Collaborating with your Biostatistician

Dr. Monica Ahrens – CBHDS

Dr. Tonja Locklear – Carilion Clinic HART

Learning Objectives

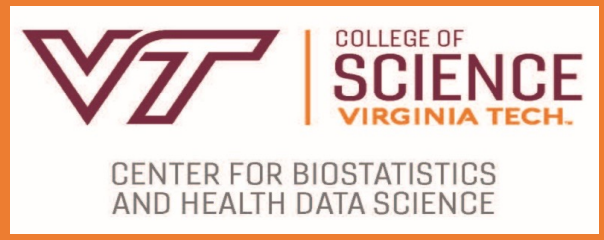
- Identify Carilion and Virginia Tech statistical analysis resource teams, including how to connect with them and the services each offers
- Recognize the importance of a statistical analysis plan.
- Describe how biostatisticians can assist within all stages research.

CARILION CLINIC

COLLEGE OF SCIENCE VIRGINIA TECH

CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE

Statistical Analysis Resources



HART

Health Analysis Research Team

- Director: Mattie Tenzer
- Employed by Carilion Clinic
- Email: hart@carilionclinic.org

CBHDS

Center for Biostatistics and Health Data Science

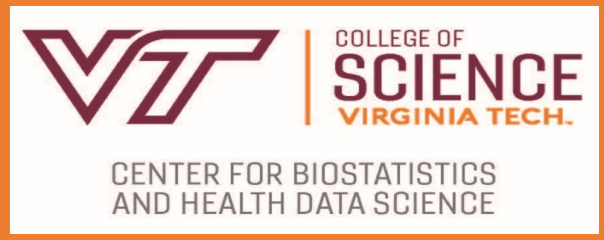
- Director: Alex Hanlon
- Employed by Virginia Tech
- Email: biostats@vt.edu
- Website: biostat.centers.vt.edu

CARILION CLINIC

VT | COLLEGE OF SCIENCE VIRGINIA TECH

CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE

Statistical Services Offered



HART

- 20 team members providing end to end support
- Mix of grant funded / institution supported
- 250+ combined years in healthcare/research
- 165+ combined years at Carilion
- 83% diversity (non-white (33%), female (72%), veteran (5%))
- Full support for Research Informatics, Biostatistics, Data Science, Epidemiology, Research Design, data extraction and research navigation services
- Carilion component of NIH CTSA iTHRIV iBERDI group
- Innovation Department support (app development)
- Virtual Office Hours: Tuesday 10 am – Noon (*Copy Link from Here*)

hart@carilionclinic.org

CARILION CLINIC

HART

End to End Research Support

hart@carilionclinic.org

Design

- Study design support
- Biostatistical consultation
  - Data discussion
    - Extract, Chart Review and/or Survey Data
  - Power Analysis
  - Feasibility – TriNetX
- Survey consultation
- Epic Research Access & Clinical Trial Support

Data

- Data extracts
  - Epic
  - TriNetX reidentification
  - Registries
- Data exploration, organization, and manipulation
- Data Management and Surveys (REDCap)
  - Survey Creation, Distribution & Support
- Secure Cloud-Based Data Storage (SPARC)

Analysis

- Data analysis support
  - Statistical Presentations
  - Guided and complete statistical analysis
  - Qualitative & Quantitative Analyses
  - Data Analysis in R-Studio, SAS Studio, Python, NVIVO
  - Data Science Methodologies
- Presentation of Findings
  - Poster, Manuscript, etc.
    - Images/Graphs
    - Methods/Results

CARILION CLINIC

Long-Term Collaboration

A biostatistician serving as a collaborator is an academic biostatistician who works with an investigator to study a research question or research agenda, preferably from study conception through dissemination.

