Preorientation Curriculum: An Approach for Preparing Students with Heterogenous Backgrounds for Training in a Master of Biostatistics Program

Megan L. Neely[1,*], Jesse D. Troy[1], Gerald T. Gschwind[2], Gina-Maria Pomann[1], Steven C. Grambow[1] & Gregory P. Samsa[1]

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

[2]President, Symphony Learning Partners LLC, Raleigh, NC, USA

*Correspondence: 2424 Erwin Road, Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC, 27705, USA. Tel: 919-684-8783. E-mail: megan.neely@duke.edu

Received: March 21, 2022 Accepted: April 13, 2022 Online Published: April 21, 2022

doi:10.5430/jct.v11n4p77 URL: https://doi.org/10.5430/jct.v11n4p77

Abstract

We describe an innovative preorientation curriculum (POC) for a Master of Biostatistics (MB) program. The goal of the POC is to fill critical skills gaps for students entering the MB program from heterogeneous backgrounds so they are prepared to engage in the program's rigorous, fast-paced training upon arrival. To achieve this goal, we introduce a structured approach to thinking that forms the foundation of a sound mental map of Biostatistics, which will assist students in their subsequent efforts to master the discipline. Based on constructivist principles, the POC covers mathematical and statistical theory, data analysis methods, programming, and statistical practice through a sequence of instruction that encourages reflection, extension, and connection between topics. The POC is modular, self-paced, and offered online via a cloud-hosted interactive learning management system (LMS). Students are required to complete the curriculum prior to the MB program orientation. We describe the rationale, design, features, and initial evaluation of the POC. Finally, to help programs interested in designing similar curricula, we provide a detailed instruction sequence description of one topic.

Keywords: biostatistics, preorientation curriculum, constructivism, curriculum development, curriculum evaluation

1. Introduction

The primary mission of our Master of Biostatistics (MB) program is to train collaborative biostatisticians who will successfully integrate into multi-disciplinary scientific teams after completion of the program. Successful students will be effective in data management and analysis, problem definition, problem solving, and communication (Pomann et al., 2020; Samsa, 2018a; Samsa, 2018b; Zapf et al., 2019) When developing and evaluating our program, we use as our organizing structure the "ABCs of biostatistics":

- **Analytical skills** include mastery of statistical analysis methods and statistical programming, as well as general skills in problem definition and problem solving.
- **Biology skills** include a combination of subject matter proficiency in biology and medicine with development of critical thinking skills related to the integration of biological concepts into research study design and data analysis.
- **Communication skills** include taking in information from outside the discipline and effectively sharing statistical information with others, both of which are fundamental to team science.

To achieve this mission, we have developed a fast-paced, rigorous 2-year graduate program. The ABCs are targeted in the four first-year core sequences, each consisting of two full-credit semester-long courses required of all students. Specifically, the core sequences are mathematical statistics, applied statistical methods, statistical programming, and statistical practice. This foundation is further developed in the second year where elective coursework builds on these foundations and fosters connections between the ABCs. A distinctive feature of our program is the emphasis on biology and especially communication; the first is often overlooked in graduate-level biostatistics training programs, and the second is often given minimal coverage in the form of seminar or consulting courses. For more details about the MB program curriculum, see Appendix A. Our MB program is a 2-year graduate degree housed in a School of Medicine and matriculates approximately 50 students every fall.

Given the accelerated nature of our MB program, it is essential that students enter the program with an “actionable foundation” in the necessary pre-requisites, which we define as training in calculus and linear algebra and exposure to biology, medicine, or other scientific areas through courses or work experiences. An “actionable foundation” goes beyond knowing facts; it also requires that students can synthesize information to formulate the problem and sort through known facts to determine which are relevant for solving the problem. Based on insights gained from a program-wide review (described in the next section) and our experience developing, administering, and teaching in the MB program, we found that good marks in pre-requisite coursework are imperfect surrogates for an “actionable foundation.” Students tend to be relatively stronger at the concrete elements of the pre-requisite training (for example, how to integrate a probability density function over the support) and relatively weaker on applying concepts in practice (for example, recognizing the integrand is the desired answer and why that number is meaningful in the context of the problem).

