Biomedical Researchers Should be Taught Statistics Differently than Biostatisticians in Training: Illustration of a Module within a Clinical Research Seminar Course

Gregory P. Samsa[1,*], Steven C. Grambow[1], Megan L. Neely[1], Gina-Maria Pomann[1], Clemontina Davenport[1], Marissa Ashner[1] & Jesse D. Troy[1]

[1]Department of Biostatistics and Bioinformatics, School of Medicine, Duke University, Durham, NC, USA

*Correspondence: Department of Biostatistics and Bioinformatics, School of Medicine, Duke University, Durham, NC, USA. Tel: 1-919-668-2932. E-mail: jesse.troy@duke.edu

Received: February 7, 2024 Accepted: March 10, 2024 Online Published: April 30, 2024

doi:10.5430/jct.v13n2p33 URL: https://doi.org/10.5430/jct.v13n2p33

Abstract

The predominant model for biomedical research is team science. Two critical members of the team are the clinical investigator and the biostatistician. Typically, the biostatistician performs statistical analyses and the clinical investigator interprets the results. Clinical investigators have different background and interests than biostatisticians, and should be taught statistics differently. Concepts should be phrased in plain language, illustrations should replace mathematical derivations, and underlying statistical concepts should be explicitly named. Consistent with basic principles of constructivism, clinical investigators and biostatisticians will (and should) have different but overlapping mental maps of statistics. Our approach is illustrated through the description of a module within a research seminar course for clinical investigators.

Keywords: clinical researchers, constructivism, mental maps, statistical education

1. Introduction

The predominant model for biomedical research is team science (Pomann et al., 2021). Two critical members of the team are the domain expert and the biostatistician. The domain expert might be a laboratory scientist, a psychologist, a physician, a decision modeler, or a hospital administrator, among others. For concreteness and simplicity of exposition, we assume that the research will be performed using humans, and that the domain expert is a clinical investigator (CI) such as a physician-scientist.

We have taught statistics to CIs and biostatisticians in multiple settings (Neely et al., 2022; Pomann et al., 2022; Samsa et al., 2012; Samsa, 2018; Samsa et al., 2018, Samsa, 2020, Samsa, 2021; Troy et al., 2021; Troy et al, 2022a, Troy et al., 2022b; Troy et al., 2023; Troy et al., 2024). Among them are two masters' programs. One program, the Master of Biostatistics Program (MB), focuses on biostatisticians in training. The MB program, among others, prepares its students to effectively collaborate with CIs. Another program, the Clinical Research Training Program (CRTP), focuses on domain experts in biomedical science such as physician-scientists who are early in their academic careers (e.g., fellows). The CRTP, among others, prepares its students to effectively collaborate with biostatisticians. One premise of both programs is that team science proceeds more effectively when, instead of only mastering a single "silo" of knowledge, practitioners can span multiple disciplines. -- for example, when biostatisticians have some knowledge of the underlying science and CIs have some knowledge of statistics.

Established in 1986, the CRTP was one of the first programs of its kind, and its longevity has allowed us to accumulate experience in teaching both basic and advanced research competencies to physician-scientists. Considering advanced statistical competencies, we discovered that CRTP students are a group with unique characteristics that distinguish them from more typical students of statistics: for example, their domain knowledge and general scientific sophistication are exceptional, yet (in large part because of time constraints) their effective mastery of mathematical argumentation and computer programming is modest. Although there is a literature on methods for teaching basic statistics to this audience (e.g., Bland, 2004; Miles et al., 2010), we discovered that it does not directly apply to teaching advanced statistical competencies, which require not only mastering technical and

conceptual information, but integrating the two.

The pedagogic goals for the MB program are (among others) to: (1) assist students in learning to think about statistics like a biostatistician (i.e., to develop a sound mental model of the discipline); (2) provide students with the ability to implement statistical tools; and (3) provide students with structured exposure to effective collaboration with CIs. Curriculum development was aided by a natural starting point: namely, the traditional mathematically-based materials for teaching the fundamentals of statistical inference, and relied upon students' facility with mathematics as a core language of instruction.

This report considers curriculum development within the CRTP. For example, we ask whether the statistical components of the CRTP should be delivered similarly to the MB curriculum or differently. Indeed, should we begin with the same mathematically-based road map used as a starting point for designing a curriculum for biostatisticians or should we conceptualize the curriculum development task differently? Our approach is not a formal evaluation of the CRTP curriculum, but instead is reflection and identification of general principles which might also be applicable to teaching statistics to CIs in other settings. The application of these principles is illustrated through the description of a module within a clinical research seminar course.