Proposal submission:

- statistics analysis plan
- specific aims
- research strategy
- budget

Multi-year project period:

- analytical and data management support
- weekly team meetings
- continuation proposals
- interim reports

Dissemination of results:

- manuscript preparation
- abstracts, posters, oral presentations
- conferences

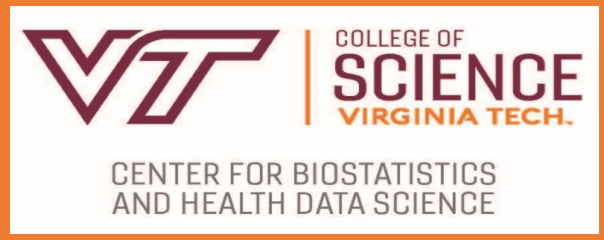
Short-term Consulting

A statistician serving as a consultant provides input on a straightforward statistical question that can be solved in a single meeting lasting for a relatively short period of time, say 30 minutes to an hour. As with many things, there is no clear or definitive marker that distinguishes when general advice transitions to impactful collaboration. We typically use time as the metric for making this distinction.

- Drop-in hours Monday 10am to 2pm and Wednesday 3-7pm. In person option during regular working hours and zoom option for all 8 hours each week.

- 4 Riverside Circle Room 1220
- https://biostat.centers.vt.edu/index.html#Zoomhours

Stages of a Research Study

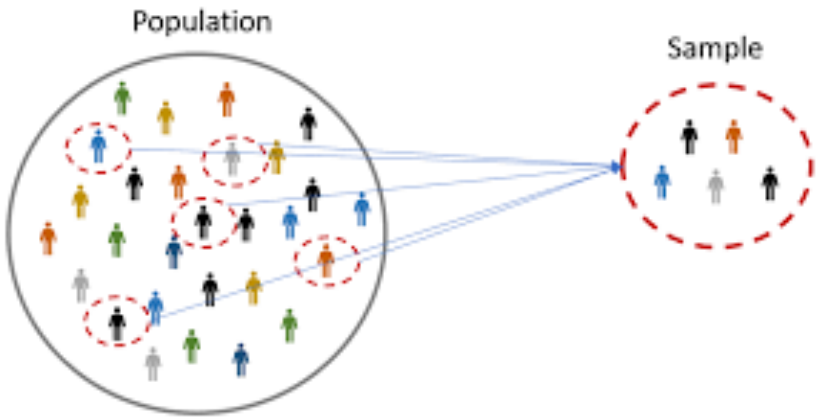


Stage I:

CARILION CLINIC

COLLEGE OF SCIENCE VIRGINIA TECH

CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE



Power and Sample Size Calculation

- For NIH Grants a Power Calculation is always needed when applying for funding
- The IRB at Carilion requires that a statistician reviews your power calculation

CARILION CLINIC

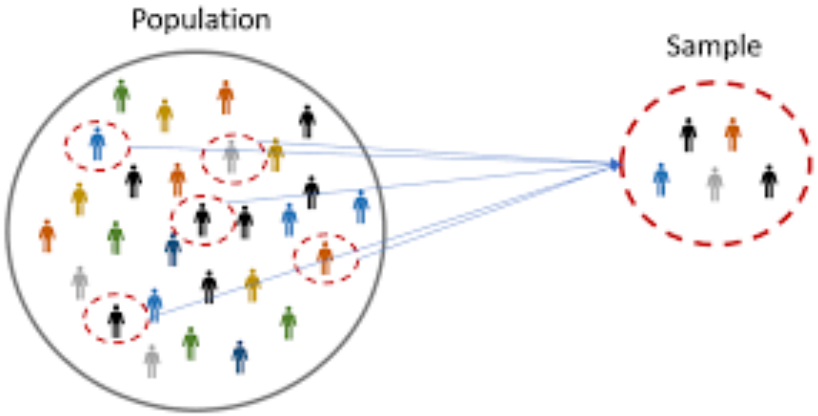
COLLEGE OF SCIENCE VIRGINIA TECH

CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE

Sample Size Estimations

The key components needed to estimate sample size:

- Research question or questions
- Desired significance level and power level
- Effect size of clinical relevance
  - What's the smallest change you *care* about from a clinical point of view?
- Variability
  - What range of measurements can we expect by chance?