As the need for the integration of biostatisticians to collaborate on research teams across medicine continues to grow (DeMets et al., 2006; Davidson et al., 2020; Sharp et al., 2021), graduate programs need to have the ability to efficiently train students from variety of backgrounds. Therefore, a secondary mission of our MB program is to make the training available and accessible to a broad range of students with quantitative acumen and potential. Admitted students typically have undergraduate degrees in subjects like mathematics, statistics, biology, and public health. Most admitted students are moving directly from undergraduate to graduate training, although some enter with several years of work experience or enter the program while maintaining their current position, often in a related field. As such, our student population is heterogenous, and our experience is that students from different backgrounds tend to be stronger in different parts of the pre-requisite training – math-related majors challenged more by the biology material and vice versa for the biology-related majors. Although not required pre-requisites, we have also noted an increasing number of students entering the MB program with exposure to introductory statistics and data science courses. In parallel, an increasing number of student comments in the first-year course evaluations suggest that course content is introduced too slowly. Based on these comments and the program-wide review, we felt this more introductory material could be streamlined, allowing us to include additional content in the first-year core courses focused on higher-order skills like interpreting data, selecting an analysis plan, and choosing between statistical tests. Doing so would provide students with a more comprehensive set of skills by the end of the program’s first-year training and would better position them to receive competitive internships and research opportunities within and outside the university.

To address these issues, we have developed a preorientation curriculum (POC)that (1) ensures students from heterogenous backgrounds enter the program with the same, requisite, actionable body of knowledge and (2) allows first-year course instructors to eliminate some introductory content so that more instruction time can be spent on higher-order skills like conceptualization, communication, and advanced applications. We accomplished this by introducing “how to think like a biostatistician” and ensuring that students have a working knowledge of a core set facts prior to entering the program. We describe here the development and initial evaluation of the POC, beginning with a description of the design process. Then, we describe the structure and features of the POC. Finally, we describe our evaluation efforts to date and reflect on what we have learned about developing and implementing a POC and its potential benefits.

2. Design and Development of the POC

The POC was designed and developed from October 2019 through May 2020 by an experienced team of three faculty members, the Director of Graduate Studies for the MB program, and an instructional designer. Team meetings were held weekly and agreement on the curriculum and instructional approach were arrived at through an informal, iterative consensus process. The first iteration of the POC was offered in June 2020 for students matriculating in the Fall 2020 semester.

The POC was designed to be a modular, self-paced curriculum offered online via a cloud-hosted interactive learning management system (LMS), TalentLMS (https://www.talentlms.com/). The POC consists of 4 content modules, one for each of the first-year core course sequences: statistical practice, statistical programming, mathematical statistics, and applied statistical methods. Each module contains content covering a set of topics, short practice exercises, and a cumulative assessment. Students are required to complete the curriculum prior to the MB program orientation.

A preliminary outline of the POC content was developed by reviewing topics covered in the fall semester of first-year core courses and identifying those most naturally shifted to the POC. This was part of an extensive program review effort which was intended to better align program goals, admission criteria, and teaching methods. Step 1 being to remove unnecessary content from the first-year courses; Step 2 being to divide the remaining content into POC, fall semester and spring semester; Step 3 being to add new content, primarily to the spring semester; and Step 4 being to better integrate the content across the core course sequences. The following POC modules were created (one for each of the 4 core course sequences):

- Module 1: Biology and Communication
- Module 2: R Programming
- Module 3: Mathematical Statistics
- Module 4: Applied Statistical Methods

Module 1 was entitled “Biology and Communication” instead of “Statistical Practice” because the first course in the statistical practice sequence focuses on biological concepts and communication skills. It makes the connection between biological sciences and biostatistics more explicit by teaching students how the underlying science impacts statistical aspects of the research study and how oral and written communication are vital in the team science setting (Troy et al., 2021). Similarly, Module 2 was entitled “R Programming” instead of “Statistical Programming” because the first course in the statistical programming sequence focuses on the R programming environment. Through R, students are introduced to what statistical programming is and why it is essential to the practice of biostatistics.

Topics embedded within each module were carefully balanced to present a core set of facts students need to be successful in the first-year courses in combination with an introduction to information that teaches students “how to think like a biostatistician.” This approach ensures information presented in the POC is actionable, not just a collection of imparted facts, and teaches students how to translate biostatistical facts into biostatistical practice. The goal of this balanced approach is to help students create a sound mental map of biostatistics that prepares them to learn the higher-order skills taught in the MB program while equalizing their statistical knowledge base regardless of their heterogenous backgrounds.

