CUBE: A Collaborative Undergraduate Biostatistics Experience To Bring Diversity And Awareness To The Field Of Collaborative Biostatistics

Alicia J. Lozano, Monica L. Ahrens, Genevieve R. Lyons, Sarah J. Ratcliffe, Alexandra L. Hanlon, Jennie Z. Ma
Virginia Tech
University of Virginia

Abstract

Background: The field of collaborative biostatistics plays a critical role in translating scientific discoveries into practical applications by providing sound statistical analysis, ensuring data integrity, and facilitating interdisciplinary collaboration in translational science and decision-making in healthcare. The role of a collaborative biostatistician is particularly important in addiction and mental health research given the use of complex study designs and statistical methodology, along with large amounts of missing data and study dropout. Despite growth in the demand for collaborative biostatisticians and the number of graduate degrees in biostatistics earned in the United States in recent years, the percentage from underrepresented minorities remains low. To increase awareness and bring diversity to the field of collaborative biostatistics, the Collaborative Undergraduate Biostatistics Experience (CUBE) was developed. CUBE is built on four pillars: training in introductory biostatistics, R programming, professional development, and mentorship through a collaborative biostatistics research project focused in the areas of addiction and mental health. This paper provides an overview of the key components of the CUBE curriculum, evaluation metrics, and results from the CUBE cohort of summer 2023 (N = 5).

Results: Eighty percent of participants were female students entering their fourth year of undergraduate study. From pre- to post-program, participants demonstrated increased knowledge in the overall research process, ethical and professional aspects of research, basic concepts in research, data analysis, data management, report preparation, and presenting research. With respect to future education and career plans, most participants felt more confident in planning for their graduate school education and pursuing a career in biostatistics or research.

Conclusions: Results from CUBE 2023 highlight its positive impact on participants, particularly in terms of knowledge enhancement and future aspirations. Most participants experienced significant improvements in their understanding of various aspects of the research process. The program played a pivotal role in fostering confidence, and thus positively influencing their future. These results affirm the effectiveness of the CUBE program in equipping aspiring researchers with the skills and confidence needed for successful endeavors in the fields of biostatistics, research, and translational science, with an emphasis on addiction and mental health.

Introduction

Biostatistics refers to the application of mathematical and statistical methods to the biological and medical sciences. Collaborative biostatistics is a specialized area of biostatistics that focuses on the collaboration between biostatisticians and researchers from various disciplines, such as addiction and mental health. While numerous academic training programs and courses in statistics and biostatistics are offered across the United States (US), these programs do not traditionally emphasize the collaborative aspects of the field (Begg & Vaughan, 2011; Pomann et al., 2022). As a result, students may not be adequately exposed to the real-world applications of biostatistics in interdisciplinary research during their academic training and lack awareness of the field of collaborative biostatistics and related career opportunities.

Approaches to clinical and health-related research have become increasingly interdisciplinary, leading to an increase in the need to integrate collaborative biostatisticians within research teams to meet this demand (Boulware et al., 2021). The field of collaborative biostatistics plays a fundamental role in clinical and health sciences research given its impact on clinical practice and decision-making, as well as health outcomes (Boulware et al., 2021).

Although true in all disciplines, the role of a collaborative biostatistician is particularly important in addiction and mental health research given the types of problems commonly encountered, such as the use of complex study designs, large amounts of informative missing data and study dropout, along with the need to quantify complex mechanisms of action or underlying relationships (King et al., 2020; MacKinnon & Lockwood, 2003). In the planning stages of a research project, a collaborative biostatistician is responsible for ensuring that the study design and analytic approach are both rigorous and appropriate based on the specific research question(s) or aim(s), variable types, and data collection timing and approach. During data collection, a collaborative biostatistician can ensure data quality and provide guidance on data management and storage. After the data have been collected, a collaborative biostatistician works closely with the research team to provide high-quality analyses and to effectively communicate, interpret results appropriately, and disseminate the results of the study in an interdisciplinary environment.

Employment in the field of biostatistics is projected to grow 30% over the next ten years *(Mathematicians and Statisticians*: Occupational Outlook Handbook: US Bureau of Labor Statistics). Despite steady increases in the number of graduate degrees earned in biostatistics in the US since 2011, the percentage of Black, Latinx and Indigenous people earning master's degrees (4.8%, 5.5%, 0.3%, respectively) and PhDs (4.4%, 4.0%, 0.8%, respectively) in biostatistics has remained relatively low and unchanged (Benn et al., 2020; *Survey of Earned Doctorates (SED)*, 2022). The number of graduate degrees earned in science, technology, engineering, and mathematics (STEM) fields among students from underrepresented or minoritized backgrounds is consistently low (Benn et al., 2020; *Black Students in STEM and Health Graduate Programs Increase But a Large Racial Gap Remains,* 2021; Golbeck, 2020; *Survey of Earned Doctorates (SED*), 2022). Per the National Institutes of Health (NIH), the following groups are underrepresented in biomedical research: racial and ethnic minorities (Blacks or African Americans, Hispanics or Latinos, American Indians or Alaska Natives, Native Hawaiians, and other Pacific Islanders), women, persons with disabilities, and individuals from disadvantaged backgrounds *(Notice of NIH's Interest in Diversity,* 2019). Increasing diversity among collaborative biostatisticians can bring a range of perspectives, experiences, and ways of thinking to the field. This can help to ensure that cultural nuances and considerations are taken into account when designing studies, collecting data, and interpreting results (Slade et al., 2023).

Key factors contributing to the retention of students in STEM majors include creating inclusive and supportive learning environments, as well as establishing a close mentoring relationship with invested faculty and peers. Additionally, meaningful undergraduate research experiences and internships have shown to enhance student engagement and retention of student interest in future careers in STEM-related fields (Benn et al., 2020; Caldwell

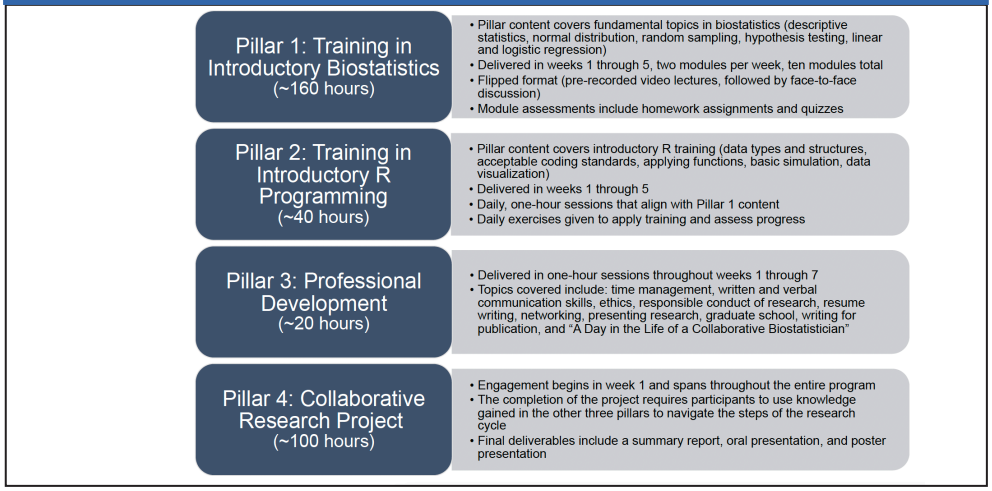
et al., 2021; Estrada et al., 2018).

