# Minding the p-Values and Quartiles:

# Data Analysis, Research Study Design and the IRB

**Don Allensworth-Davies, MSc**
**Research Manager, Data Coordinating Center**
**Boston University School of Public Health**
**IRB Reviewer, Panel Purple**

*“I always find that statistics are hard to swallow and impossible to digest. The only one I can ever remember is that if all the people who go to sleep in church were laid end-to-end they would be a lot more comfortable.”*

**— Mrs. Robert A. Taft**

# Data Analysis and Human Subjects Protection

**Research means a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge.**

***§46.102(d) Definitions, Federal Regulations Title 45, Public Welfare, Department of Health and Human Services, Part 46, Protection of Human Subjects***

# Data Analysis and Human Subjects Protection

**Risks to subjects are minimized: (i) By using procedures which are consistent with sound research design and which do not unnecessarily expose subjects to risk, and (ii) whenever appropriate, by using procedures already being performed on the subjects for diagnostic or treatment purposes.**

***§46.111 Criteria for IRB Approval of Research, Federal Regulations Title 45, Public Welfare, Department of Health and Human Services, Part 46, Protection of Human Subjects***

# Good Study Design and Analysis Plans as Features of Ethical Research with Humans

BY JANICE M. WEINBERG AND KEN P. KLEINMAN

**IRB: Ethics and Human Research, September-October 2003**

**Available on the IRB Website**

# What the IRB Considers in Reviewing an Analysis

- Is the sample size adequate to answer the research question?
  - Is the sample size large enough to ensure sufficient data to answer the research question?
  - Is the sample size justified?

# What the IRB Considers in Reviewing an Analysis

- Is the analysis plan clearly described and adequate to answer the research question?
  - Is it clear which variables will be analyzed?
  - Is it clear what comparisons will be made?
  - Are the methods being used appropriate for the study design and the information that was collected?
  - How will the researcher know whether or not the study objectives have been met?

# What Methods Are Commonly Used for Analysis?

- **Methods Used Are Dictated by:**
  - Study design
  - Type of information being collected (e.g., single observation vs. multiple observations, categorical vs. continuous)
  - Whether bias or confounding is expected
- **Most Common Study Designs**
  - **Cross-sectional studies:** surveys, questionnaires collecting information at a single point in time
  - **Longitudinal studies:** Information is collected on subjects over time

# What Methods Are Commonly Used for Analysis?

- **Most Common Study Designs (cont'd)**
  - **Randomized clinical trials:** subjects are randomly assigned to treatment(s)/ interventions or placebo/standard of care
  - **Case-control studies:** all cases with a disease or condition are identified and controls are randomly selected for comparison
  - **Case-crossover studies:** Each subject acts as their own control and information is collected on each subject under a treatment and placebo condition

# What Methods Are Commonly Used for Analysis?

- Different statistical tests are used for categorical vs. continuous variables
- Different regression models are used to adjust for bias/confounding depending on how the study outcome is measured

# What Methods Are Commonly Used for Analysis?

- **Cross-Sectional Studies:**
  - Basic descriptive statistics including counts, averages (means), midpoints (medians), standard deviations and percentages (proportions)
  - Categorical Data: Chi-square or Fisher's exact test (for comparisons with cell counts < 5) with p-values (traditional cut-off for significance $p < 0.05$)

# What Methods Are Commonly Used for Analysis?

- **Cross-Sectional Studies:**
  - **Continuous Data:**
    - **If NORMALLY distributed: t-test, analysis of variance (ANOVA) with p-values, Pearson correlation**
    - **If NON-NORMALLY distributed: Wilcoxon or Kruskal-Wallis test with p-values, Spearman correlation**
  - **These methods may also be used with other study designs**

# What Methods Are Commonly Used for Analysis?

- **Longitudinal Studies:**
  - Incidence rates, cumulative incidence (%), rate (risk) ratios with 95% confidence intervals and p-values
  - **IMPORTANT:** Longitudinal studies almost always involve repeated observations per subject; methods for "repeated measures" or, in the case of only two observations per subject, "pairs" should be described

# What Methods Are Commonly Used for Analysis?

- Longitudinal Studies:
  - If time to an event is outcome of interest, may also include
    - Survival Curves (Kaplan-Meier)
    - Cox Proportional Hazard Models

# What Methods Are Commonly Used for Analysis?

- Randomized Clinical Trials:
  - **IMPORTANT:** An "intention-to-treat" analysis should be described
    - *"Once randomized, always analyzed"*
    - Preserves the benefits of randomization

# What Methods Are Commonly Used for Analysis?

- **Case-Control Studies:**
  - Odds, odds (risk) ratios with 95% confidence intervals and p-values
  - **IMPORTANT:** If the study is **MATCHED** special methods to "take the matching into account" should be described (e.g., conditional logistic regression)

# What Methods Are Commonly Used for Analysis?

- Case-Crossover Studies ("Within-Groups"):
  - **IMPORTANT:** Case-crossover studies always involve repeated observations per subject; methods for "repeated measures" or, in the case of only two observations per subject, "pairs" should be described
- Focus groups and open-ended subject interviews (i.e., analysis of recordings or transcripts)

# What Methods Are Commonly Used for Analysis?

- Regression Models
  - Aid in controlling bias/confounding
  - Simultaneously adjust for all variables in model
  - Outcome of Interest = Dependent variable
  - Variables in Model = Independent variables
    - Predictors
    - Adjustors

# What Methods Are Commonly Used for Analysis?

- Types of Regression Models
  - Linear Regression
    - Dependent variable (outcome) is continuous
    - "Simple regression" = Only one independent variable used as predictor in model
    - Assumes linear relationship (i.e., if means of dependent and independent variables plotted against each other would fall on straight line)

# What Methods Are Commonly Used for Analysis?

- Types of Regression Models
  - Logistic Regression
    - Dependent variable (outcome) is categorical and dichotomous (only two levels)
    - Provides adjusted odds ratios, 95% CIs and p-values
    - Does NOT assume a linear relationship between variables

# Characteristics of a Good Analysis Plan

- **Parsimonious**
  - Only includes the variables needed to answer the research question
  - Avoids (or adjusts for) multiple comparisons
- **Clearly Described**
  - Comparison(s) to evaluate each study objective
  - Methods that will be used
  - How the researcher will know if the objective has been reached and their hypothesis proved/disproved

# Analytic Resources for BUMC Researchers

## Data Coordinating Center, BUSPH

http://www.bu.edu/dbin/sph/research_centers/dcc.php

## Department of Biostatistics, BUSPH

http://www.bu.edu/dbin/sph/departments/biostatistics

# General Clinical Research Center Investigators

**If you are currently a GCRC researcher**

**–OR–**

**If you are preparing an application to use the GCRC**

**The GCRC can provide assistance with development of the study design and analysis plan**

http://dccwww.bumc.bu.edu/gcrcweb/abouttheGCRC.htm

# Questions and Discussion