Once module topics were determined, modules were then assigned to individual faculty committee members to develop the instructional content, including lecture slides and/or packets, video lecture and demonstration scripts, practice exercises, and cumulative module assessments. Faculty then worked with an instructional designer and his production team to turn their module content into interactive e-learning content delivered to students using the LMS. All instructional content was reviewed by the committee and the first-year instructors in the relevant courses. Consensus with first-year instructors was particularly important to ensure a coordinated transition between the POC and first-year fall courses. Videos were pilot tested by two student volunteers enrolled in the MB program and their feedback incorporated as appropriate.

3. Features of the POC

3.1 Content Overview

The POC is offered as a 12-week self-paced online course that starts in late May and finishes prior to new student orientation in the third full week of August. Because they provide prerequisite information for the others, students must complete the Biology and Communication and R Programming modules prior to starting the Mathematical Statistics and Applied Statistical Methods modules. However, students may work asynchronously on the Biology and Communication and R Programming modules until both are complete. To help students plan and stay on track, the program recommends that students complete the first two modules no later than 6 weeks after course activation. The expected time commitment is no more than 10 hour per week. Students may progress through the material at their own pace; there are no restrictions on how early the POC can be completed. Instructors are available to answer questions, grade assignments, and send written feedback to students through email.

3.2 Content Structure

Each module starts with an overview, which describes the information to be covered, how that information is used in biostatistics (and, thus, why it is important), and how the information relates to the content of the other modules (and, thus, how they fit together). For example, the Applied Statistical Methods module is organized around a complete data analysis exercise, beginning with the conceptualization of the study question and biological basis for the research (linking with the Biology and Communication module), including directed data analyses (linking with the R

Programming module), and ultimately proceeding to interpretation of results. Once the stage is set for each module, students proceed through a series of module-specific topics. Topics covered in each module are displayed in Figure 1.