The Collaborative Undergraduate Biostatistics Experience (CUBE) is an eight-week summer program designed to help close this gap by providing a unique opportunity for students from underrepresented or minoritized backgrounds in STEM from around the US to work together in a vibrant, collaborative academic environment with active researchers, under the mentorship of recognized experts from a variety of disciplines. By working closely with experienced collaborative biostatisticians, leading experts in various areas of research, peer and graduate student mentors, participants are given the opportunity to learn about the challenges of communicating and working collaboratively with researchers in a multi-disciplinary team science environment, with an emphasis on addiction and mental health. These activities provide a foundation for participants to build the confidence and knowledge needed to apply to, and succeed in, graduate school and careers in biostatistics and quantitative research.

The CUBE program was developed in spring of 2021, when Virginia Tech's (VT) Department of Statistics participated in an event hosted by the Department of Mathematics and Economics at Virginia State University (VSU), a historically Black college and university (HBCU). While there was a large amount of student interest in statistics and biostatistics, no evidence-based training opportunities in collaborative statistics or biostatistics existed at VT; programs that introduce students who come from underrepresented or minoritized backgrounds to the variety of methods and tools necessary to analyze data in biomedical research certainly did not exist. As a result of this experience, one VSU undergraduate student from an underrepresented or minoritized background was offered a summer internship position at VT's Center for Biostatistics and Health Data Science (VT CBHDS) and received informal training in introductory biostatistics, statistical programming, professional development, and was engaged in a collaborative research project under the mentorship of an experienced biostatistician.

Student and mentor feedback from this initial summer internship was applied to the pilot launch of the CUBE program in the summer of 2022, where four students from universities across the US were offered internship positions at VT CBHDS and the University of Virginia's Department of Public Health Sciences (UVA PHS), with two at each site. This paper focuses on the key components of the CUBE program curriculum, design, and evaluation metrics delivered in summer of 2023, after receiving funding from the NIH's National Institute on Drug Abuse (NIDA) and National Institute on Alcohol Abuse and Alcoholism (NIAAA) to support up to ten students per summer at VT CBHDS for five years, and additional funding from the National Center for Advancing Translational Sciences of the NIH to support two students at UVA PHS.

Figure 1. Four Pillars of the CUBE Program



CUBE Program Design

Program Goals and Curriculum The goals of the CUBE program are to bring awareness to, and promote diversity, equity, and inclusion to the field of collaborative biostatistics. Guided by Lent, Brown, and Hackett's (1994) Social Cognitive Career Theory (SCCT) conceptual framework (Lent et al., 1994; Stewart et al., 2020), the CUBE program is built on the following four pillars: 1) training in introductory biostatistics; 2) training in R programming; 3) professional development; and 4) a collaborative research project addressing research questions in addiction and mental health.

The four pillars revolve around the core competencies of an effective collaborative biostatistician, including technical and analytical skills, strong communication, and problem solving and framing, in combination with professional development (Samsa, 2018). The professional development pillar includes mentorship from, and exposure to, various graduate researchers and scientific programs, as well as hands-on direction in graduate school applications. Details on the components of each CUBE pillar are provided in ***Figure 1***.

Program Sites and Delivery The CUBE program is delivered at two sites: VT CBHDS (lead site), and UVA PHS (partner site). Biostatistics and clinical mentors (i.e., content experts) are staffed at each site to facilitate one-on-one discussions and advising for the collaborative research projects. The clinical mentors located at each site provide the research questions and data for these projects. Participants at both sites are given the same training in introductory biostatistics and R programming, where biostatisticians at the lead site deliver the content in person to their participants and offer this remotely to those at the partner site. All professional development sessions are done jointly and are held either in person (lead site only) or remotely to allow more flexibility for invited speakers.

CUBE Mentors Effective mentorship and advising is critical to engaging, recruiting, retaining, and training students in STEM (Diggs-Andrews et al., 2021). As such, CUBE participants are mentored in each program pillar by faculty members with expertise in biostatistics at VT and UVA, along with individuals holding positions at other academic institutions, government entities, and the pharmaceutical industry from across the US who are committed to diversity, equity, and inclusion (DEI) efforts. The CUBE program also incorporates leading experts in addiction and mental health research for the Collaborative Project pillar. The clinical mentors are specific to the program

Figure 2. CUBE Collaborative Research Project Titles in Summer 2023

site (VT CBHDS, UVA PHS), are well-published and highly respected researchers in the biomedical workforce, and provide personalized guidance, support, and feedback throughout the program.

Participants are assigned to their clinical mentors and collaborative projects based on their rankings of project titles and descriptions. The number of students assigned to each project may vary from year to year, depending on the number of available projects and participating students. In summer 2023, students worked independently as there were five identified collaborative projects and five students (see ***Figure 2*** with project titles in summer 2023). In future years, with a larger student cohort, groups will be assigned to work on each project.

Participant Eligibility Criteria and Selection Undergraduate students from underrepresented or minoritized backgrounds in their second and third years of study are eligible to apply for the CUBE program, with preference given to those majoring in STEM fields, as well as those who are considering a career in data science, or a related field. All prospective students interested in applying to the CUBE program go through a formal application process, including an online application in REDCap (Harris et al., 2019; Harris et al., 2009) to provide demographic information, a copy of their unofficial university transcript, current resume, two references from faculty members at their home institution, and a personal statement describing their interest in CUBE and how the program may contribute to their future education and career goals.

Applications are reviewed by a selection committee, where eligible students are ranked on a five-point Likert-type scale on each criterion (year of study, undergraduate major/program, coding experience, expressed interest in data science/statistics, undergraduate GPA) based on their submitted materials. Students invited to participate in the program are given two weeks to make their final decisions.

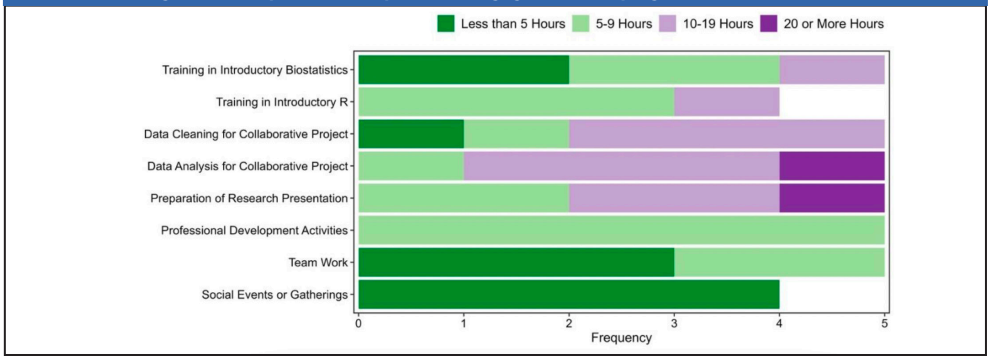
Participant Recruitment and Advertising Several strategies are taken to recruit participants and advertise the program each summer, including the use of social media platforms, online recruiting platforms specific for undergraduate students (e.g., Handshake), professional association job boards (e.g., American Statistical Association), word-of-mouth experiences from past participants, as well as automated email advertisements to heads of statistics, mathematics, and other STEM-related departments from universities across the US.

