EDUCATION
SPECIAL COMMUNICATION

The ASIBS Short Course: A unique strategy for increasing statistical competency of junior investigators in academic medicine

Emma K. T. Benn[1]*, Chengcheng Tu[1], Ann-Gel S. Palermo[2], Luisa N. Borrell[3],
Michaela Kiernan[4], Mary Sandre[1] and Emilia Bagiella[1]

[1] Center for Biostatistics and Department of Population Health Science & Policy, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[2] Center for Multicultural and Community Affairs and Office of Diversity and Inclusion, Icahn School of Medicine at Mount Sinai, New York, NY, USA
[3] Department of Epidemiology and Biostatistics, Graduate School of Public Health and Health Policy, City University of New York, New York, NY, USA
[4] Stanford Prevention Research Center, Stanford School of Medicine, Stanford, CA, USA

As clinical researchers at academic medical institutions across the United States increasingly manage complex clinical databases and registries, they often lack the statistical expertise to utilize the data for research purposes. This statistical inadequacy prevents junior investigators from disseminating clinical findings in peer-reviewed journals and from obtaining research funding, thereby hindering their potential for promotion. Underrepresented minorities, in particular, confront unique challenges as clinical investigators stemming from a lack of methodologically rigorous research training in their graduate medical education. This creates a ripple effect for them with respect to acquiring full-time appointments, obtaining federal research grants, and promotion to leadership positions in academic medicine. To fill this major gap in the statistical training of junior faculty and fellows, the authors developed the Applied Statistical Independence in Biological Systems (ASIBS) Short Course. The overall goal of ASIBS is to provide formal applied statistical training, via a hybrid distance and in-person learning format, to junior faculty and fellows actively involved in research at US academic medical institutions, with a special emphasis on underrepresented minorities. The authors present an overview of the design and implementation of ASIBS, along with a short-term evaluation of its impact for the first cohort of ASIBS participants.

Introduction

As clinical researchers at academic medical institutions across the United States increasingly become the gatekeepers of complex clinical databases and registries, they often lack the statistical expertise needed to utilize the data for research purposes. This statistical inadequacy prevents junior investigators from disseminating clinical findings in top-tier, peer-reviewed journals and makes it very difficult for them to successfully obtain research funding, which in turn hampers the academic promotion process.

* Address for correspondence: E. K. T. Benn, DrPH., M.P.H., Center for Biostatistics and Department of Population Health Science & Policy, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
(Email: emma.benn@mountsinai.org)

Underrepresented minorities (URMs), in particular, confront unique challenges as clinical investigators. First, since URMs are less likely to have participated in research-related activities during their graduate medical education, they often have more difficulty acquiring full-time faculty appointments in academic medicine than Whites [1]. Second, URMs are less likely to be awarded National Institutes of Health (NIH) R01 grants than Whites [2]. Third, URM faculty with assistant and associate professorial ranks are 32% and 19%, respectively, less likely to be promoted than similarly ranked White faculty at US medical centers [3].

Increasing recruitment and retention of diverse leaders and investigators in academic medicine requires comprehensive, multilevel solutions. One mechanism by which academic medical institutions can better ensure the success of both URM trainees and faculty members is by providing them with frequent opportunities to improve their methodological skills.

To fill this major gap in the statistical training of junior faculty and fellows, we developed the Applied Statistical Independence in Biological Systems (ASIBS) Short Course, a 5-year educational initiative funded by the National Institute of General Medical Sciences in 2015. The overall goal of ASIBS is to provide formal applied statistical training, via 7 weeks

of distance learning and 1 week of in-person statistical computing and instruction, to junior faculty and fellows actively involved in research at US academic medical institutions, with a special emphasis on URMs.

In this paper, we present an overview of the design and implementation of ASIBS, along with our short-term evaluation of its impact from the first cohort of ASIBS participants.

Methods

Target Population

ASIBS was developed for junior faculty at the rank of Instructor or Assistant Professor and postdoctoral researchers and clinical fellows from academic medical institutions nationwide who are: (1) actively participating or will begin participating within 1 year in at least 1 bio-logical-, social-, or behavioral-related research project, and (2) US citizens or eligible non-citizens. Up to 40 participants out of all eligible applicants are accepted each year. To ensure geographic and racial/ethnic diversity, 80% of accepted applicants are expected to be from outside of ASIBS' academic home, the Icahn School of Medicine at Mount Sinai (ISMMS), and 50% should be URMs as defined by the NIH.

General Participant Recruitment Strategies

In the first year of ASIBS, we employed a 2-pronged participant recruitment strategy. Flyers were emailed to Centers for Medical Education and Clinical and Translational Science Award (CTSA) hubs nationwide. Our goal with this approach was to have partner Centers for Medical Education and CTSAs circulate and disseminate information about the ASIBS Short Course through their own internal communication channels. To increase visibility and outreach online, we established an easily accessible, informative and attractive Web site (currently accessible at www.asibs-statistics.com) that would appear in key-word triggered internet searches. The ASIBS Web site provides detailed information about the mission and core objectives of the course, acknowledgment of NIGMS funding, dates of instruction, eligibility, curriculum, instructors and staff, application instructions, the registration fee (i.e., $375) and deadlines, diversity fellowships, and lodging.

URM Participant Recruitment Strategies

To ensure that we could meet our recruitment goal of at least 50% URMs, we established the Diversity Recruitment Advisory Board (DRAB), which is composed of faculty (Drs A.S.P., L.N.B., and M.K.) with a long-standing commitment to mentorship and successful promotion of URM clinical and translational investigators. The primary tasks of the Diversity Recruitment Advisory Board members are as follows:

(1) Identify barriers to recruitment of URMs with the understanding that such barriers may differ by rank or stage in academic medical career, specialty or track, level of institutional commitment to diversity, and geographic region.