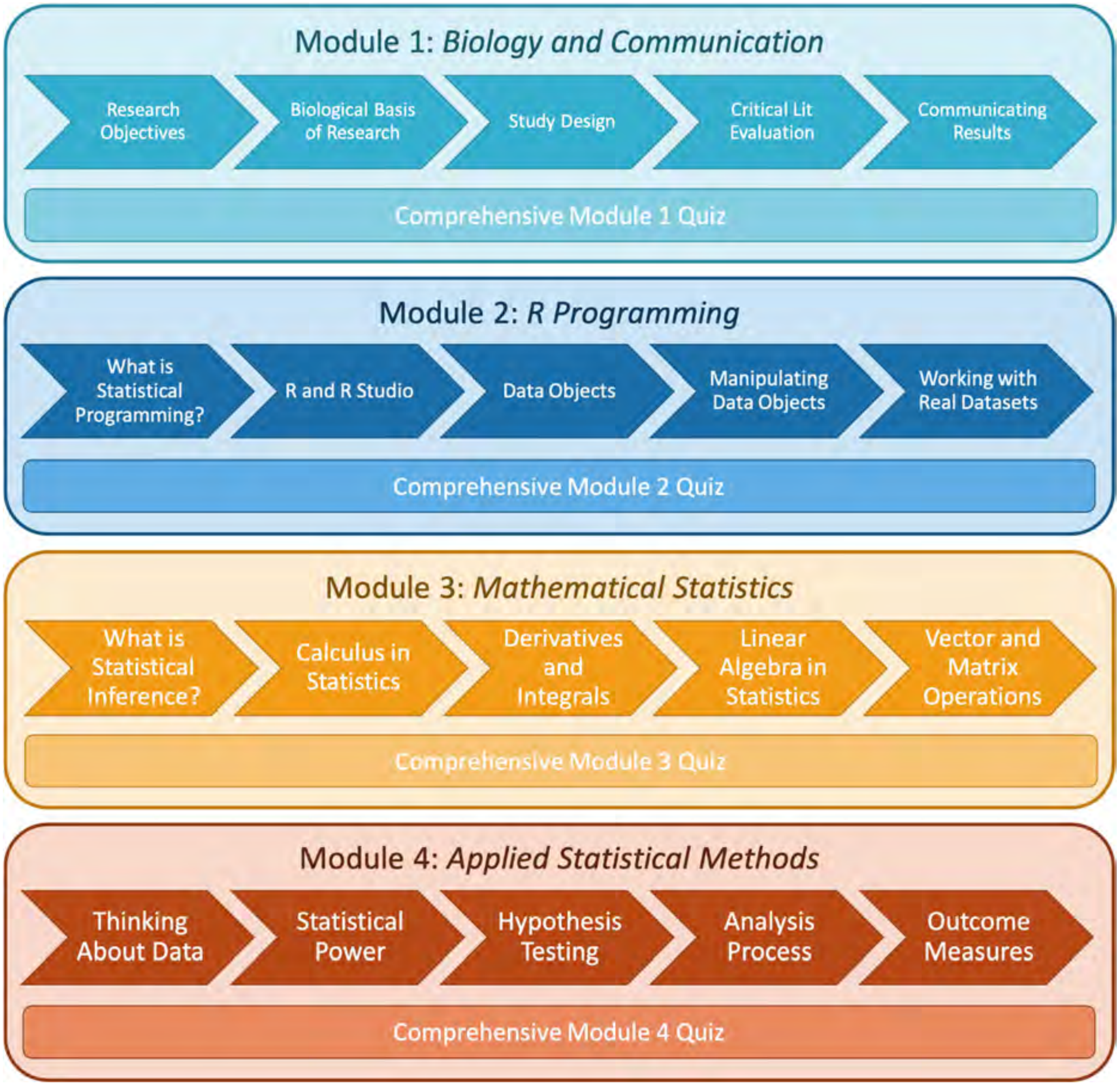

Figure 1. Topics Covered in Each of the Four Preorientation Curriculum Modules to Prepare Students with Heterogeneous Backgrounds to Engage in the Rigorous, Fast-Paced Training of the Master of Biostatistics Program Upon Arrival

Based on the principles of constructivism (Biggs, 1996), each topic within a module is presented with the following instructional sequence: begin with a conceptual introduction, follow with illustrations, reinforce through exercises that extend and require reflection, and conclude with connections to content presented in other topics within and across modules. An example of the instructional sequence is presented in Figure 2 for the Statistical Power topic in the Applied Statistical Methods module. The presentation begins at a conceptual level, with a contextual working definition as well as an intuitive mathematical definition of power. Once the concepts are clarified, a power calculation is illustrated using an R function which builds upon information from the programming module. This provides students with content that teaches both "how to think" and "how to do." Students then complete exercises that both reinforce the content and challenges them to extend and integrate these newly-learned skills in service of solving a practical biostatistics problem. Finally, connections to other topics and modules are presented, offering students an opportunity to reassess and realign their mental map of the material.