2. Conceptual Model

2.1 A straw Man

Before discussing what ought to be done, we illustrate what ought not to be done. Consider, for example, an "introduction to statistics" course provided to medical students or a similar audience. This course often begins with a module on t-tests to determine whether two group means are significantly different from one another. This module describes a protocol by which a t-test is performed, one of the steps being "check that the assumption of normality is satisfied: if so, calculate a p-value using a specific formula (or software)". Perhaps the logic of hypothesis testing is explained within the context of a t-test, and perhaps it was explained previously.

We would argue that this approach, as typically implemented, has a number of deficiencies:

• It gives too much emphasis to hypothesis testing, thus perpetuating a misplaced emphasis within the biomedical literature.

• It fails to introduce the concept of a sampling distribution, which is central to statistical inference in general and is often misunderstood.

• It wrongly presents a statistical analysis as being fundamentally similar to a laboratory protocol -- that is, as a set of mechanical operations. For example, it treats "check that the assumption of normality is satisfied" as a binary decision, although what should actually be done is more nuanced -- namely: "visualize the data and decide, based on both the data and the underlying science, whether comparing group means is appropriate".

• Except for an unadjusted analysis associated with a randomized trial, t-tests ought to be relatively uncommon. If this course is to be the entirety of statistical training provided to the student, then space within the curriculum is at a premium, and learning the technical mechanics of performing a t-test is an inefficient use of that space.

According to our observation, CIs are at especial risk of treating statistics as an oversimplified protocol.

2.2 Competencies

In designing statistical curricula for CIs, a principled way to proceed begins with the competencies which we wish them to develop (and also the specific skills which those competencies imply). A natural place to start is the literature on statistical education which, for example, lists various competencies for biomedical researchers pertaining to statistics. (e.g., Enders, 2011; Enders et al., 2017; Hornung et al., 2018; Oster et al., 2015; Oster & Enders, 2018; Sonstein et al., 2014; Sonstein et al., 218, Sonstein et al., 2020) Illustrative competencies include identifying the strengths and limitations of study designs and recognizing when to involve a quantitative expert. A list of statistical competencies for CIs contains some, but not all, of the competencies which biostatisticians must master.

A critical question in curriculum design might be framed as follows: If a curriculum for biostatisticians is intended to develop competencies A-J and a curriculum for CIs is intended to develop competencies A-E, is it sufficient to take the curriculum for biostatisticians and simply leave out F-J? Or, instead, would another approach be preferable?

2.3 Overlapping Mental Maps

All disciplines have multiple organizing structures and, indeed, one of the hallmarks of an expert practitioner is the ability to meld multiple ways of framing a problem, and also to select an organizing structure which is well suited to

the problem at hand. A principle of constructivism (Begg, 2015) is that these organizing structures eventually become internalized as "mental maps (mental models)" of the discipline, and also that one of the responsibilities of educators is to provide experiences which help students develop sound mental maps (Biggs, 1996).

A particular challenge for teaching statistics to students outside the discipline, especially CIs whose mathematical training isn't necessarily extensive or current, is that the primary organizing structure used to teach biostatisticians is that of mathematics. Assuming that the student in question has sufficient mathematical training this way of organizing the information is quite natural. Moreover, this is the way that statistical faculty learned their discipline, and so the default way to teach statistics to non-statisticians is to use mathematics as the primary organizing structure, with some of the more complex and technical details omitted. In effect, the default is to select competencies A-E and teach them in the usual way, perhaps with a bit of watering down.

Experience – not just qualitative but consistently supported by student evaluations – demonstrates that, for many CRTP students, mathematics simply does not work well as a central organizing structure, and thus that different approaches are needed. A primary reason is that within this context the processing task of converting mathematical symbols to usable information is too demanding, and thus neither intuition nor insight result. In other words, applying the traditional approach to delivering statistical content to CIs relies on a critical and faulty assumption: namely, that CIs do (and should) learn statistics in the same fashion as the mathematically-oriented students who are the target audience for the traditional approach.

To productively collaborate with biostatisticians in team science projects CIs require two linked mental maps. The first is their own mental map around data analysis. This mental map will be simpler and different from than that of a biostatistician, but it should nevertheless be sound. The second is a general understanding of how biostatisticians think. CIs need not be proficient enough in applying this latter mental map to accomplish the same analytic tasks as biostatisticians, but they should have enough "tourist-level functionality" to communicate with them. In other words, CIs should be familiar with the structure of the mental map used by biostatisticians.