Participant Costs To support and provide an immersive research and professional development training program experience, participants receive a stipend of $4,800, along with housing in a shared, furnished apartment with a full kitchen at their respective site (~$2,400). Participants are expected to cover their own meals and travel costs. The program additionally pays for a one-year membership to both the American Statistical Association and the Society for Advancement of Chicanos/Hispanics and Native Americans in Science (SACNAS) (~$60 per student). Other participant costs not covered by federal funding include social events attended by both students and mentors, such as ice cream outings, hiking trips, movie nights, and group lunches and dinners (~$1,000).

Participant Daily Experience During the eight-week program, CUBE participants spend the morning receiving instruction in biostatistics and R programming. The content is delivered in a flipped format via recorded videos. Specifically, educational videos are viewed asynchronously at the students' leisure, with interactive in-person question and answer discussions held during the morning hours. Additionally, participants complete weekly homework assignments to reinforce the concepts, and timely feedback is provided to enhance learning. Participants attend group lunch-and-learn sessions focused on professional development topics, informal networking with research professionals or social lunches with other undergraduate and graduate students in STEM participating in other summer programs at each site. Students also engage in a two-day workshop in partnership with Mount Sinai's Center for Scientific Diversity (CSD), covering critical topics in DEI, such as scientific identity, imposter syndrome, and sense of belonging within the scientific community. This workshop is a distinctive feature that sets CUBE apart from other undergraduate summer programs in biostatistics. These sessions aim to equip CUBE participants with the necessary tools to navigate and thrive in diverse academic and professional environments.

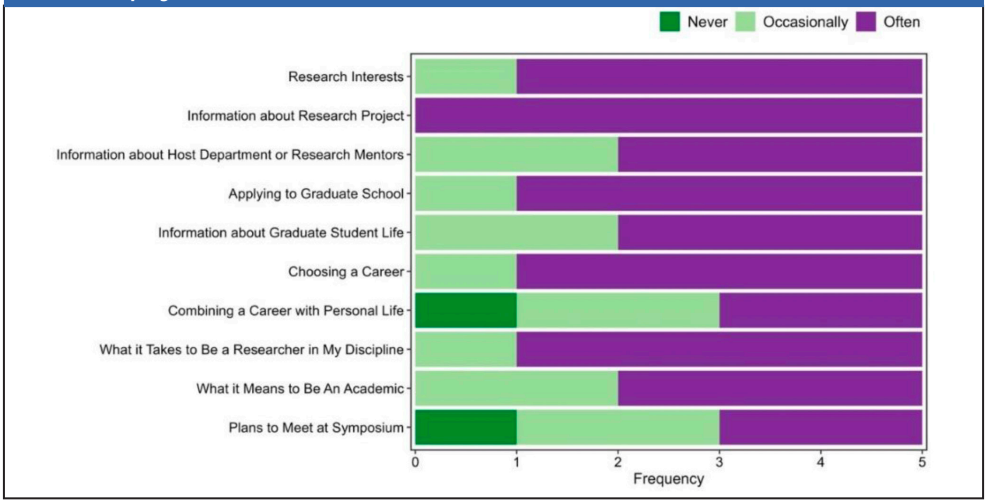
Afternoons include tours of behavioral and addiction research labs to give participants exposure to the type of research being carried out and the researchers with whom they will collaborate. These tours help broaden CUBE participants' understanding of research activities, such as identifying challenges with data collection methods, clarifying research questions, building relationships, and enhancing methodological knowledge. As the program goes on, participants gradually transition from biostatistics learning to application as they progress through their collaborative research projects. Participant contributions to the project include data cleaning, creating a statistical analysis plan and "Table 1", and performing analyses using appropriate statistical methodology (e.g., linear or logistic regression, survival analysis, mixed effects modeling, mediation modeling). Biostatistics mentors may cover additional statistical topics when more complex methods are used that go beyond their instruction in the Introductory Biostatistics pillar. At the end of the 8-week program,

Figure 3. Bar plot of hours per week engaged in CUBE program activities* (N = 5)



*Question: During your CUBE program experience, how many hours each week did you typically spend on the following activities (Response Options: Less than 5 hours; 5-9 hours; 10-19 hours; 20 or more hours)

Figure 4. Bar plot of topics discussed during Professional Development Pillar* delivered during the CUBE program (N = 5)



*Question: During the Professional Development sessions delivered during the CUBE program, how often did you discuss the following topics with others? (Response Options: Often; Occasionally; Never)

Table 1. Demographic characteristics of CUBE participants (N = 5)

Variable		Median (IQR)/n (%)
Age (years), Median (IQR)		21.08 (20.83, 21.33)
Gender, n (%)	Female	4 (80.0%)
	Male	1 (20.0%)
Race/Ethnicity, n (%)	Asian	1 (20.0%)
	Black/African-American	1 (20.0%)
	Hispanic/Latinx	1 (20.0%)
	White	2 (40.0%)
Underrepresented Status*, n (%)	Disability	1 (20.0%)
	Disadvantaged Background	3 (60.0%)
	Gender	4 (80.0%)
	LGBTQIA+	2 (40.0%)
	Racial/Ethnic Minority	3 (60.0%)
Undergraduate Year of Study, n (%)	Rising Third Year	1 (20.0%)
	Rising Fourth Year	4 (80.0%)
Expected Semester & Year of Graduation, n (%)	Spring 2024	4 (80.0%)
	Spring 2025	1 (20.0%)
Undergraduate Major, n (%)	Applied Statistics/Statistics	2 (40.0%)
	Biology	1 (20.0%)
	Economics	1 (20.0%)
	Global & Public Health Sciences	1 (20.0%)
CUBE Program Site, n (%)	UVA PHS	2 (40.0%)
	VT CBHDS	3 (60.0%)

IQR: Interquartile Range; LGBTQIA+: lesbian, gay, bisexual, transgender, queer or questioning, intersex, asexual, and more; UVA PHS: University of Virginia Department of Public Health Sciences; VT CBHDS: Virginia Tech Center for Biostatistics and Health Data Science; *Participants were able to select more than one category related to underrepresented status.

Table 2. Descriptive statistics of activities to prepare for the CUBE program (N = 5)

Question: Please rate your program with regard to your following activities to prepare for your summer experience.

	Excellent	Good	Fair	Poor
Prompt notification from program organizers regarding your acceptance into the program	2 (40.0%)	3 (60.0%)	0 (0.0%)	0 (0.0%)
Timely information on program logistics (e.g., key contact personnel, schedule, housing, transportation, local community, etc.)	4 (80.0%)	0 (0.0%)	1 (20.0%)	0 (0.0%)
Timely communication with mentor information (e.g., mentor contact information, department, etc.)	4 (80.0%)	1 (20.0%)	0 (0.0%)	0 (0.0%)
On-site orientation to overall summer program (e.g., program expectations, final deliverables, campus facilities and policies, etc.)	2 (40.0%)	3 (60.0%)	0 (0.0%)	0 (0.0%)

students present their research products, including a written report, poster, and 15- to 30-minute oral presentation attended by the biostatistics and clinical mentors, as well as family and friends of the participants (~10-15 attendees).

Program Evaluation Metrics To evaluate the success of the CUBE program, quantitative and qualitative evaluations were administered to participants prior to, during, and at the conclusion of the eight-week program. Specifically, a survey developed by The Leadership Alliance (Ghee et al., 2016) measuring pre- versus post-program perspectives on knowledge development, future plans, impact of mentoring and support was modified for the CUBE program. In addition, weekly reflections, both quantitative and qualitative, are collected from each participant and used to continually refine the CUBE program for maximum success. Lastly, qualitative, structured exit interviews are completed by all participants at the close of the program asking them to provide feedback on their experience in each of the four pillars of the CUBE program (e.g., pace, content, topics covered, final deliverables), how CUBE may have impacted their future goals (e.g., new skills/techniques learned and their usefulness), as well as their overall experience in the program (e.g., relationships with mentors, most/least rewarding aspect of program, suggestions for improvement). These sessions are recorded and transcribed for informal descriptive analyses that are used to improve future program delivery.