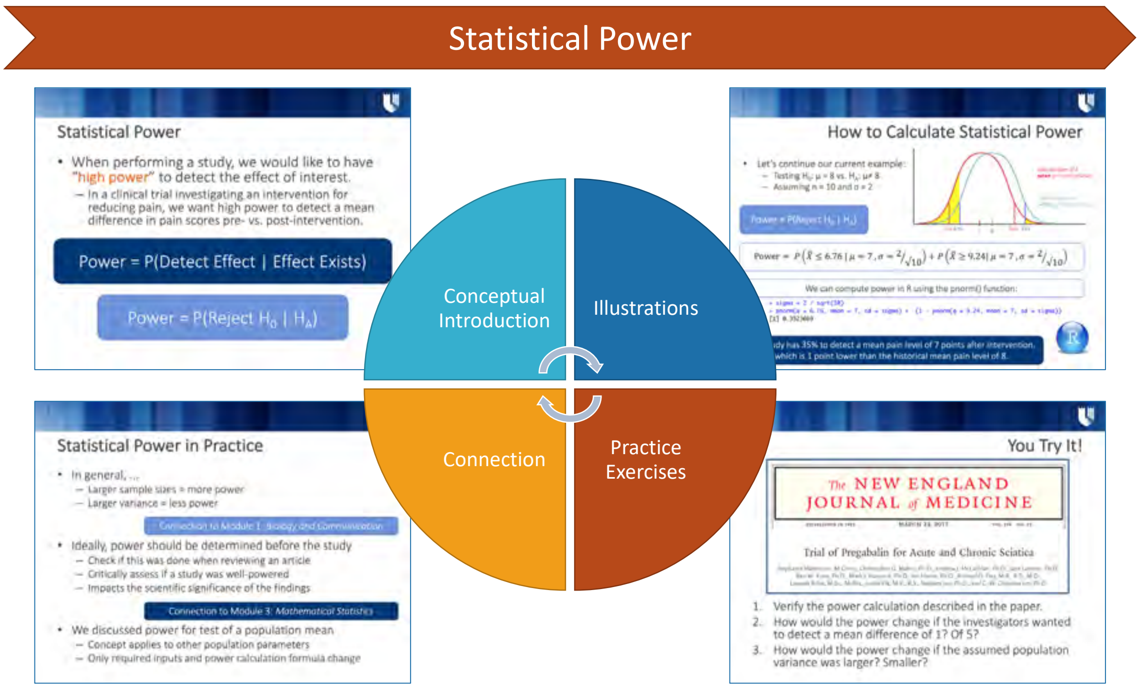

Figure 2. Example of the Instructional Sequence for the Statistical Power Topic in the Applied Statistical Methods Module to Teach Students both "how to think" and "how to do" Biostatistics

Students successfully complete a module when they earn a passing score on the comprehensive module assessment. The assessments are comprised of multiple-choice questions, which are graded automatically by the LMS, and/or short answer questions that requiring writing and explanation, which are graded by the instructor assigned to each module. There is no limit on the number of attempts, and assessments were carefully designed to prevent students from succeeding through iterative guessing. Students may proceed directly to the module assessment without viewing the content if they feel confident they can pass without reviewing the material. However, students must pass the module assessments for the Biology and Communication and R Programming modules prior to attempting the module assessments for the Mathematical Statistics and Applied Statistical Methods modules.

3.3 Content Delivery

All POC content is delivered to students using a cloud-hosted interactive LMS which can automate course assignment and user management via triggering of specific actions according to specified parameters. Using this automation functionality, the POC content is delivered to students following the structure shown in Figure 1. This allows students to proceed through the content at their own pace, rather than waiting for instructors to manually release or upload new content at regular intervals. The only aspect of the POC that is not fully automated is the grading of short answer questions on the cumulative module assessments. These questions, which require writing and explanation, are graded by the instructors who provide written feedback to the students. The LMS notifies instructors when an assessment has been submitted by a student, enabling the instructor to quickly complete the grading so that students can continue their progress through the POC.

One key aspect of the POC delivery is the use of different media formats, allowing each instructor to tailor their delivery method to the one best aligned with their module's content. For example, the R Programming module interweaves written content, used to introduce programming terminology and concepts, with narrated video content, used to provide programming demonstrations of those concepts in the R environment. Other modules use narrated PowerPoint presentations supplemented by more detailed materials and hands-on examples based on real-world scenarios. Switching delivery methods helps refocus the students' attention and maintain engagement as they proceed through the material (David et al., 2014).

Another key aspect of the POC is the integration of feedback opportunities built into the content at every step along the way. This helps ensure students are prepared for the cumulative module assessments and ultimately for them entering the MB program after completing the POC. Short, automatically scored assessments, typically multiple-choice and short answer questions, are embedded in a just-in-time fashion at the end of each topic within a module. These assessments provide immediate feedback to the student about their level of understanding and can guide the student to module topics that should be reviewed prior to starting the cumulative module assessment. User groups, hosted by an instructor, provide students with another avenue of feedback by allowing the opportunity to discuss POC content amongst themselves and with an instructor. This progressive sequence of assessments helps students gradually build confidence with the material and maintain motivation while working independently (Keller, 2010).

4. Evaluation of the POC

We directly evaluated the POC as (1) software, (2) a mechanism for transmitting specific information, and (3) a mechanism for ensuring students have an actionable foundation based on that specific information. We completed the evaluation by examining student performance metrics from the Summer 2020 offering of the POC and by soliciting feedback from the Fall 2020 incoming students (i.e., the target audience) at the end of their first semester in the MB program. This approach allowed us to obtain student input on the POC's utility after completing the first half of the first-year core sequences.