Sample size is a balance of statistical and practical considerations.

Statistical Analysis Plan (SAP)

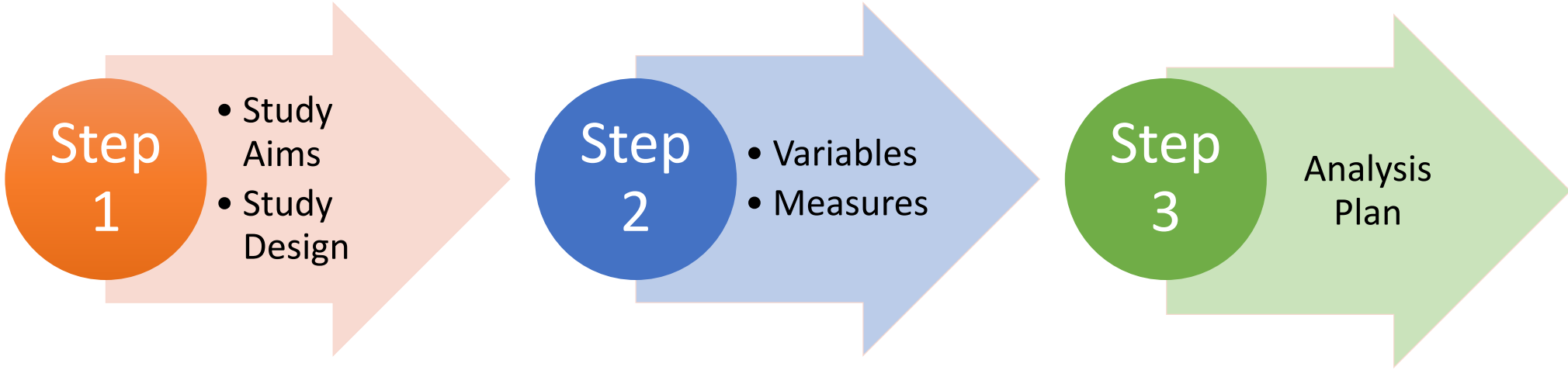
- The SAP describes the *planned analysis* for a study
  - Background information of the research study
  - Design and framework of the study
  - Inclusion/Exclusion criteria for study participants
- Serves as a plan and a *checklist*
  - Facilitates thoughtful evaluation
  - Ensures fine details are not missed
- *Everything but the science.*

CARILION CLINIC

COLLEGE OF SCIENCE VIRGINIA TECH

CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE

What Drives the SAP?



Some pitfalls to avoid

- Feasibility
  - Is the data available for the study?
- Data collection
  - Does the data answer the research question?
- Sample size
  - Are there enough subjects to power the analysis?

CARILION CLINIC

COLLEGE OF SCIENCE VIRGINIA TECH

CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE

Step 1

- Study Aims
- Study Design

Study Aims

- Study *aims* are the main objectives of the research, they point you at your *target*.
- Everything about the SAP and study flows from the aims.

Study Design

- Study *design* refers to how the data are collected and structured.
- Observational Studies
  - Retrospective
  - Survey
- Controlled Trials
  - Treatment
  - Assignment
  - Randomization