Statistical Analysis Demographic information collected from CUBE participants was summarized using descriptive statistics. Participant responses to quantitative surveys were described and visualized with bar plots, recognizing our small sample size (N = 5). All statistical analyses were performed in SAS V9.4 (SAS Institute Inc., Cary, NC), with visualizations generated using the ggplot2 package in R (v4.3.1) (Wickham, 2016).

Program Results and Outcomes

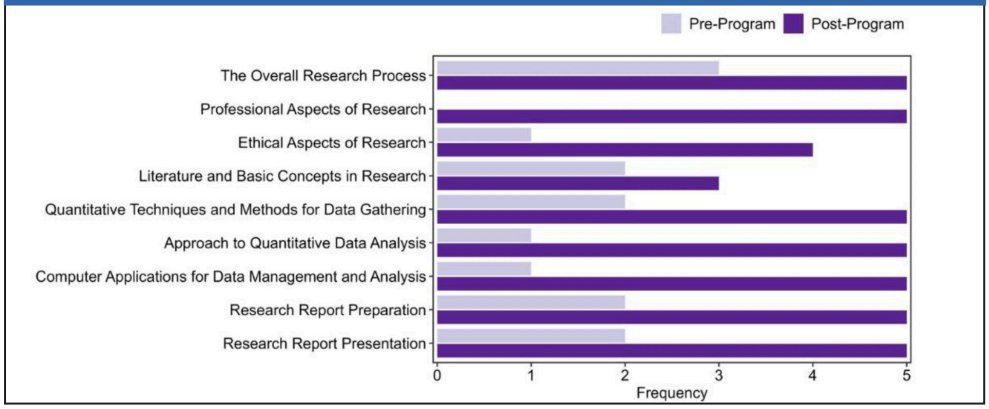
Participant Demographics CUBE participant demographics are presented in ***Table 1***. Five undergraduate students participated in the CUBE program in summer of 2023, with three hosted at VT CBHDS and two at UVA PHS. Participants reported a median age of 21 years at the start of the program. The majority of CUBE participants were female students (80%) entering their fourth year (80%) of undergraduate study. All five participants identified as belonging to at least one underrepresented group per the NIH's definition (Notice of NIH's Interest in Diversity, *2019)*, including: racial/ethnic minority (60%), gender (80%), disadvantaged background (60%), LGBTQIA+ (40%), and disability (20%).

Pre-Program Logistics and Communication ***Table 2*** summarizes participant onboarding activities performed prior to start of CUBE program, where participants were asked to rate the delivery of pre-program logistics and communication. All participants felt that notification of their acceptance into the program was prompt and that they received information on program logistics and mentors in a timely manner and were overall satisfied with the on-site program orientation (***Table 2***).

Program Activities and Curriculum CUBE participants were asked about the amount of time spent on program activities, including the four CUBE pillars (i.e., training in introductory biostatistics, R programming, professional development, data cleaning and analysis for their collaborative projects, preparation of poster and oral presentations), as well as teamwork and social events or gatherings. As demonstrated in ***Figure 3***, participants spent most of their time in the first five weeks engaged in their introductory biostatistics and R programming; in the final two weeks of the program, participants reported being mostly engaged in the data cleaning and analysis for the Collaborative Research Project pillar, along with the preparation of their poster and oral presentations. Participants spent the least amount of time each week on social events or gatherings (***Figure 3***). Regarding the Professional Development pillar, CUBE participants were asked to rate how often specific topics in research were discussed during these sessions (***Figure 4***). The top three topics rated as being discussed most often included: information about the collaborative research project, research interests, and career paths (***Figure 4***).

Pre- versus Post-Program Research-Related Knowledge Participants were asked to rate their knowledge of research-related topics at pre- and post-program. Frequencies of responses rated as either 'intermediate' or 'high' at pre- and post-program are provided in Figure 5. CUBE participants reported increased knowledge in various areas, including the overall research process (pre: 3/5 vs. post: 5/5), ethical (pre: 1/5 vs. post: 4/5) and professional (pre: 0/5 vs. post: 5/5) aspects of research, literature and basic concepts in research (pre: 2/5 vs. post: 3/5),

Figure 5. Frequency of research-related topics rated as intermediate or high at pre- and post-program *# (N = 5)



*Question: Please indicate your level of knowledge about each of the following topics (Response Options: None; Beginner; Intermediate; High).
#Note that none of the 5 participants rated their knowledge in Professional Aspects of Research as 'Intermediate' or 'High' at pre-program.

data analysis (pre: 1/5 vs. post: 5/5), data management (pre: 1/5 vs. post: 5/5), report preparation (pre: 2/5 vs. post: 5/5), and presenting research (2/5 vs. post: 5/5) (***Figure 5***).

Experiences and Interactions with Mentors At the start of the program, CUBE participants were asked about the importance of various experiences and interactions with their mentors (***Table 3***). The most prominent categories identified as "very important" for their summer research experience included: availability to discuss and respond to questions, offering guidance and advice, showing interest in the participant's research, making participants feel like an integral part of the research project, introducing participants to a variety of research methodological techniques, and providing them with constructive feedback on their academic career development (***Table 3***). CUBE participants were asked to rate their actual experiences and interactions with their mentors during the CUBE program from 'excellent' to 'poor'. As shown in ***Figure 6***, participants felt that their overall experiences and interactions with their CUBE mentors were excellent in the areas that were rated as most important to them at the start of the program. Additionally, all participants reported it was very likely they would keep in touch with their CUBE mentors in the next academic year (data not shown).