2.4 A Reconceptualization of Curriculum Goals

Table 1. Based on the above, We Began with Three Curriculum Goals for the MB Program, and Translated Them into Goals for CRTP

Goal in MB curriculum	Goal in CRTP curriculum
assist students in learning to think soundly about statistics as a biostatistician	assist students in learning to think soundly about statistics as a CI
provide students with the ability to implement statistically-related tools	provide students with the ability to effectively use the results of statistical analyses
provide students with structured exposure to effective collaboration with CIs	provide students with structured exposure to effective collaboration with biostatisticians

The first goal, namely assisting students in thinking soundly about statistics, is effectively the same for CRTP and MB. However, the goal is implemented differently, because the mental models of CIs and biostatisticians differ. The second goal is intended to lead to the production and effective use of statistical analyses, with biostatisticians primarily responsible for their production and CIs primarily responsible for interpretation. The third goal, namely providing structured exposure to effective interactions between clinicians and biostatisticians, is effectively identical for the two programs. Implicit within the third goal for CIs is exposure to mental maps of how biostatisticians conceptualize the discipline of statistics. In practice, this involves exposure to big-picture concepts such as the visualization, analysis, implementation (VAI) cycle, the distinction between signal and noise, and the fundamental elements of a statistical analysis plan, among others.

3. Illustration

3.1 A clinical Research Seminar Practicing High-level Statistical Skills

Appendix 1 presents excerpts from the syllabus for CRP247, a clinical research seminar held in the second year of the CRTP program. Briefly, its goals include (1) illustrating the mental maps of biostatisticians around high-level statistical constructs about which CIs often struggle; and (2) helping CIs develop their own mental maps around these constructs, while navigating various constraints. These constraints include (1) mathematics is not a natural language of instruction; (2) students are generally unwilling or unable to engage in significant computer

programming; and (3) background knowledge from previous introductory statistics courses might not yet be fully mastered.

Our general pedagogic approach centers on (1) explicitly identifying and naming the construct being studied; (2) providing a definition using non-technical language with mathematical notation deemphasized; (3) replacing derivation with simple illustration; and (4) describing the mental map of the construct used by biostatisticians.

3.2 Material from the Session on Subgroup Analysis and Interaction

Appendix 2 presents the slides which accompany the pre-class video on subgroup analyses and interaction. We highlight how these slides illustrate the above pedagogic approach.

Slide 1 describes the problem using non-technical language, namely: *subgroup analyses tend to be overinterpreted*. It also explicitly names a related construct, namely regression toward the mean.

Slide 2 lists potential causes of overinterpretation, in order of seriousness. The supplemental readings for this module provide examples of each.

Slides 3-6 provide supplemental information, using (1) very simple examples; and (2) intuitive explanations. For example, in slide 3: *reporting only the most significant results (or only the results which favor your hypothesis) induces a bias in favor of your hypothesis* is intended to be an intuitive explanation. In slide 5, a counterexample is provided to a typical fallacy in the medical literature, namely that a treatment being statistically significant in one group but not another must imply that the treatment impact is different across the groups. Slide 6 is intended to be an entrée into a discussion about a common questionable research practice: namely, attempting to turn a negative study into a positive one by focusing on the subgroup which happened to most strongly favor the intervention.

Slide 7 provides concrete plain-language recommendations for performing subgroup analyses. In the spirit of "knowing what you know and what you don't know", the videos highlight that the last two recommendations require assistance from the biostatistician to implement. For example, the final recommendation is essentially "*ask your biostatistician to help account for the falsely positive conclusions induced by performing multiple statistical tests"*.

Slide 8 provides a plain-language definition of an interaction: *the impact of X on Y depends on the level of Z* and a concrete example: *the impact of treatment (X) on survival (Y) depends on the presence or absence of a genetic signature (Z).* Slide 9 relates this definition to something outside the CI's mental map -- namely, how the biostatistician will actually implement a test for interaction within their statistical software. This is intended as aid to communication across disciplines, to help the CI understand some of the technical details of the biostatistician's task, but without teaching the CI how to actually perform the task.

Slide 10 introduces confounding, a construct which is often confused with interaction. Plain-language descriptions are used, consistent with the how the CI will encounter these constructs in practice.

Slide 11 introduces a general statistical principle -- namely, that *focusing on the most dramatic subgroup effect, even with caveats, is dangerous* -- in a fashion which can be integrated into the mental map of the CI. Ideally, the CI could extend this principle to other circumstances -- for example, to recognize that a variable selection application where 100 potential predictor variables are screened and only 5 are found to be statistically significant is essentially the same thing. This slide also names a related concept, *regression toward the mean*.

Slide 12 includes a plain-language description of a general statistical principle -- namely, that *the more subgroup analyses the greater the number of false positives* -- a simple numerical example, and an actionable conclusion (i.e., that *subgroup analyses based on small sample sizes are particularly unstable)* which, in turn, implies that they should be approached with especial skepticism (Andrade, 2021).

Slide 13 introduces the construct of regression toward the mean by example rather than by a less accessible mathematical formula. Slide 14 provides illustrative data. The video notes that, when reading across a row of the accompanying table the results are consistent with intuition, but when reading down a column the regression toward the mean is observed. This provides the intuition for the analysis in slide 15, which concludes with the plain-English description of the underlying principle: *extremely good performance is usually part skill and part luck.* Slide 16 lists some applications of this principle that CIs are likely to encounter.