CARILION CLINIC

COLLEGE OF SCIENCE VIRGINIA TECH

CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE

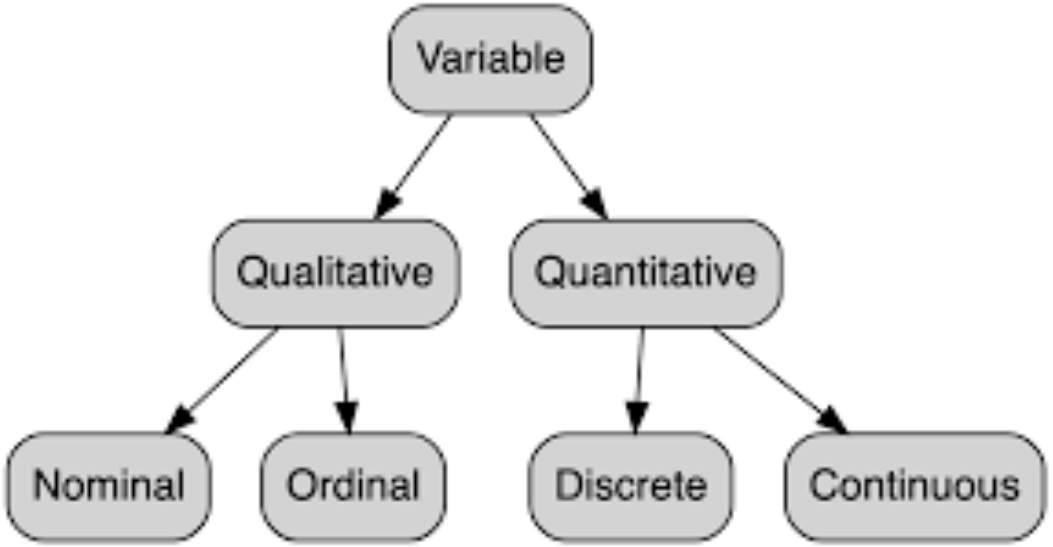
Variables and Measures

Step 2

- Variables
- Measures

Examples:

- Qualitative/Categorical
  - Nominal: sex, eye color
  - Ordinal: height as short, medium, tall
- Quantitative/Numeric
  - Discrete: number of siblings
  - Continuous: BMI



The variable type determines the appropriate statistical model and possible inferences.

Analysis Plan

The analysis plan consists of three main components

1. Preliminary analyses
   - Table 1, basic descriptive statistics, plots of the data
   - Discuss how to handle missing data (imputation, complete case, etc.)
2. Analyses by aim
   - Specify the statistical technique to be used in each aim
   - Where appropriate, discuss parameters and settings for the model

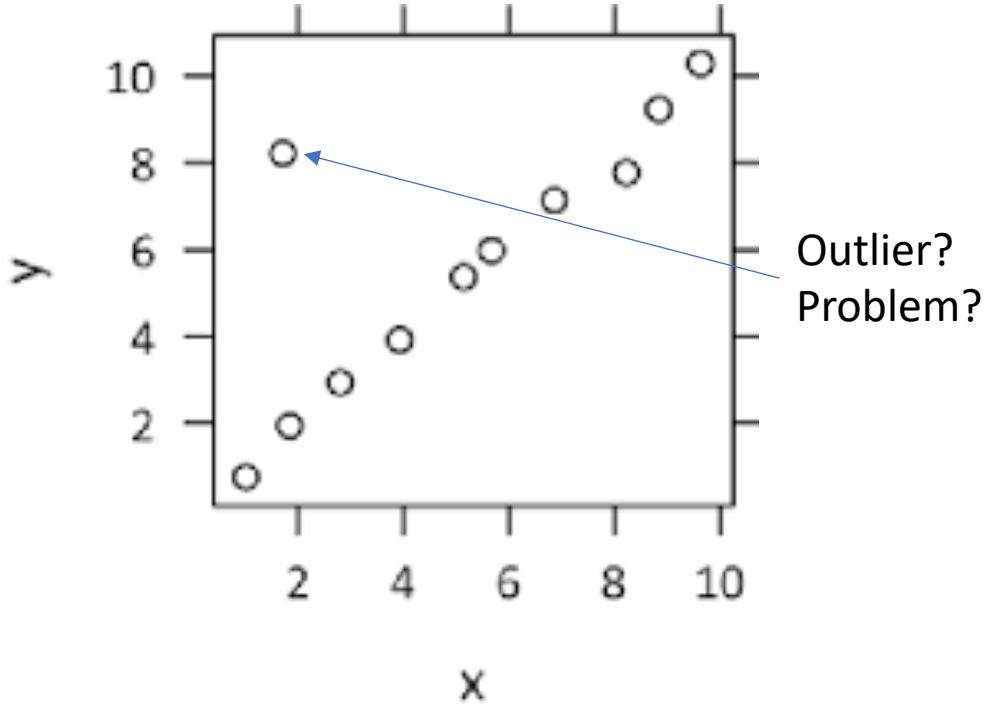
Stage II: Data Visualization and Preliminary Analyses