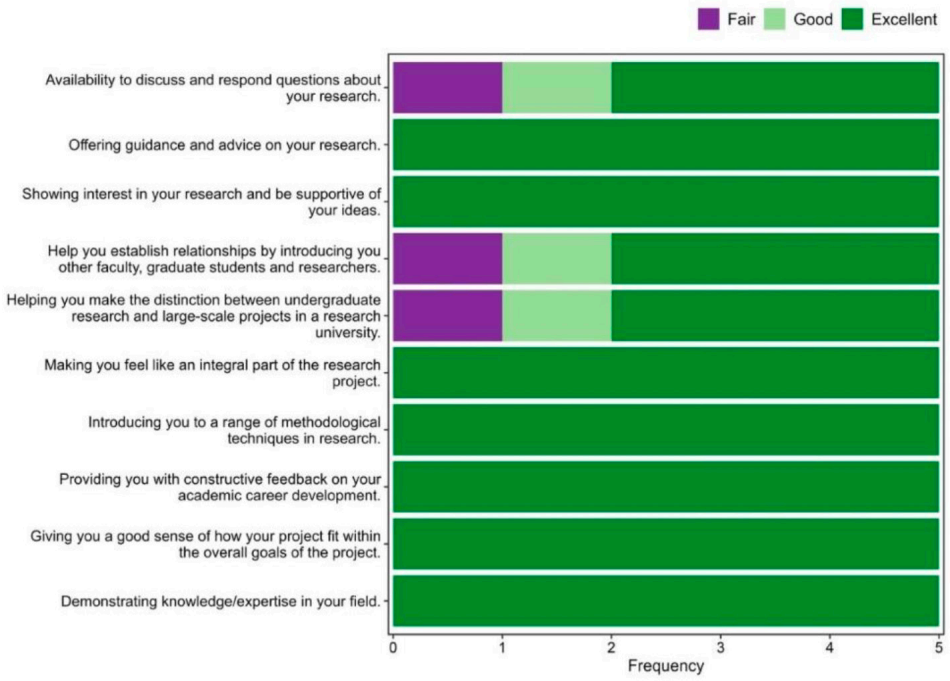
Impact and Utility of CUBE on Research Experience The relevance and importance of the participants' research experiences are summarized in ***Table 4***. All participants felt that their CUBE experience will help them with their future coursework, senior thesis/research projects and/or papers, and will enhance their overall research experience. CUBE participants were also asked about the impact related to the methodological techniques and skills introduced during the CUBE program (***Table 5***). Overall, participants saw the most impact in their critical thinking skills (4/5), ability to learn on their own (5/5), writing skills (5/5), and analytical and critical reading skills (5/5); they also expressed that they learned new methodological (5/5) and adaptive (5/5) skills and vocabulary (5/5) by the end of the program (***Table 5***). Regarding utility of skills gained during their summer experience, CUBE participants felt the program was rated as useful in these categories: increasing their readiness to engage in research, improving their overall confidence as a researcher, becoming part of a learning community, and

Table 3. Descriptive statistics of expected experiences and interactions with mentors rated prior to the start of the CUBE program (N = 5)

Question: In thinking about the research experience you will be engaged in this summer, please consider the following ways your CUBE mentor(s) may interact with you and indicate the importance of each to you.

	Very Important	Somewhat Important	Not very/Not at all Important
Availability to discuss and respond questions about your research.	5 (100.0%)	0 (0.0%)	0 (0.0%)
Offering guidance and advice on your research.	5 (100.0%)	0 (0.0%)	0 (0.0%)
Showing interest in your research and be supportive of your ideas.	5 (100.0%)	0 (0.0%)	0 (0.0%)
Helping you establish relationships by introducing you other faculty, graduate students, and researchers.	3 (60.0%)	2 (40.0%)	0 (0.0%)
Helping you make the distinction between undergraduate research and large-scale projects in a research university.	2 (40.0%)	3 (60.0%)	0 (0.0%)
Making you feel like an integral part of the research project.	5 (100.0%)	0 (0.0%)	0 (0.0%)
Introducing you to a range of methodological techniques in research.	5 (100.0%)	0 (0.0%)	0 (0.0%)
Providing you with constructive feedback on your academic career development.	5 (100.0%)	0 (0.0%)	0 (0.0%)

Figure 6. Bar plot of actual experiences and interactions with mentors during the CUBE program* (N = 5)



*Question: Based on the contact with your CUBE mentor(s), how would you rate him/her/them on the following qualities? (Response Options: Excellent; Good; Fair; Poor)

Table 4. Descriptive statistics of relevance and importance of research, workshops, and seminar experiences during the CUBE program (N = 5)

Question: Thinking about your research experience this summer, how true would you say the following statements are for you?			
	Very True	Somewhat True	Not Very/Not at all True
My experiences this summer are relevant to my research work at my current undergraduate institution.	3 (60.0%)	2 (40.0%)	0 (0.0%)
The skills I learned this summer will help me with future coursework.	5 (100.0%)	0 (0.0%)	0 (0.0%)
The skills I learned this summer will enhance my senior thesis/research projects and/or papers.	5 (100.0%)	0 (0.0%)	0 (0.0%)
My experiences this summer have inspired me to take new courses that I had not considered before.	2 (40.0%)	3 (60.0%)	0 (0.0%)
My experiences this summer will enhance my overall undergraduate experience.	5 (100.0%)	0 (0.0%)	0 (0.0%)

Table 5. Descriptive statistics of the impact related to the methodological techniques and skills introduced during the CUBE program (N = 5)

Question: Based on your experiences, indicate a response to each of the following statements:				
	Strongly Agree	Agree	Disagree	Strongly Disagree
My critical thinking skills were enhanced through my research and dialogue with fellow students and faculty.	4 (80.0%)	0 (0.0%)	1 (20.0%)	0 (0.0%)
I increased my ability to learn on my own, pursue ideas, and find information I need.	3 (60.0%)	2 (40.0%)	0 (0.0%)	0 (0.0%)
I further developed and improved my research writing skills.	3 (60.0%)	2 (40.0%)	0 (0.0%)	0 (0.0%)
My analytical and critical reading skills were sharpened.	4 (80.0%)	1 (20.0%)	0 (0.0%)	0 (0.0%)
I am better equipped to put ideas together, see relationships, similarities, and differences between ideas.	4 (80.0%)	1 (20.0%)	0 (0.0%)	0 (0.0%)
I learned new techniques that enhanced my methodological skills.	5 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
I have developed the necessary vocabulary of my discipline to engage in critical dialogue.	5 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
I learned how to adapt to change (e.g., new technologies, different jobs, or personal circumstances, etc.).	5 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

understanding what it takes to have a career in research (***Figure 7***). When asked to rate the utility of the training and information provided to them during the CUBE program, participants identified the instruction and skill building for presenting and publishing research, along with professional guidance on resume development and career development as the most useful (***Figure 8***).

Participant Graduate School and Future Career Plans. Participants reported on their graduate school and future career plans at pre- and post-program. At pre-program, few CUBE participants had plans to attend graduate school to pursue either a Master's (4/5) or PhD (3/5) (***Table 6***). At the end of the program, most participants felt more confident (responses of strongly agree/agree) in planning for their graduate school education (5/5) and a career in research (4/5) and had a more clarified career path (4/5). Overall, CUBE participants felt a stronger commitment towards pursuing a research career (4/5), with no such plans prior to the start of the program (5/5) (***Table 6***). Participants were asked which graduate entry exams they were planning to take, with most seeking to take the Graduate Record Exam (GRE) (3/5) and one participant interested in taking the Medical College Admission Test (MCAT) to pursue a medical degree (data not shown).

Figure 9 depicts frequencies of responses rated as either 'strongly agree' or 'agree' to questions about knowledge and experiences related to graduate school and future careers pre- versus post-program. CUBE participants showed increases in their understanding of the graduate school application process (pre: 1/5 vs. post: 3/5), graduate school life (pre: 1/5 vs. post: 3/5), careers in statistics or biostatistics (pre: 3/5 vs. post: 5/5). Decreases in plans to pursue academic careers and non-academic research careers were also observed (***Figure 9***).

CUBE Participant Outcomes. Of our five CUBE participants in summer of 2023, none came into the program with an interest in biostatistics or pursuing a career in research despite expressing interest in applying to graduate school at the start of the program. Through the exit interview sessions, all but one of the participants expressed that they are now on a trajectory that has been directly impacted by CUBE, where four participants have plans to attend graduate school in biostatistics or research-related fields. One student said of the field of collaborative biostatistics, "It feels like a hidden gem of a career", and another stated, "I do like it. . .it's definitely what I've been wanting in a career path". The single participant whose trajectory was not impacted by CUBE is preparing for medical school; the participant came into the program with the same plans.