We evaluated the POC as software by performing standard usability testing throughout the development process. This began with POC and first-year core course instructors testing the POC to uncover any issues in the LMS implementation, including errors in the content, mistakes in the navigation control, and incorrect answers coded in the assessment rubrics. The final step was soliciting comment from intended users, the incoming students. The usability reports were favorable. Both students and faculty found the LMS easy to navigate. In the student surveys, evaluation of how easy the POC software was to use had a mean (median) of 8.2 (9.0) on a scale of 0 (terrible) to 10 (wonderful). In their open-text feedback, students mentioned appreciation of the short introduction video produced by the instructional design team on how to use the LMS. Several students mentioned that it took a long time for the LMS webpage and videos to load. We believe this is most likely caused by variability in students' home internet speed. In response, we updated the program documentation on the POC to include this observation and to suggest students find access to high-speed internet when competing the POC if possible. Students also mentioned they had to complete an entire section for their work to be saved in the LMS; they could not pause in the middle of a section and pick up their work later. This is a limitation of the LMS we chose; however, prior to releasing the Summer 2021 offering of the POC we carefully reviewed each module's elements to break them down into their smallest, meaningful sub-units to help mitigate this issue.

We evaluated the POC as a mechanism for transmitting specific information by examining student performance on the four comprehensive module assessments. The student performance results were encouraging. The mean (standard deviation) score vs. median (interquartile range) score was 98.0% (2.5%) vs. 100% (5.0%) on the Biology and Communication module, 100% (0%) vs. 100% (0%) on the R Programming module, 97.8% (2.8%) vs. 100% (5.0%) on the Mathematical Statistics module, and 98.6% (1.4%) vs. 97.5% (2.5%) on the Applied Statistical Methods module. The high average scores and low variability suggest that students left the POC with a solid understanding of the core content presented in each module and that this held true across all students, regardless of their backgrounds.

We evaluated the POC as a mechanism for ensuring students have an actionable foundation in that specific information by examining the quality of student responses on the short answer questions from the cumulative module assessments. Following the principles of constructivism, these short answer questions are open-ended to encourage critical thinking and emphasize reflection, extension, and connection between topics within and across modules. The results of this evaluation were encouraging. Student responses to the short answer questions were generally correct, and the quality of their written communication (a programmatic focus) generally good. Moreover, the questions contained sufficient nuance to help assess which concepts were understood superficially rather than deeply. After turning in the assignment, students received example answers written by the instructor in addition to written feedback on the answers they submitted. In total, these observations suggest that students left the POC with an actionable body of knowledge, demonstrating that they were able to synthesize the content presented in the POC and apply it to answer complex questions that required integration of knowledge gained from each of the four POC modules. These observations also indirectly suggest that the POC is a useful tool for laying the foundation of sound mental map of biostatistics.

The POC was generally well-received by students. Student evaluation of the POC on a scale of 0 (terrible) to 10 (wonderful) had a mean (median) of 7.74 (8.25) across the modules, and 80% of students responded “I see the connection between my current coursework and the material covered in the POC” when asked how relevant they found the POC material to their current coursework. When asked how many hours they spent per week completing each module, 70%, 70%, 100%, and 100% of students responded they spent less 10 hours to complete the Biology and Communication module, the R Programming module, the Mathematical Statistics module, and the Applied Statistical Methods module, respectively. These findings suggest our target of 10 hours per week, which translates to 30 hours per module if students split their time evenly across modules during the 12-week course, was reasonable. This suggests students from all backgrounds were able to complete the POC by the program deadline with a low time pressure, while receiving high marks on the cumulative module assessments, making the POC a relatively stress-free educational experience.

5. Discussion

Much has been written about the\features of successful collaborative biostatisticians that should form the basis of biostatistics training programs (DeMets et al., 2006; Perkins et al., 2016; Pomann et al., 2020; Tobi et al., 2001; Van Steen et al., 2001). Many have emphasized the need for “integration skills” and a “broad perspective” of the interdisciplinary research field (Begg & Vaughan, 2011). Our approach of focusing on developing a sound mental map of the discipline being studied is consistent with this recommendation for a strong foundation in the ABC’s of biostatistics, that can be expanded to incorporate new methods and domain-specific knowledge. By explicitly teaching the thought process behind biostatistical practice in conjunction with essential knowledge, the POC lays the foundation of students’ mental maps, better positioning them to engage in the program curriculum and creates the conditions for accelerated learning in the fall. We anticipate this will change the way students process the "facts" that we teach them. Rather than perceiving these facts as dangling bits of information that they are required to absorb without context, we hope that they will interpret them as knowledge that enables critical thinking. In turn, this might encourage students to be more active participants in their studies rather than passive (and perhaps reluctant) consumers of information that faculty present to them. We also anticipate the POC will reduce the number of “struggling students,” operationally defined as students requiring an academic remediation plan after their first semester and will result in better performance among the higher-performing students. Finally, we expect the POC will allow us to broaden our admission criteria, expanding the pool of well qualified applicants with quantitative acumen and potential. This is a good thing: the demand for excellent collaborative biostatisticians far exceeds the supply, and there aren’t nearly enough mathematics and statistics majors to fill the pressing needs in the field.

