

Collaborating With a Statistician: Bringing a Clinical Perspective to Statistics

Matt Hall, PhD,^a Troy Richardson, PhD,^a Jay G. Berry, MD, MPH,^b Lilliam Ambroggio, PhD,^c Samir S. Shah, MD, MSCE,^c Kenneth B. Roberts, MD^d

Peruse the author lists of articles in any medical journal, and you will quickly appreciate that research is performed by teams. Similar to any team that you might encounter in sports, there is a distribution of labor among the research team members that makes them operate effectively and efficiently. Although statisticians are critical members of the research team, they are frequently engaged to assist with analysis well after the study has been designed and executed. The statistical methods used to answer a research question directly affects the study design, the data collection instruments, and the operationalization of the variables. Therefore, the consequence of delaying a statistician's engagement is that the collected data may have significant limitations that statistical techniques alone cannot fix to answer the intended questions. However, involving a statistician from the onset on a research team does not relegate all statistical responsibilities to the statistician. It is still incumbent on the clinical study members to lend their insights and content expertise to guide the statistician in developing and executing the analysis. Consequently, dialogue between the clinicians and statisticians is paramount to making appropriate decisions and reaching sound conclusions from a study in the most efficient manner. In this article, which is intended for clinicians who conduct research, we explore this critical interface and discuss how clinicians can better collaborate with a statistician throughout the research process.

DEVELOPING THE ANALYTIC PLAN: THE PLANNING PHASE

Familiarity with the study and the field helps a statistician engage in your research, learn the language and statistical techniques commonly used in your area of expertise, and operate most efficiently. Although some statisticians may have been through training programs that expose them to some basics in medicine (ie, biostatistics), many will not be familiar with even the basic clinical concepts that make your topic relevant. Spend time explaining the science, providing an overview of what has been done in the past, explaining how your research fills critical gaps in our knowledge, and providing gold-standard papers for the statistician to review. The statistician will be able to better guide you through the analysis so that you have a better understanding of how the results were derived.

At the beginning of a study, the statistician will be a key player to help the team develop an analytic plan. This will likely include deciding on an appropriate study design, power and sample size considerations, and analytic methods to be used. All of these should be decided on through dialogue to weigh the various options available. There is rarely one best way to do a study, so advantages and disadvantages should be thought through before deciding on any of these issues.

www.hospitalpediatrics.org

DOI:10.1542/hpeds.2016-0054

Copyright © 2016 by the American Academy of Pediatrics

Address correspondence to Matt Hall, PhD, Children's Hospital Association, 6803 W. 64th St, Overland Park, KS 66202. E-mail: matt.hall@childrenshospitals.org

HOSPITAL PEDIATRICS (ISSN Numbers: Print, 2154-1663; Online, 2154-1671).

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: No external funding.

POTENTIAL CONFLICT OF INTEREST: The authors have indicated they have no potential conflicts of interest to disclose.

^aChildren's Hospital Association, Overland Park, Kansas; ^bDivision of General Pediatrics, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts; ^cDivision of Hospital Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; and ^dDepartment of Pediatrics, University of North Carolina School of Medicine Greensboro, North Carolina

Determining how many subjects you will need in your study (ie, sample size) avoids having insufficient power to detect important clinical differences. The number of subjects needed for a study depends on several factors, including the level of confidence that you would like to have in your findings and the magnitude of difference you consider clinically meaningful. For example, we may be doing a study that compares the efficacy of 2 drugs on treating gastroesophageal reflux disease. If you consider that an additional benefit of 20% would be a clinically meaningful difference, we will need significantly fewer children to participate compared with a scenario in which we believe the difference will be 5%. As this detectable difference decreases, the sample size will increase.

When developing the analytic plan, statisticians have many different options in their toolbox, and selecting from among them can be challenging if they do not have enough information. Two critical inputs that the clinical team can provide the statistician at this stage include (1) how to use variables in a manner that makes clinical sense and (2) the nature of the dependent variable and its relationship to the independent variables.

First, the statistician will need your guidance on how variables should be used analytically so that they have clinical relevance and meaning. For example, if your study involves length of stay and discharge timing, to the statistician, it may seem logical that the 8-hour blocks of time between noon and 8 PM and midnight and 8 AM are similar because they contain the same number of hours. However, these are very different blocks of time in a hospital: things being very active during the day, with people coming and going, but virtually no one being discharged at 3 AM. If a patient is in the hospital at 3 AM, chances are that the patient will still be there at 6 AM, even if the patient is ready for discharge. As another example, to a statistician, blood pressure may be used analytically as a continuous variable, but it might make more sense clinically to categorize it (eg, hypertensive, normotensive, or hypotensive). Similarly,

body temperature may appear to a statistician to be continuous, but there are biologic pressures on both ends of the scale, with the difference between 99 and 100°F being very different clinically from 106 and 107°F. Knowing these types of clinical nuances in the variables will allow the statistician to use them appropriately in subsequent analyses.

Second, understanding the nature of the dependent variable (ie, the outcome) and its relationship to the independent variables is essential to selecting an appropriate type of model. Statisticians have many types of models to select from, but there are assumptions of these models that may not be met or clinically justifiable. Returning to our example of changing body temperature, as the dependent variable, a linear model might not be the best option because a change of 1 degree has different clinical meaning at different areas of the temperature scale. In these cases, nonlinear models may be more appropriate.