Conclusions

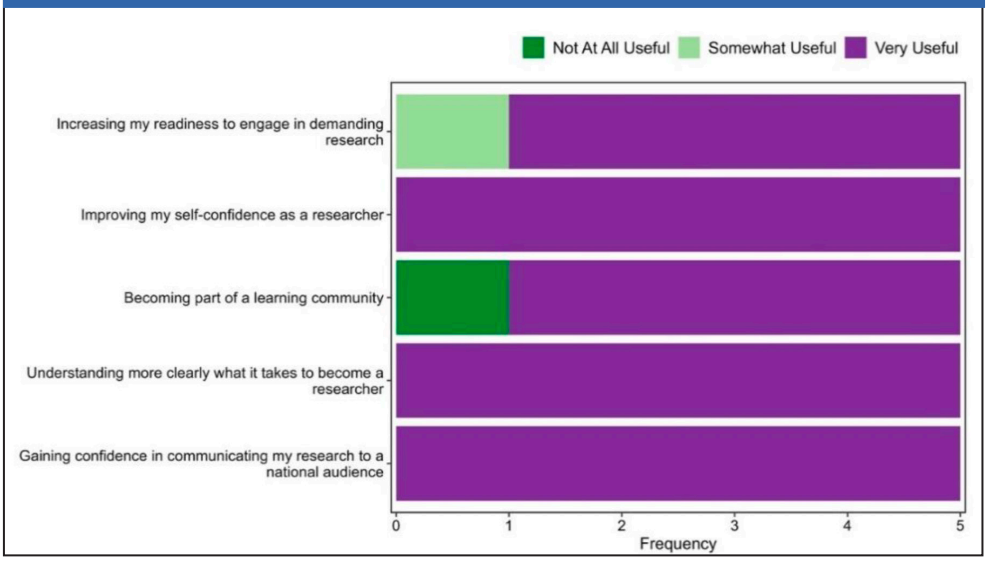
Results from participant survey and qualitative exit interviews in summer 2023 demonstrated an overall positive impact on the students, specifically with respect

to enhancing their knowledge in statistics, programming, and research, as well as their future aspirations. Notably, four of the five participants had plans to attend graduate school in biostatistics or research-related fields by the end of the program despite having no plans at the start of the program. The CUBE program played a critical role in fostering confidence in participants and thus positively influencing their future in STEM. This affirms the CUBE program equipped aspiring researchers with the skills and confidence needed for successful endeavors in the fields of biostatistics, research, and translational science, with an emphasis in the areas of addiction and mental health.

Challenges and Strategies The CUBE team has encountered some major challenges including: issues with securing both office space and housing, as the program has grown over the years; recruitment and enrollment; ensuring consistent mentoring of participants across a variety of clinical mentors; problems with late timing of stipend payouts, where students with financial needs and potential food insecurity are particularly negatively impacted; and struggling with the "clock-puncher" mentality, where students check-in at 9am and check-out at 5pm, viewing the CUBE opportunity as an hourly job. This mentality can limit engagement and the potential for learning, and the CUBE team is working on messaging to view the program as an opportunity and foundation for their future pathway. Some of the strategies and solutions to these challenges include: VT CBHDS requesting an administrative supplement from the NIH to support an extensive recruitment tour to increase robustness of recruitment, along with developing training materials for the clinical mentors in the mentoring of students who come from underrepresented or minoritized backgrounds; fully automating the recruitment email process, as well as the process for gathering letters of recommendation from applicants; and finally, working with Human Resources at both sites on payment schedule alternatives so that students can be paid earlier in the summer.

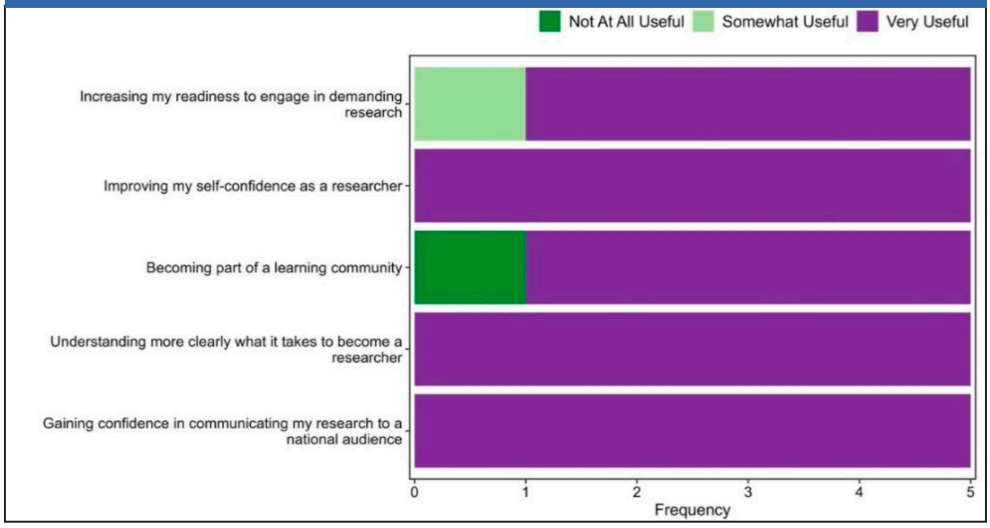
Limitations These results should be interpreted in the context of acknowledged limitations. The first relates to the small overall sample size (N = 5). Despite being funded by the NIH to support up to ten students per year at VT CBHDS, only three students were enrolled in the CUBE program for summer 2023 at this site due to the notice of award being received a few weeks prior to the start of the program. Further, only two students could be supported at UVA PHS due to funding constraints. No statistical comparisons from pre- to post-program could be made due to the limited sample size, and thus this paper focuses on describing findings from participant surveys and interviews. Additionally, survey questions from The Leadership Alliance (Ghee et al., 2016) were not entirely specific to activities conducted in the program delivered in summer 2023. Moving forward, surveys specific to the CUBE curriculum will be created and validated to assess the program more accurately.

Figure 7. Bar plot of student-rated utility of the CUBE program experience* (N = 5)



*Question: Please indicate how useful your CUBE program experience was in the following areas (Response Options: Very Useful; Somewhat Useful; Not at all Useful; Not applicable)

Figure 8. Utility of training and information provided to the CUBE participants* (N = 5)



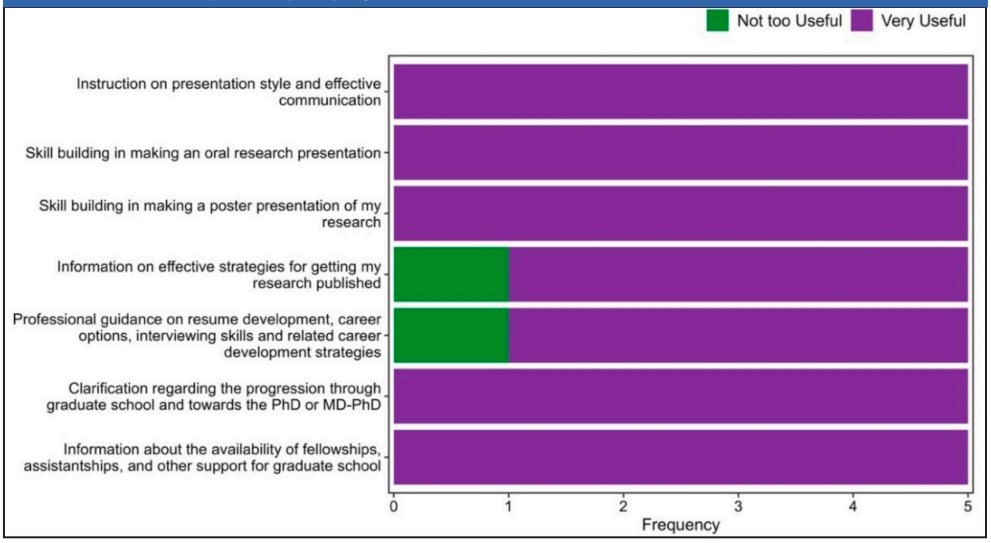
*Question: Please indicate how useful your CUBE program experience was in the following areas (Response Options: Very Useful; Somewhat Useful; Not at all Useful; Not applicable)