4. Evaluation

There are multiple ways to evaluate an element of a curriculum, ranging from unsystematic observation to formal experiments. Examples of the latter include randomized assignment of students to different approaches, pre-post

designs where student outcomes are compared before and after implementation of a curriculum change, etc. While experimentation results in the most direct inference, the choice of method should be based on context (Sullivan, 2011). For example, randomized experimentation should be considered when the curriculum change is significant, there is equipoise about whether or not it will be successful, students are willing to be randomized, there is an objective measure of student performance that can serve as a natural outcome variable, etc.

In the present circumstance, we chose to base our evaluation on student comments and instructor observation. Our rationale for evaluating our approach in this fashion was based on three considerations. First, as part of their professional responsibilities, our students are experienced in providing critical comment on similar constructs (e.g., quality of clinician-patient communication during a clinical encounter), and could reasonably be considered to be reliable reporters. Second, the construct being evaluated was salient to students, as the initial class session established that a key student goal in fact was to learn high-level statistical skills to help support their professional development. Thus, we anticipated that student comments would be thoughtful. Student comments were consistently positive: for example, "now I understand how this fits together with other topics", "this will improve the quality of collaboration with my biostatistician", etc.

Third, and in contrast to a specific technical element of statistical knowledge such as a formula that is amenable to a fact-based assessment, the utility of the information in this course will only become apparent over time, and will manifest itself in elements of professional advancement (e.g., success in grants and publications) that have multiple causes. Thus, cause and effect are difficult to definitively separate. Instructor observation included two initial manifestations of high-level statistical skills that were observable during class time: namely, (1) student summaries of the Zoom video that presented the content for the session; and (2) quality of the discussion when journal articles were reviewed. The instructor rated both as highly encouraging, especially so when students recognized the limits of their knowledge.

5. Discussion

Biomedical science is team-based and multidisciplinary. While the lion's share of work related to data analysis is left to the biostatistician, all other team members are expected to contribute to the interpretation and discussion of the findings. This requires all team members to be conversant at some level with what the biostatistician has done. In other words, team science requires a common paradigm of thought and understanding to enable fruitful communication concerning the results of data analyses. On the other hand, different disciplines necessarily bring different approaches to thinking about the same problem. Loosely speaking, based on constructivist theory, we can call these approaches to thinking separate and overlapping mental maps.

The notion of separate and overlapping mental maps is based on our experience taking the role of the biostatistician within biomedical team science and teaching statistics to future biostatisticians and CIs, and is in part a response to an evaluation we consistently received from those CIs, namely: "I'm relatively confident that I can do what we covered in class but am uncertain when to do so. For example, without additional guidance I won't know whether this is a regression problem or an analysis of variance problem, or whether to engage in variable selection or testing a prespecified hypothesis." In the language of constructivism, this is a plea from the CI to make the mental map used by the biostatistician more explicit and actionable.

These anecdotal observations led us to ask, when teaching CIs about statistics within the context of team science, should we be helping them to develop the *same* mental map as biostatisticians, only with less detail, or instead should these two audiences have *different* mental maps? Upon reflection, our experience with effective team science is that the collaborators from different disciplines don't utilize the "same mental map only less" but, instead, rely on mental maps which are different yet overlapping. To make a metaphor, one might understand these as two maps of the same geographic area, one with fine-grained navigational details desired by the local inhabitants and the other with highlights of landmarks of interest to tourists. For example, a tourist map of biostatistics for CIs might provide a minimum set of facts and focus instead on what is essential to facilitate communication with biostatisticians.

One argument in support of the conclusion that CIs should be taught differently from biostatisticians pertains to background knowledge. CIs typically lack the mathematical background to benefit from mathematically rigorous proof-based instruction, nor do they necessarily find formulae to be illuminating and intuitive. Instead, this audience responds well to concrete demonstration. Another argument in support of this conclusion pertains to coherence. Removing content from the standard curriculum for biostatisticians in deference to CIs might cause the remaining content to no longer fit together.

Our use case was a clinical research seminar for CIs, whose goal was to provide directed practice in applying high-level statistical concepts such as confounding, but bypassing much of the derivation, detail, and mathematical orientation by which this material would be taught to biostatisticians. In designing this seminar, at a big-picture "strategic" level, our approach included (1) naming the underlying constructs; (2) explicitly describing how biostatisticians conceptualize them (i.e., their mental map); (3) providing simple demonstrations; and (4) discussing journal articles which use those constructs and provide CIs with directed practice in applying them. At a more "tactical" level are the pedagogic techniques used to make those demonstrations effective. For example, while a MB student would gain insight by writing a computer program to explore how changing the inputs to a statistical technique would affect the outputs, a CRTP student might instead use an app that illustrates the same relationship, and thus provide insight about how the analytical "machine" functions.