The POC presented in this paper is like a bridge program. Bridge programs have traditionally focused on improving career pathways for students from underrepresented backgrounds to the science, technology, engineering, and math (STEM) fields. These programs are often focused on providing additional training opportunities that will help students achieve success at the next step of their academic career, usually from high school to undergraduate programs (Bradford et al., 2021; Kitchen et al., 2018). Graduate bridge programs have been developed to ensure that students succeed in doctoral programs by exposing a variety of STEM students to appropriate materials before attending a graduate program (Purdue University, n.d.; University of California Merced, n.d.). However, to date these programs have focused on areas of graduate training applicable to all fields (for example, navigating advisor-advisee relationship, grant/dissertation writing, teaching), not on field-specific content like the POC presented in this paper. Future applications of the POC presented here could be integrated within such a program for the relevant students.

6. Conclusion

We have described the rationale, development, structure, and evaluation of an innovative POC for a 2-year MB program whose goal is to fill critical skills gaps for students entering the program from heterogeneous backgrounds, so they are prepared to engage in the program’s rigorous, fast-paced training upon arrival. Specifically, the POC ensures all students entering the MB program have the same, requisite, actionable body of knowledge prior to starting the program, which in turn allows first-year course instructors to eliminate some introductory content so that more instruction time can be spent on higher-order skills like conceptualization, communication. We achieved this by introducing students to “how to think like a biostatistician” in addition to presenting a core set of facts. This structured approach to thinking forms the foundation of a sound mental map of biostatistics, which will assist students in their subsequent efforts to master the discipline.

References

Begg, M. D., & Vaughan, R. D. (2011). Are biostatistics students prepared to succeed in the era of interdisciplinary science? (and how will we know?). *American Statistician*, *65*(2), 71-79. https://doi.org/10.1198/tast.2011.10222

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, *32*(3), 347-364. https://doi.org/10.1007/BF00138871

Bradford, B. C., Beier, M. E., & Oswald, F. L. (2021). A meta-analysis of university stem summer bridge program effectiveness. *CBE—Life Sciences Education*, *20*(2), ar21. https://doi.org/10.1187/cbe.20-03-0046

Davidson, M.A., Dewey, C.M., & Fleming, A.E. (2018). Teaching communication in a statistical collaboration course: A feasible, project-based, multimodal curriculum. *The American Statistican, 73*(1), 61-69. https://doi.org/10.1080/00031305.2018.1448890

Davis, J., Balda, M., Rock, D., McGinniss, P., & Davachi, L. (2014). The science of making learning stick: An update to the AGES model. *NeuroLeadership Journal*, *5*, 1-15.

DeMets, D. L., Stormo, G., Boehnke, M., Louis, T. A., Taylor, J., & Dixon, D. (2006). Training of the next generation of biostatisticians: A call to action in the U.S. *Statistics in Medicine*. https://doi.org/10.1002/sim.2668

Keller, J. M. (2010). The Arcs model of motivational design. In *Motivational design for learning and performance* (pp. 43-74). Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-1250-3_3

Kitchen, J. A., Sadler, P., & Sonnert, G. (2018). The impact of summer bridge programs on college students' STEM career aspirations. *Journal of College Student Development*, *59*(6), 698-715. https://doi.org/10.1353/csd.2018.0066

Perkins, S. M., Bacchetti, P., Davey, C. S., Lindsell, C. J., Mazumdar, M., Oster, R. A., Peduzzi, P. N., Rocke, D. M., Rudser, K. D., & Kim, M. (2016). Best Practices for Biostatistical Consultation and Collaboration in Academic Health Centers. *The American Statistician*, *70*(2), 187-194. https://doi.org/10.1080/00031305.2015.1077727

Pomann, G.-M., Boulware, L. E., Cayetano, S. M., Desai, M., Enders, F. T., Gallis, J. A., Gelfond, J., Grambow, S. C., Hanlon, A. L., Hendrix, A., Kulkarni, P., Lapidus, J., Lee, H.-J., Mahnken, J. D., McKeel, J. P., Moen, R., Oster, R. A., Peskoe, S., Samsa, G., … Thomas, S. M. (2020). Methods for training collaborative biostatisticians. *Journal of Clinical and Translational Science*. https://doi.org/10.1017/cts.2020.518