While engaging students in the third year of their undergraduate program is optimal for creating pathways to graduate programs or the workforce, it may also be viewed as a limitation of the CUBE program. Admission to graduate programs in biostatistics may require students to take more advanced mathematics courses, such as multivariate calculus and linear algebra. If students have not initiated this course sequence before the end of their third year of undergraduate study, they will not have time to complete the prerequisite coursework to begin graduate school in biostatistics immediately after graduation. Of the five students participating in summer 2023, one had already taken all of these courses, and the remaining four had taken a portion of these classes and could complete the required courses before graduation, suggesting that this may not be a major limitation of the program.

Future Work. The CUBE program delivered in summer 2023 laid the foundation for programs of its kind to be developed at other institutions. As such, future work will include not only expanding the program at both sites, but also developing partnerships with leaders in the field at other universities across the US to establish similar programs seeking to provide a pipeline for equity and diversity in the practice of collaborative biostatistics. Additional long-term program evaluation metrics are expected be captured in CUBE alums, including degrees earned, number of publications, as well as the number of conference abstracts and grants submitted and accepted. The authors plan to summarize survey responses and interviews from additional CUBE participants and their long-term career trajectories after the program in a subsequent paper. Future study and a larger sample would allow leaders in the

field to examine the impact of summer program experiences such as CUBE on bringing diversity and awareness to collaborative biostatistics and health data science.

Figure 9. Frequency of increased knowledge (strongly agree/agree) related to graduate school and future careers at pre- and post-program* (N = 5)



*Question: Indicate your level of agreement with each of the following statements about graduate school (Response Options: Strongly agree; Agree; Disagree; Strongly disagree).

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?). *The American Statistician, 65*(2), 71-79. http://www.jstor.org/stable/23020498

Benn, E., Tabb, L. P., Exum, P., Moore, R. H., Morales, K. H., Simpson, F., Lawrence, S. A., & Bellamy, S. L. (2020). Creating and sustaining effective pipeline initiatives to increase diversity in biostatistics: the ENAR Fostering Diversity in Biostatistics Workshop. *J Stat Educ, 28*(3), 295-303. https://doi.org/10.1080/10691898.2020.1820409

Black Students in STEM and Health Graduate Programs Increase But a Large Racial Gap Remains. (2021). https://jbhe.com/2021/04/black-students-in-stem-and-health-graduate- programs-increase-but-a-large-racial-gap-remains/

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., Pomann, G.-M., Samsa, G., . . . Wruck, L. (2021). Methods for training collaborative biostatisticians. *Journal of Clinical and Translational Science, 5*(1), e26, Article e26. https://doi.org/10.1017/cts.2020.518

Caldwell, C. H., Thomas, D., Hoelscher, H., Williams, H., Mason, Z., Valerio-Shewmaker, M. A., & Panapasa, S. V. (2021). Tailoring Recruitment and Outreach Strategies for Underrepresented Students in Public Health Pipeline Programs. *Pedagogy in Health Promotion, 7*(1_suppl), 36S-43S. https://doi.org/10.1177/23733799211047517

Diggs-Andrews, K. A., Mayer, D. C. G., & Riggs, B. (2021). Introduction to effective mentorship for early-career research scientists. *BMC Proceedings, 15*(2), 7. https://doi.org/10.1186/s12919-021-00212-9

Estrada, M., Hernandez, P. R., & Schultz, P. W. (2018). A Longitudinal Study of How Quality Mentorship and Research Experience Integrate Underrepresented Minorities into STEM Careers. CBE—Life Sciences Education, 17(1), ar9. https://doi.org/10.1187/cbe.17-04- 0066

Ghee, M., Keels, M., Collins, D., Neal-Spence, C., & Baker, E. (2016). Fine-Tuning Summer Research Programs to Promote Underrepresented Students' Persistence in the STEM Pathway. CBE Life Sci Educ, 15(3). https://doi.org/10.1187/cbe.16-01-0046

Golbeck, A. L. B., T.H., Rose, C.A. (2020). *Report on the 2017–2018 New Doctorate Recipients* (Mathematical and Statistical Sciences Annual Survey, Issue. A. M. Society. http://www.ams.org/profession/data/annual-survey/2018Survey-NewDoctorates- Report.pdf

Table 6. Descriptive statistics for topics related to graduate school and future career plans at pre- and post-program (N = 5)

	Program Period	Strongly Agree	Agree	Disagree	Strongly Disagree
I plan to attend graduate school to pursue a Master's degree.	Pre-Program*	1 (20.0%)	3 (60.0%)	1 (20.0%)	0 (0.0%)
I plan to attend graduate school to pursue a PhD	Pre-Program*	1 (20.0%)	2 (40.0%)	1 (20.0%)	1 (20.0%)
I plan to attend graduate school to pursue a MD-PhD degree.	Pre-Program*	1 (20.0%)	0 (0.0%)	2 (40.0%)	2 (40.0%)
Overall, I am more confident about planning for my graduate school education.	Post-Program**	4 (80.0%)	1 (20.0%)	0 (0.0%)	0 (0.0%)
Overall, I am more confident about planning for a career in research.	Post-Program**	1 (20.0%)	3 (60.0%)	1 (20.0%)	0 (0.0%)
I have clarified my career path.	Post-Program**	2 (40.0%)	2 (40.0%)	1 (20.0%)	0 (0.0%)
I had a plan to pursue a research career before I began the CUBE program, and the plan has not changed.	Post-Program**	0 (0.0%)	0 (0.0%)	5 (100.0%)	0 (0.0%)
My commitment to pursue a research career is stronger.	Post-Program**	1 (20.0%)	3 (60.0%)	1 (20.0%)	0 (0.0%)
		Pursue Master's Degree	Pursue PhD	Pursue MD-PhD	
Which best describes your plans for graduate school within 2-3 years after completing your undergraduate degree:	Post-Program	3 (60%)	1 (20%)	1 (20%)	--

*Question (Pre-Program): Based on your knowledge and experiences to date, indicate your response to each of the following statements about graduate school
**Question (Post-Program): Based on your experiences and knowledge gained through the CUBE program, indicate your level of agreement with each of the following statements about graduate school

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., & Duda, S. N. (2019). The REDCap consortium: Building an international community of software platform partners. *J Biomed Inform, 95*, 103208. https://doi.org/10.1016/j.jbi.2019.103208

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform, 42*(2), 377-381. https://doi.org/10.1016/j.jbi.2008.08.010

King, C., Englander, H., Priest, K. C., Korthuis, P. T., & McPherson, S. (2020). Addressing Missing Data in Substance Use Research: A Review and Data Justice-based Approach. *J Addict Med, 14*(6), 454-456. https://doi.org/10.1097/adm.0000000000000644