An additional benefit of reconceptualizing the pedagogic task in this fashion is that we are in a better position to help clinicians “more confidently decide what they should do” when performing a data analysis – that is, to allow us to address this consistent concern expressed by our students. Interestingly, this is accomplished not by training them to actually “decide” by themselves, but instead to be well prepared to communicate with a biostatistician and arrive at a decision collaboratively.

6. Conclusion

Clinical investigators have different background and interests than biostatisticians, and should be taught statistics differently. Concepts should be phrased in plain language, illustrations should replace mathematical derivations, and underlying statistical concepts should be explicitly named. As a result, clinical investigators and biostatisticians will (and should) have different but overlapping mental maps of statistics.

References

Andrade C. (2021). HARKing, cherry-picking, P-hacking, fishing expeditions, and data dredging and mining as questionable research practices. *Journal of Clinical Psychiatry, 82*(1), 20f13804. https://doi.org/10.4088/JCP.20f13804

Begg A. (2015). Constructivism: an overview and some implications. ACE papers. Retrieved from https://api.semanticscholor.org/CorpusID:124317876

Biggs, J. (1996). Enhancing teaching through constructive alignment. *High Educ, 32*, 347-364. https://doi.org/10.1007/BF00138871

Bland, J. (2004). Teaching statistics to medical students using problem-based learning: the Australian experience. *BMC Med Educ, 4*(31). https://doi.org/10.1186/1472-6920-4-31

Enders, F. (2011). Evaluating mastery of biostatistics for medical researchers: need for a new assessment tool. *Clin Transl Sci, 4*(6), 448-454. https://doi.org/10.1111/j.1752-8062.2011.00323.x.

Enders, F. T., Lindsell, C. J., Welty, L. J., Benn, E. K. T., Perkins, S. M., Mayo, M. S., Rahbar, M. H., Kidwell, K. M., Thurston, S. W., Spratt, H., Grambow, S. C., Larson, J., Carter, R. E., Pollock, B. H., & Oster, R. A. (2017). Statistical competencies for medical research learners: What is fundamental? *Journal of Clinical and Translational Science*, *1*(3), 146-152. https://doi.org/10.1017/cts.2016.31

Hornung, C. A, et al. (2018). Competency indices to assess the knowledge, skills and abilities of clinical research professionals. *International Journal of Clinical Trials, 5*(1), 46-53.

Miles, S., Price, G. M., Swift, L., Shepstone, L., & Leinster, S. J. (2010, Nov). Statistics teaching in medical school: opinions of practising doctors. *BMC Med Educ, 4*(10), 75. https://doi.org/10.1186/1472-6920-10-75

Neely, M. L., Troy, J. D., Gschwind, G. T., Pomann, G-M., Grambow, S. C., & Samsa, G. P. (2022). Preorientation curriculum: an approach to preparing students with heterogeneous backgrounds for training in a Master of Biostatistics program. *Journal of Curriculum and Teaching, 11*(4), 120-138. https://doi.org/10.5430/jct.v11n4p120

Oster, R. A., Lindsell, C. J., Welty, L. J., Mazumdar, M., Thurston, S. W., Rahbar, M. H., Carter, R. E., Pollock, B. H., Cucchiara, A. J., Kopras, E. J., Jovanovic, B. D., & Enders, F. T. (2015). Assessing statistical competencies in clinical and translational science education: one size does not fit all. *Clin Transl Sci*, *8*(1), 32-42. https://doi.org/10.111/cts.12204

Ong, J. C., Ulmer, C. S., & Manber, R. (2012). Improving sleep with mindfulness and acceptance: a metacognitive model of insomnia. *Behav Res Ther, 50*(11), 651-660.

Onken, L. S., Carroll, K. M., Shoham, V., Cuthbert, B. N., & Riddle, M. (2014). Reenvisioning clinical science: unifying the discipline to improve the public health. *Clin Psychol Sci., 2*(1), 22-34.

Oster, R. A., & Enders, F. T. (2018). The importance of statistical competencies for medical research learners. *J Stat Educ, 26*(2), 137-142. https://doi.org/10.1080/10691989.2018.1484674.

Pomann, G-M, Boulware, L. E., Cayetano, S. M., Dasai, M., Enders, F. T., Gallis, J. A., Gelfond, J., Grambow, S. C., Hanlon, A., Hendrix, A., Kulkarni, P., Lapidus, J., Lee, H-J., Mahnken, J. D., McKeel, J. P., Moen, R., Oster, R. A., Peskoe, S. P., Samsa, G., Stewart, T. G., Troung, T., Wruck, L., & Thomas, S. M. (2021). Methods for training collaborative biostatisticians. *Journal of Clinical and Translational Science, 5*(1), E26. https://doi.org/10.1017/cts.2020.518

Pomann, G-M., Boulware, L. E., Chan, C. C., Grambow, S. C., Hanlon, A. L., Neely, M. L., Peskoe, S. B., Samsa, G., Troy, J. D., Yang, L. Z., & Thomas, S. M. (2022, September 9). Experiential learning methods for biostatistics students: A model for embedding student interns in academic medical centers. *Stat, 11*(1). https://doi.org/10.1002/sta4.506

Samsa, G. P., Thomas, L., Lee, L. S., & Neal, E. M. (2012). An active learning approach to teach advanced multi-predictor modeling to clinicians. *Journal of Statistics Education, 20*(1), 1-31.