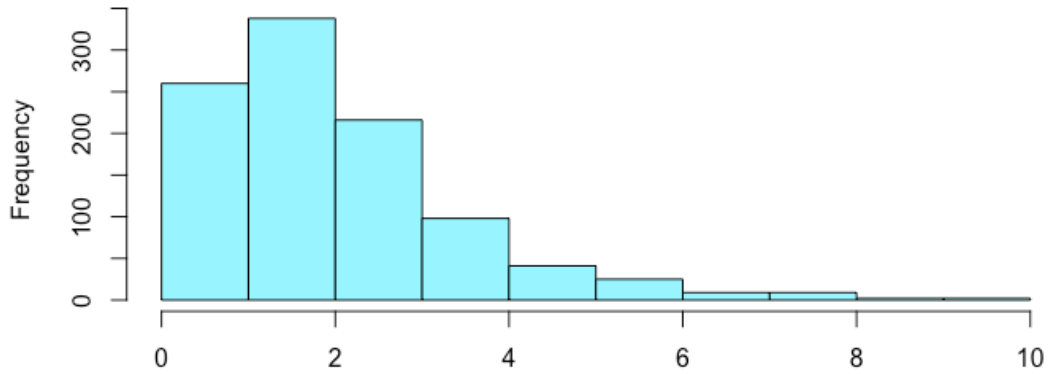
CARILION CLINIC

COLLEGE OF SCIENCE VIRGINIA TECH

CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE

Why Should We Visualize Data?





- Visualizing your data can reveal obvious data errors and give you a feel for data distributions.
  - In the real world, there *will* be errors in your data.

Most statistical methods make assumptions about data and visualizations are a good way to check those assumptions

- Are the data symmetric/normally distributed?

CARILION CLINIC

COLLEGE OF SCIENCE VIRGINIA TECH

CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE

Characterizing the Sample

- "Table 1" in most clinical papers describes the sample used in your study

Table 1: Patient Demographics

	Overall	No Disease	Disease	p-value
Age				<0.0001
	46.4 (7.71)	47.1 (7.96)	43.2 (5.49)	
Sex				0.0008
Male	171 (48.9%)	152 (53.3%)	19 (29.2%)	
Female	179 (51.1%)	133 (46.7%)	46 (70.8%)	
Location				<0.0001
Rural	70 (20%)	61 (21.4%)	9 (13.8%)	
Suburban	186 (53.1%)	167 (58.6%)	19 (29.2%)	
Urban	94 (26.9%)	57 (20%)	37 (56.9%)	
Time				0.0074
Pre-COVID	150 (42.9%)	112 (39.3%)	38 (58.5%)	
Post-COVID	200 (57.1%)	173 (60.7%)	27 (41.5%)	
Survival				0.5815
Yes	257 (73.4%)	207 (72.6%)	50 (76.9%)	
No	93 (26.6%)	78 (27.4%)	15 (23.1%)	