EXECUTING THE PLAN: THE ANALYTIC PHASE

Once the analytic plan has been developed and the data collected, it is time to execute the analytic plan. Although there may be justifiable reasons to deviate from the analytic plan (eg, the underlying assumptions of the selected model were not substantiated with the data), for the most part, a solid plan will provide the foundation for the analytic sequence.

Typically, the analysis will begin with an exploration of the data. This will help the team understand the empirical nature of the variables, and how this might deviate from the original intent of using the data analytically. As an example, the analytic plan might propose using age as a categorical variable with 5 strata. However, once the data are collected, there may not be sufficient data across the strata, and they may need to be redefined with fewer groups. Exploration of the data will also identify issues in the collected data, including missing data and potential outliers. Generic statistical outlier detection methods are typically based on the observed distribution of the data, and do not include clinical plausibility. For example,

a review of body temperatures may define outliers as temperature values $<97^{\circ}\text{F}$ and $>105^{\circ}\text{F}$. Although values outside of this range may be real, they are expected to be rare and may warrant review. Consequently, a statistical approach may provide a list of potential outliers, but these should be viewed through a clinical lens before treating them as actual outliers.

After understanding the data descriptively, addressing missing data and outliers, the statistician will generally explore relationships between all of the variables (independent and dependent) in a bivariate fashion. This will help the team determine which adjustments to make for interrelatedness of independent variables and relationships between independent variables and the dependent variable. Graphical representations of the relationships (eg, scatter plots between 2 continuous variables or box plots between continuous and categorical variables) are helpful to determine linear versus nonlinear relationships in the data. At times, highly related independent variables will need to be combined (eg, through a factor analysis) or reduced to ensure appropriate modeling. Clinical rationale should drive these decisions so that the clinical meaningfulness of the resulting analyses does not get lost in the statistical manipulations.

Next, the underlying assumptions of the models need to be verified with the actual data, and appropriate adjustments made. Statistics and statistical significance (eg, P values) are driven by many factors. They are useful tools for guiding scientifically sound decisions, but they should not be used as a definitive tool void of clinical judgment. As an example, suppose we are comparing 2 groups of children on their weight loss with 2 randomly assigned drugs. After we collect the data and do various adjustments, we find that children on drug A lost, on average, 5.5 kg, and children on drug B lost 6.0 kg. With a large enough sample size, we might conclude that this difference is statistically significant (ie, $P < .05$), but this may be a clinically meaningless difference.

The resulting models need to be challenged by testing underlying assumptions (statistical and clinical). This typically takes the form of sensitivity analyses to ensure that the models are robust to permutations of different decisions that have been made along the way. Although the statistician can suggest various modeling alternatives to assess this from an analytic perspective, the clinical team can provide plausible clinical alternatives that should be explored. Examples of alternatives include altering inclusion/exclusion criteria, treating outliers differently, redefining strata for categorical variables, or stratifying a continuous outcome.

Last, the statistician is typically charged with drafting the methods portion of presentations, reports, and publications resulting from the study. Sometimes, these sections of text can be filled with highly technical language that is unapproachable by a nonstatistical audience. Although there are important details that should be included to explain the methodology, the clinical team should provide input on making the methods understandable by a clinical audience.

CONCLUSIONS

Developing an ongoing relationship with a statistician can be beneficial for your research career. The more a statistician is

engaged in your research, the greater their investment and accountability for its ongoing success will be. Statisticians are key players on any successful research team, and as such, should be engaged throughout the research process and be an integral part of the team. However, developing and executing the analytic plan does not lie solely on the shoulders of the statistician. Rather, there should be ongoing dialogue between the statisticians and the clinicians to ensure that the results are clinically meaningful and scientifically sound. As such, it is as imperative that you bring a clinical perspective to statistics as a statistical perspective to your research.

Collaborating With a Statistician: Bringing a Clinical Perspective to Statistics

Matt Hall, Troy Richardson, Jay G. Berry, Lilliam Ambroggio, Samir S. Shah and
Kenneth B. Roberts

Hospital Pediatrics 2016;6;750

DOI: 10.1542/hpeds.2016-0054 originally published online November 7, 2016;

Updated Information & Services	including high resolution figures, can be found at: http://hosppeds.aappublications.org/content/6/12/750
Subspecialty Collections	This article, along with others on similar topics, appears in the following collection(s): Medical Education http://classic.hosppeds.aappublications.org/cgi/collection/medical_education_sub Research Methods & Statistics http://classic.hosppeds.aappublications.org/cgi/collection/research_methods_-_statistics_sub
Permissions & Licensing	Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at: https://shop.aap.org/licensing-permissions/
Reprints	Information about ordering reprints can be found online: http://classic.hosppeds.aappublications.org/content/reprints



Collaborating With a Statistician: Bringing a Clinical Perspective to Statistics

Matt Hall, Troy Richardson, Jay G. Berry, Lilliam Ambroggio, Samir S. Shah and
Kenneth B. Roberts

Hospital Pediatrics 2016;6;750

DOI: 10.1542/hpeds.2016-0054 originally published online November 7, 2016;

The online version of this article, along with updated information and services, is
located on the World Wide Web at:

<http://hosppeds.aappublications.org/content/6/12/750>

Hospital Pediatrics is the official journal of the American Academy of Pediatrics. A monthly publication, it has been published continuously since 2012. Hospital Pediatrics is owned, published, and trademarked by the American Academy of Pediatrics, 345 Park Avenue, Itasca, Illinois, 60143. Copyright © 2016 by the American Academy of Pediatrics. All rights reserved. Print ISSN: 2154-1663.

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™