Samsa G. (2018). A day in the professional life of a collaborative biostatistician deconstructed: implications for curriculum design. *Journal of Curriculum and Teaching, 7*(1), 20-31. https://doi.org/10.5430/jct.v7n1p20.

Samsa, G. P., LeBlanc, T. W., Locke, S. C., Troy, J. D., & Pomann, G. M. (2018, August 3). A model of cross-disciplinary communication for collaborative statisticians: implications for curriculum design. 2018. *Journal of Curriculum and Teaching, 7*(2), 1-11. https://doi.org/10.5430.jct.v7n2p1

Samsa G. (2020). Using coding interviews as an organizational and evaluative framework for a graduate course in programming. *Journal of Curriculum and Teaching, 9*(3), 107-140.

Samsa, G. (2021). Evolution of a qualifying examination from a timed closed-book format to a collaborative take-home format: a case study and commentary. *Journal of Curriculum and Teaching, 10*(1), 47-55. https://doi.org/10.5430/jct.v10n1p47

Sonstein, S. A. et al. (2014). Moving from compliance to competency: a harmonizing core competency framework for the clinical research professional. *Journal of Clinical Research Best Practices, 10*(6).

Sonstein, S. A., & Jones, C. T. (2018). Joint task force for clinical trial competency and clinical research professional workforce development. *Frontiers in Pharmacology, 9*, 1148. https://doi.org/10.3389/fphar.2018.01148

Sonstein, S. A., et al. (2020). Leveling the joint task force core competencies for clinical research professionals. *Therapeutic Innovation & Regulatory Science, 54*(1), 1-20. https://doi.org/10.1007/s43441-019-00024-2

Sullivan, G. M. (2011, Sep.). Getting off the "gold standard": randomized controlled trials and education research. *J Grad Med Educ., 3*(3), 285-9. https://doi.org/10.4300/JGME-D-11-00147.1. PMID: 22942950; PMCID: PMC3179209.

Troy, J. D., Neely, M. L., Grambow, S. C., & Samsa, G. P. (2021). The Biomedical Research Pyramid: A model for the practice of biostatistics. *Journal of Curriculum and Teaching, 10*(1), 10-17. https://doi.org/10.5430/jct.v10n1p10

Troy, J. D., Granek, J., Samsa, G. P., Pomann, G-M., Updike, S., Grambow, S. C., & Neely, M. L. (2022, April 21). A course in biology and communication skills for master of biostatistics students. *Journal of Curriculum and Teaching, 11*(4), 77- https://doi.org/10.5430/jct.v11n4p77

Troy, J. D., McCormack, K., Grambow, S. C., Pomann, G-M., & Samsa, G. P. (2022). Redesign of a first-year theory sequence in biostatistics. *Journal of Curriculum and Teaching, 11*(8), 1-12. https://doi.org/10.5430/jct.v11n8p1

Troy, J. D., Pomann, G-M., & Neely, M. L., Grambow, S. C., Samsa, G. P. (2023). Are simulated coding interviews a fair and practical examination format for non-professional programmers enrolled in a master's degree program in biostatistics? *Journal of Curriculum and Teaching, 11*(15), https://doi.org/10.5430/jct.v12n6p253

Troy, J. D., Neely, M. L, Grambow, S. C., Pomann, G-M., & Samsa, G. P. (2024). A curriculum review of programming courses in a master of biostatistics program. *Journal of Curriculum and Teaching, 13*(1), 405-420.

https://doi.org/10.5430/jct.v13n1p405

Appendix 1: Excerpts from CRP247 Syllabus (Lightly Edited for Clarity)

Your general goals in the CRTP

Most of you will be collaborating with a statistician rather than performing statistical analyses entirely on your own.

To qualify this statement: some investigators nevertheless find it helpful to perform data visualization, preliminary analysis, etc., in order to get in better touch with their data and inform their discussions with their statistician. That can be helpful, although unless you practice these skills regularly they will atrophy. In any case, differentiating between what you know and what you don't know is essential.

What all of you are trying to develop is "high-level" statistical skills -- for example, how to choose the correct analysis, how to interpret the results (e.g.., are they biased?), how to design studies, how to describe the outline of a statistical analysis plan in a grant proposal, etc. This is in contrast to the "low-level" technical skills such as learning the mathematical formula for a t-test, the R code for that t-test, etc. This isn't to denigrate the low-level skills, but simply to observe that you'll typically be delegating these technical tasks, and that your research career depends on the development of high-level rather than low-level statistical skills.