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a Unifying Social Cognitive Theory of Career and Academic Interest, Choice, and Performance. *Journal of Vocational Behavior, 45*(1), 79-122. https://doi.org/https://doi.org/10.1006/jvbe.1994.1027

MacKinnon, D. P., & Lockwood, C. M. (2003). Advances in statistical methods for substance abuse prevention research. *Prev Sci, 4*(3), 155-171. https://doi.org/10.1023/a:1024649822872

Mathematicians and Statisticians: Occupational Outlook Handbook: US Bureau of Labor Statistics. Bureau of Labor Statistics. https://www.bls.gov/ooh/math/mathematicians- and-statisticians.htm?external_link=true

Notice of NIH's Interest in Diversity. (2019). National Institutes of Health. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-20-031.html

Pomann, G. M., Boulware, L. E., Chan, 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). Experiential Learning Methods for Biostatistics Students: A Model for Embedding Student Interns in Academic Health Centers. *Stat, 11*(1). https://doi.org/10.1002/sta4.506

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). https://doi.org/https://doi.org/10.5430/jct.v7n1p20

Slade, E., Brearley, A. M., Coles, A., Hayat, M. J., Kulkarni, P. M., Nowacki, A. S., Oster, R. A., Posner, M. A., Samsa, G., Spratt, H., Troy, J., & Pomann, G. M. (2023). Essential team science skills for biostatisticians on collaborative research teams. J Clin Transl Sci, 7(1), e243. https://doi.org/10.1017/cts.2023.676

Stewart, J., Henderson, R., Michaluk, L., Deshler, J., Fuller, E., & Rambo-Hernandez, K. (2020). Using the Social Cognitive Theory Framework to Chart Gender Differences in the Developmental Trajectory of STEM Self-Efficacy in Science and Engineering Students. Journal of Science Education and Technology, 29. https://doi.org/10.1007/s10956-020-09853-5

Survey of Earned Doctorates (SED). (2022). National Science Foundation. https://www.nsf.gov/statistics/srvydoctorates/

Wickham, H. (2016). Ggplot2: Elegant graphics for data analysis (2 ed.). Springer International Publishing.

Acknowledgements

The authors would like to thank and acknowledge our CUBE Collaborative and Professional Development mentors and partners to date, including Drs. Charlotte Baker, Warren Bickel, Emma Benn, Anne Brown, Brooks Casas, Paulette Ceesay, Pearl Chiu, Jody Ciolino, Maricela Cruz, Mario Davidson, Tina Davidson, Alexandra DiFeliceantonio, Christopher Grubb, Marieke Jones, Ji-Hyun Lee, Matilde Kam, Tracy Layne, Virginia LeBaron, Nihal Mohamed, Emmanuel Nartey, Oluwatosin Ogunmayowa, Gina-Maria Pomann, Dionne Price, Megan Price, Julia Scialla, Jamilia Sly, Jeff Stein, Marina Walther-Antonio, and Li Wang, as well as Alicia Arneson, Tanner Barbour, Matthew Fritts, Alexandre Goebel, Jennifer Heinold, Ryann Kolb, Muyao (Jenny) Lin, Christianna Lindsay, Amanda MacDonald, Tracy Truong, and Kayla Williams. We also acknowledge Benjamin Zeitlin and Wenyan Ji for assisting with the data analysis presented in this publication. Finally, we would like to express our gratitude to our CUBE students for their participation and interest in this program.

Funding

The delivery of the Collaborative Undergraduate Biostatistics Experience **(CUBE)** program and results reported in this publication were supported by the National Institutes of Health's (NIH) National Institute on Drug Abuse (NIDA) and National Institute on Alcohol Abuse and Alcoholism (NIAAA) (Award Number: R25DA058482), the National Center For Advancing Translational Sciences of the NIH (Award Number: UL1TR003015), as well as supplemental funding from Virginia Tech's College of Science, Fralin Life Sciences Institute (FLSI), the Institute for Society, Culture, and Environment (ISCE), the University of Virginia's Department of Public Health Sciences, AbbVie, and Merck & Co., Inc.

Ms. Alicia J. Lozano is the Assistant Director of the Center for Biostatistics and Health Data Science (CBHDS), and a Senior Research Associate in the Department of Statistics at Virginia Tech. She has extensive collaborative experience supporting numerous clinical or biomedical research projects and has mentored various undergraduate and graduate students in collaborative biostatistics. Ms. Lozano serves as both a biostatistics mentor and confidant to students participating in the NIH-funded Collaborative Undergraduate Biostatistics Experience (CUBE) program.

Dr. Monica L. Ahrens is a Research Scientist at the Center for Biostatistics and Health data Science at Virginia Tech. She has worked as a biostatistician on a variety of content areas including psychology, neuroscience, and physiology. Dr. Ahrens has served as a mentor for medical, PhD, and undergraduate students; she has worked with more than twenty medical students through required research projects in a variety of clinical research areas. Dr. Ahrens serves as both a biostatistics and R programming mentor in the NIH-funded Collaborative Undergraduate Biostatistics Experience (CUBE) program.

Ms. Genevieve R. Lyons is a Senior Biostatistician in the Department of Public Health Sciences at the University of Virginia. She has extensive experience as a collaborator on a range of research projects in clinical, public health, and nursing research, and she has developed expertise in methods and techniques for analyzing observational and secondary data. Genevieve has mentored undergraduate, graduate, and medical students as well as research fellows. She teaches statistical programming in UVA's MPH program, and she serves as a biostatistics mentor and confidant to students participating in the NIH-funded Collaborative Undergraduate Biostatistics Experience (CUBE) program.

Dr. Jennie Z. Ma is a Professor of Biostatistics in the Department of Public Health Sciences at the University of Virginia. She is a co-director of the Biostatistics, Epidemiology, and Research Design (BERD) within the Research Methods Core for the NIH-funded integrated Translational Health Research Institute of Virginia (iTHRIV) CTSA. Dr. Ma has collaborated extensively with clinical investigators for 30+ years, adeptly integrating innovative statistical methodology with subject-specific expertise in outcomes research and clinical trials. Dr. Ma serves as a biostatistics mentor for medical students, residents, fellows, junior faculty, and graduate/undergraduate students, including those in the NIH-funded Collaborative Undergraduate Biostatistics Experience (CUBE) program.

Dr. Sarah J. Ratcliffe, is Professor and Director of Biostatistics at the University of Virginia. She currently serves as the Senior Vice Chair for Research in the Department of Public Health Sciences, Director of the Research Methods Core within the NIH-funded integrated Translational Health Research Institute of Virginia (iTHRIV) CTSA, and Director of the UVA-Hub for the Collaborative Undergraduate Biostatistics Experience (CUBE) program. She has over 25 years of experience as a biostatistician and mentor to various faculty, undergraduate and graduate students.

Dr. Alexandra L. Hanlon is the Director of the Center for Biostatistics and Health Data Science (CBHDS) and a Professor of Practice in the Department of Statistics at Virginia Tech. She is currently the Co-Director of the Research Methods Core within the NIH-funded integrated Translational Health Research Institute of Virginia (iTHRIV) CTSA. Dr. Hanlon serves as the founder and Program Director of the NIH-funded Collaborative Undergraduate Biostatistics Experience (CUBE) program. She has over 30 years of experience as a collaborative biostatistician and is heavily engaged in diversity, equity, and inclusion efforts through her engagement with national special interest groups.