* Hypothetical data

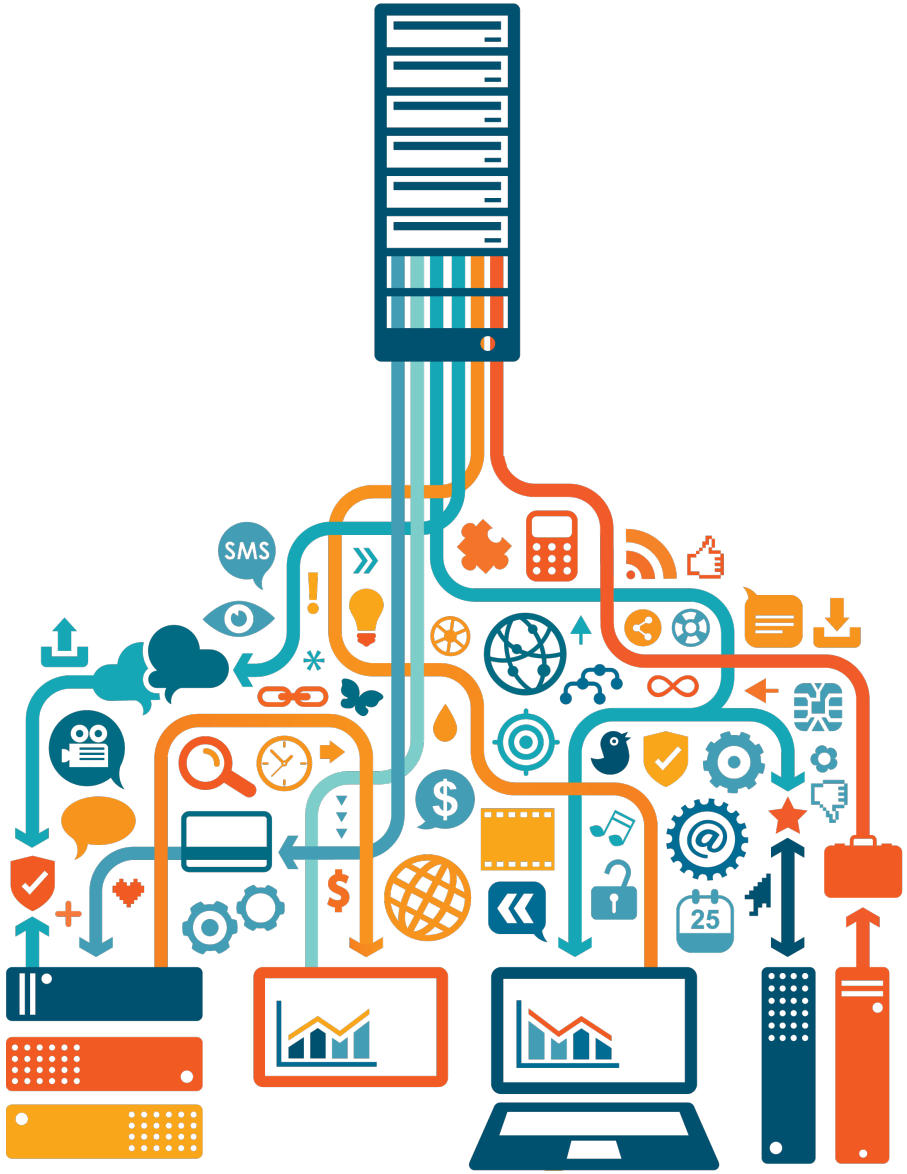
Stage III: Data Cleaning and Analysis

CARILION CLINIC

COLLEGE OF SCIENCE VIRGINIA TECH

CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE

- Data Extraction Tools
  - Epic
  - TriNetX
- Data Collection Tools
  - Surveys
  - Chart Reviews
- Data Management Tools
  - REDCap
  - Shared Drive
  - Secured Cloud-Based Storage



CARILION CLINIC

VT | COLLEGE OF SCIENCE VIRGINIA TECH

CENTER FOR BIOSTATISTICS AND HEALTH DATA SCIENCE

- Data Analysis will follow guidelines from:
  - IRB Protocol
  - SAP
  - Client Discussions
- Data Analysis Methods:
  - Qualitative and Quantitative
  - Descriptive Statistics
  - Data Visualizations
  - Bivariate analyses (t-tests, Wilcoxon rank sum tests, ANOVA, chi-square tests, Fisher's exact test, etc.)
  - Multivariable analyses (linear regression, linear mixed models)

Stage IV: Revisions and Finalization

- Manuscript and Poster Preparation
  - Written assistance with sections of the research paper or presentation
  - Assistance with tables, figures and graphs
  - Assistance with statistical formatting
- Concerns and Considerations
  - Timing issues, please allow reasonable turn around period
  - Please add us as coauthors, if appropriate.

http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html

Thank you!

hart@carilionclinic.org

biostats@vt.edu