A challenge for learning high-level statistical skills is that they are most easily taught in an environment which assumes mastery of their low-level technical counterparts. For example, analysis strategy for regression modeling is much simpler to both teach and learn if things such as the mathematical formula for a regression model, the R code for running a regression model, the meaning of the regression coefficients, etc., can all be assumed known, which allows the entire focus to be placed on the choice of which particular regression models to run and why. From the perspective of curriculum design, the fact that (in general) you lack the time, interest, and prerequisite knowledge to master the associated technical skills is one of the unique characteristics of designing statistical instruction for clinical researchers.

From the perspective of a student, what you've likely encountered to date (e.g., in your previous statistics classes) is a combination of low-level information about "how to do it" and high-level information about "what to do and how to interpret the results". By necessity, these two types of information have been provided at the same time, and so at this stage in the program students often feel that "I can more or less execute a statistical analysis if you tell me what to do but am still uncertain about how to select it, interpret the results and, more generally, am not quite certain how everything fits together". In other words, at this stage in the program students aren't fully confident about the high-level statistical skills which are of most interest to them.

I hope that the above has convinced you that a lack of complete confidence in your high-level statistical skills is normal and, indeed, a consequence of how information has been presented to you to date. For the purpose of this course, I'm assuming that your primary interest continues to be in mastering high-level statistical skills, that you are part but not all the way there, and also that you'd like another bite of this particular apple.

Goals of this class

My primary goal is to help you solidify your understanding of high-level statistical concepts through considering various topics from a big picture perspective. In the language of constructivism: we will explicitly describe the mental maps used by statisticians and practice applying those mental maps in contexts which are relevant to clinical research.

Class session organization

Pre-class

• Please watch the Zoom video, which summarizes the content for the upcoming class session

• Please read the manuscripts, with the questions for discussion in mind

During class

• Discussion of additional questions from previous class

• Quick review of Zoom video with questions, led by one of the class members

• Directed discussion of manuscripts and overall content

After class

• Forward additional questions from class, requests for additional information, etc., which will be covered at the start of the next class

Class session topics

To qualify as a topic it should be important, commonly encountered in the literature and/or research projects, is sometimes misunderstood, and qualifies as a high-level statistical competency rather than a technical one. Topics include:

1. Logic of statistical inference: what does it mean to fit a model to data?
2. What is a sampling distribution? What is bootstrapping?
3. What is confounding? How can I address it (e.g., adjustment, propensity scoring)?
4. How do I select the variables for my statistical analysis?
5. How do I validate a statistical model?
6. How do I interpret subgroup analyses? What is an interaction?
7. How should I approach missing data?
8. How do I design a pilot study? What conclusions can be drawn?
9. How do I design a definitive (i.e., non-pilot) study? How do I estimate the sample size?
10. How do I develop a statistical analysis plan? How do I write the statistical methods section of a grant?
11. How is a statistical review performed?

Appendix 2: Slides on Subgroup Analyses and Lnteraction

Module 3: Subgroup analysis and interaction

1

The problem

- Subgroup analyses tend to be overinterpreted
  - We'll illustrate why and how things can go wrong
  - We'll touch on multiple inference, interpretation of p-values
  - Related concepts include regression toward the mean
  - Much of the literature pertains to randomized trials, but the ideas extend to observational studies

2

How subgroup analyses can be reported

- Fraudulent
  - Report the most significant subgroup result, while pretending that you didn't look at the others
- Wrong
  - Report point estimates by subgroup, but not confidence intervals
- Misinformed
  - The results were statistically significant in group A ($p<.05$), not statistically significant in group B, and therefore different
- Questionable
  - Even with caveats, focus on the most significant subgroup result

3

Fraudulent

- Reporting only the most significant results (or only the results which favor your hypothesis) induces a bias in favor of your hypothesis
  - This is, without question, unethical
  - Including caveats, but then behaving as if this is all you found is marginal at best

4

Wrong

- Outcomes improved by 5 units in I but only 1 unit in C
  - Instead, you should describe the noise (i.e., the precision of the estimate) in addition to the signal

5

Misinformed (outcome is mean change)

- Group A: 5.0 (0.2 to 9.8)
- Group B: 1.0 (-3.8 to 5.8)

- Mean change in group A was statistically significant (different from 0) in group A but not in group B. Therefore, the treatment works in group A but not in group B. This reasoning is common in the biomedical literature, but wrong.

- Problem: The confidence intervals overlap. The test for interaction will be non-significant. There isn't an actual difference.

