5-12 years were enrolled between December of 1993 and September of 1995. The primary outcome of the trial was lung function as measured by the Forced Expiratory Volume at 1 second (FEV1). Secondary outcomes were bronchial responsiveness to methacholine, need for beclomethasone due to asthma symptoms, termination of assigned treatment due to cessation of symptoms, and as Asthma morbidity (frequency and severity of asthma symptoms, frequency and magnitude of PEFR measurements less than 80% of personal best, prn use of supplemental inhaled albuterol, nocturnal awakenings, days of limited activity, and absences from school, courses of steroids).*

## Hypotheses and/or research questions and/or specific aims

In order to link the overall project goals to the specific statistical analyses, describe, as specifically as possible, a clearly defined research question. For most projects, this will be some sort of research hypothesis or statistical hypothesis. This section should have enough detail to allow a person to see how the data (described below) can be employed to address the overall goals of the project (described above).

Example:

***Hypothesis****: Lung function in children with asthma as measured by FEV is significantly better in children treated with budesonide or nedocromil after 6 years.*

***Secondary Hypothesis:*** *Whether or not the parent smokes has a significant effect on the effectiveness of the treatment.*

## Data description

List variables to be used in the analysis. This should include a description of the source table/dataset/calculation. For small projects, this list may be short. For larger, more complex projects, this may include algorithms or table-linking strategies. Where relevant, this list should include:

* Outcomes
* Exposures/predictors
* Potential covariates/confounders/etc.
* Other variables of interest

Example:

***Exposure:*** *Budesonide, nedocromil or placebo.*

***Outcome:*** *Pre-bronchodilator Forced Expiratory Volume at 1 second (FEV1). There are many observations for each individual ranging from 1 – 18. Some individuals have observations extending to 10 years of follow-up. Outcome will be FEV measurements at 72 months (6 year mark), or the last measurement for those with less than 6 years of follow-up. Those with only baseline measurements will be excluded.*

***Possible Covariates:*** *Gender (male, female), Ethnicity (Black, Hispanic, Other, Non-Hispanic White), age at enrollment, age of current home in years, any pets, uses a woodstove, uses a dehumidifier, whether any parent smokes, and whether anyone (including visitors) smokes in the home. Information on age of the home, pets, woodstoves, dehumidifiers and smokers was not collected at every visit and had the potential to change over the course of the study, therefore whether or not the child ever reported the exposure during the first 6 years of follow-up will be analyzed. Age of home will be broken into two categorical variables whether or not the child ever reported living in a home older than 50 years, and older than 100 years, over the first 6 years of follow-up.*

## Analysis Methods

For the project as a whole, and for each research question, aim, or hypothesis, describe the proposed analytic method and any necessary details for its implementation. At a minimum, this should address the hypotheses mentioned above. For further detail, describe any follow up or secondary analyses, any assessments of assumptions, and sensitivity or subgroup analyses. This section should have enough detail that any statistician can recreate the analysis using any statistical software and the source data.

Example:

*A univariable analysis between outcomes and exposure will be performed using general linear regression. Assumptions of linearity, homoscedasticity, and normality will be confirmed using diagnostic plots. Relationship between possible covariates and outcomes will be analyzed. Covariates with a significant association with the outcome will be considered for a multivariable model. P values <0.05 will be considered significant. Baseline FEV, gender, and age will be included in the adjusted model based on their biological associations with the outcome. Since participants were randomly assigned, there should be no relationship between possible covariates and exposure, this will be confirmed using χ2/fisher’s exact for categorical covariates and ANOVA for continuous covariates. The secondary hypothesis will be evaluated with a partial F- test. The full model will include treatment group, whether the parent ever reported smoking and the interaction of the two, the reduced model will be the full minus the interaction term.*

## Proposed tables/figures

To supplement the analysis methods, describe the potential statistical outputs to be delivered to the investigator. These should include some form of “Table 1” describing the data itself as well as all intermediate or final tables addressing the research goals. While table shells are entirely appropriate, it is sufficient to just describe the proposed tables or graphs.

Example:

*Analyses that have significant overall F statistics will be considered for inclusion in tables.*

*Table 1. Characteristics of Study Participants by Treatment Group*

## Other information

As an optional useful supplement, include timelines, project contacts, and the intended final products (article, abstract, etc.). The included information should help focus how the results above will be described, to whom they are targeted, and how quickly they need to be produced.

Example:

* + *Exploratory Analysis – May 1st*
  + *Preliminary Report – May 10th*
  + *Comprehensive Report – May 20th*
  + *Statistical Methods/Results Manuscript Ready – May 30th*
  + *PI: Dr. Doolittle* [*doctor.doolittle@ucdenver.edu*](mailto:doctor.doolittle@ucdenver.edu)