Purdue University. (n.d.). Graduate Bridge Program. Retrieved from https://www.purdue.edu/gradschool/diversity/programs/bridge/

Samsa, G. (2021). Evolution of a Qualifying Examination from a Timed Closed-Book Format to an Open-Book Collaborative Take-Home Format: A Case Study and Commentary. *Journal of Curriculum and Teaching*, *10*(1), 47. https://doi.org/10.5430/jct.v10n1p47

Samsa, G. P. (2018a). 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. (2018b). A Model of Cross-Disciplinary Communication for Collaborative Statisticians: Implications for Curriculum Design. *Journal of Curriculum and Teaching, 7*(2), 1-11. https://doi.org/10.5430/jct.v7n2p1

Sharp, J. L., Griffith, E. H., & Higgs, M. D. (2021). Setting the stage: Statistical collaboration videos for training the next generation of applied statisticians. *Journal of Statistics and Data Science Education, 29*(2), 165-170. https://doi.org/10.1080/26939169.2021.1934202

Tobi, H., Kuik, D. J, Bezemer P. D., & Ket P. (2001). Towards a curriculum for the consultant biostatisician: Indentifcation of central disciplines. *Statistics in Medicine, 20*(24), 3921-3929. https://doi.org/10.1002/sim.1082

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

University of California Merced. (n.d.). *Competitive Edge Summer Bridge*. Retrieved from https://graduatedivision.ucmerced.edu/grad-resource-center/signature-programming/summer-bridge

Van Steen, K., Curran, D., & Molenberghs, G. (2001). Towards a curriculum for the consultant biostatistician: Identification of central disciplines. *Statistics in Medicine*, *20*(24), 3921-3929. https://doi.org/10.1002/sim.1082

Zapf, A., Huebner, M., Rauch, G., & Kieser, M. (2019). What makes a biostatistician? *Statistics in Medicine*, *38*(4), 695-701. https://doi.org/10.1002/sim.7998

Appendix A

MB Program Curriculum Overview

The MB program is a two-year graduate degree housed in a School of Medicine. Students take the same set of courses during their first year to build a foundation, and then differentiate during their second year.

The four first-year academic course sequences are:

- Mathematical statistics – the "why" behind statistical methods
- Applied statistical methods – the toolbox of standard statistical techniques
- Statistical programming – the ability to manage and analyze data using statistical software
- Statistical practice – skills and techniques for practicing statistics within the context of biomedical team science, including communication, study design, and basic biology and medicine.

Each first-year sequence begins with a fall course, followed by another course in the spring. The theory sequence is designed to be an integrated presentation with a break for the holidays, as is the data analysis sequence. The programming sequence covers the commonly used statistical programming languages, with all students taking a course in R in the fall and students choosing a course in SAS or Python in the spring. The statistical practice sequence focuses on biology, study design and clinical epidemiology in the fall (Troy et al., 2021) and elements of team science such as consulting technique and the development of statistical analysis plans and reports in the spring.

Students also participate in professional development training in a short-course sequence where they develop a resume, participate in interviewing and networking exercises, and practice other professional skills. The first year concludes with a comprehensive qualifying examination covering the four first-year core sequences (Samsa, 2021).

The second year includes various elective courses, an internship / practicum experience where students collaborate on research projects under supervision, and a masters' project. Students can choose from electives covering statistical methods that can be applied across many biomedical contexts, including survival analysis, categorical data analysis, and correlated and longitudinal data analysis. Students can also choose from electives that cover the context in which the statistical methods are applied, including clinical trials, observational studies, predictive medicine, and statistical genetics.

The selection of elective courses and master's project topic is guided by a student's choice of 1 of 3 tracks in the MB program. Tracks are chosen prior to starting the second year after students pass the qualifying exam. The 3 tracks are:

- Clinical and translational research track
- Biomedical data science track
- Mathematical statistics track

The tracks allow students to tailor their training to their interests and post-graduation goals. Each track builds on the biostatistical foundation laid in the first-year core sequences and allows students to differentiate by focusing their second-year coursework on subspeciality areas in biostatistics, emphasizing different aspects of the biostatistical training they received in the first-year core sequences.