6

Especially problematic

- Statistical test for the primary hypothesis is non-significant
- Efficacy is discovered in a subgroup which is defined *ex post*

7

Recommendations for subgroup analyses

- Consistent with the underlying science
- Specified ahead of time in the statistical analysis plan
- Preceded by a formal test for interaction
- Account for false positives associated with performing multiple analyses
  - This accounting could be formal or informal

8

Interaction ("effect modification") definition

- The impact of X on Y depends on the level of Z
  - The impact of treatment (X) on survival (Y) depends on the presence or absence of a genetic signature (Z)
  - Qualitative interaction: the impact… substantially depends …
  - Quantitative interaction: the impact… differs, but not substantially…
  - Substantial: the difference is large and/or in a different direction

9

Testing for interaction

- Include an interaction term X*Z as a predictor variable (in addition to X and Z), test whether the regression coefficient(s) associated with the interaction term equal 0
- Fit a full model with X, Z and X*Z and compare its fit with that of a reduced model with X and Z (but no interaction)
  - These 2 approaches are identical

10

Interaction and confounding

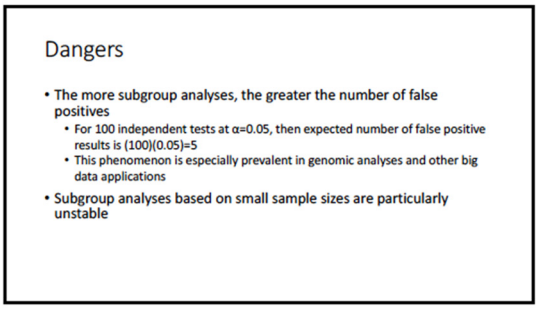
- Both interaction and confounding include a third variable, but they are different constructs
- Interaction: the impact of X on Y depends on Z
- Confounding: when assessing the impact of X on Y, the results will depend on whether or not Z is included in the model

11

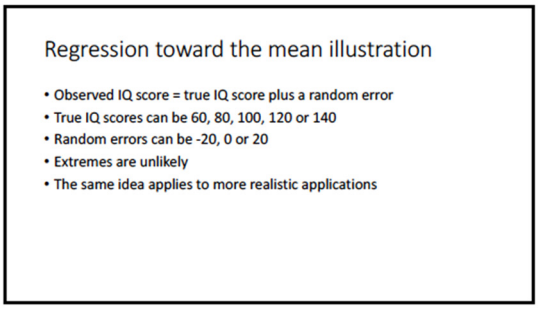
Questionable

- Focusing on the most dramatic subgroup effect, even with caveats, is dangerous
- Rationale: Such results are unstable, and will regress toward their mean

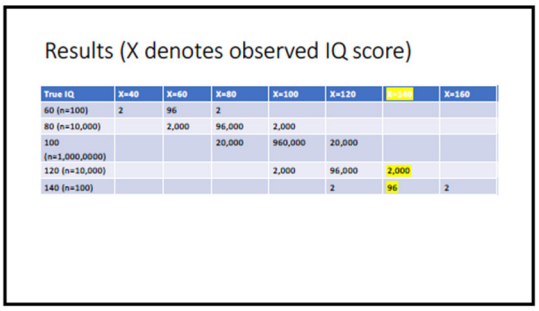
12



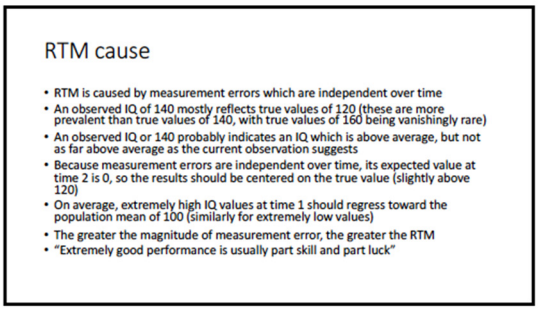
13



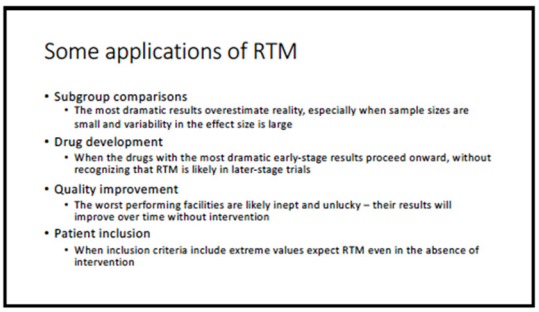
14



15



16



17

Acknowledgments

We greatly appreciate the active engagement our students demonstrated with the course described here, as well as their evaluations of its effectiveness.

Authors contributions

GPS taught the course described here and was the primary author of this communication. SCG, MLN, GMP, CD, MA and JET provided critical comment on its content. All authors are engaged in an ongoing program of pedagogic research around the methods described here. All authors read and approved the final manuscript.

Funding

Not applicable.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

Open access

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.