(2) Make effective recommendations that address the identified barriers to recruitment of URMs.

(3) Build on existing networks while developing new relationships with leaders of other Academic Medical Institutions, CTSAs, Centers of Excellence in Minority Medical Education, and national organizations and medical societies to expand nationwide URM recruitment outreach efforts.

(4) Develop methods to annually evaluate the effectiveness of URM recruitment strategies.

(5) Make improvements to URM recruitment efforts based on results of the annual evaluations.

ASIBS Curriculum

The ASIBS curriculum can be separated into 2 complementary components: a 7-week, online statistical theory module and a 5-day, in-person collaborative review and statistical computing instruction module hosted at the ISMMS. The specific topics of instruction and structure of the online and in-person curricula are described in detail in Tables 1 and 2, respectively.

ASIBS Online Curriculum

All 7 online lectures are formatted to provide a strong but not too mathematically advanced foundation in statistical methodology. Statistical concepts are introduced through motivating real-world examples. In total, 3-hour weekly lectures are delivered via web conferencing software, which allows for real-time interaction between instructors and participants.

In developing the online curriculum we considered an optimal sequence of concepts and topics that would provide the participants with a solid working knowledge of the most commonly used introductory statistical methodologies. In the first 2 weeks, ASIBS participants become familiar with the basic concepts of statistics, study design, and probability. We then introduce 1-sample and 2-sample inference with continuous data in week 3, correlations and simple linear regression in week 4, multiple linear regression in week 5,

Table 1. *Detailed schedule of the 7-week, online Applied Statistical Independence in Biological Systems (ASIBS) statistical theory curriculum*

Week 1: Introduction to statistics and study design
    Population, sample, variable, parameter, and statistic
    Cross-sectional, prospective, retrospective, and randomized study designs
    Measures of internal and external validity
    Measures of central tendency and dispersion
Week 2: Probability
    Joint, marginal and conditional probabilities. Rules of probability
    Random variables, the normal distribution, the t distribution and the binomial distribution
    Sampling distribution, standard error, critical values, confidence intervals
    Null and alternative hypotheses, type I and type II errors, power, sample size
Week 3: 1-Sample and 2-sample inference with continuous and categorical variables
    The z test, the Students t-test for 1 sample, 2 samples or paired data
    1 and 2 sample(s) test for proportions
Week 4: Correlation and simple linear regression
    The Pearson's correlation coefficient
    The regression line, the least square estimates of the regression coefficients, the ANOVA table, hypothesis testing on regression line
    Analysis of residuals, model diagnostics, goodness-of-fit test, outliers
Week 5: Confounding, effect modification, and multiple linear regression
    Definition of a confounder and an effect modifier
    Implication for model building
    The multiple regression model, model building, main effects, interaction terms, independent predictors
    Analysis of residuals, collinearity, goodness-of-fit and choice of the optimal model
Week 6: Categorical data analysis
    Contingency tables. Rules of probability. Illustration with sensitivity and specificity
    The odds ratio, risk ratio, risk difference, and confidence intervals
    The χ^2 test, the Fisher's exact test, McNemar's test, Cochran-Armitage test for trend
Week 7: Logistic regression
    The logit link function, features of a simple and multiple logistic model, the Wald test, interpretation of the regression parameters
    The likelihood ratio test, the c-statistic

Table 2. Detailed schedule of the Applied Statistical Independence in Biological Systems (ASIBS) Short Course 1-week, in-person statistical computing curriculum

	Day 1	Day 2	Day 3	Day 4	Day 5
8:00 AM–9:30 AM	Welcome breakfast	Group learning exercise	Group learning exercise	Group learning exercise	Group learning exercise
9:45 AM–12:15 PM	SAS Lecture 1 Introduction to the SAS environment Importing Data into SAS Creating a LIBNAME statement Merging data sets Viewing contents of data Generating new variables Renaming variables Formatting variables	SAS Lecture 2 Describing continuous data Graphically visualizing continuous data Assessing normality 1 sample t-test Paired t-test 2 sample t-test	SAS Lecture 3 Creating scatter plots Testing for linear correlations Simple linear regression Testing for confounding Testing for effect modification Multiple linear regression	SAS Lecture 4 Summarizing categorical data with frequency tables 1 Sample proportion test χ^2 test Fisher's exact test McNemar's test	SAS Lecture 5 Simple logistic regression Testing for confounding Testing for effect modification Multivariate logistic regression
12:15 PM–1:30 PM	Lunch	Lunch	Lunch	Lunch	Lunch
1:45 PM–2:45 PM	Office hours	Office hours	Office hours	Office hours	Office hours
3:00 PM–5:30 PM	SAS hands-on collaborative Exercise	SAS hands-on collaborative Exercise	SAS hands-on collaborative Exercise	SAS hands-on collaborative Exercise	SAS hands-on collaborative Exercise

1-sample and 2-sample inference with categorical data in week 6, and logistic regression in the last week.

The weekly online lectures are complemented by online critical thinking/problem solving assessments to prepare participants to apply their newly gained statistical knowledge outside of the classroom. Questions are supplemented with problems taken from the medical literature or from results of actual data. Participants are asked to use the knowledge acquired through the weekly lectures to comment on the statistical approach described in the medical literature, interpret the results, and propose alternative approaches and/or to critique the researchers' methodology.

ASIBS In-Person Curriculum

The in-person, statistical computing phase of ASIBS (Table 2) takes place over a 5-day period (Monday through Friday). This face-to-face instruction provides participants with intensive, hands-on SAS (Cary, NC, USA) instruction that complements the online curriculum. During the in-person statistical computing sessions, participants learn data management strategies and SAS programming instruction specific to the analytical topics covered in their online biostatistics lectures as well as how to interpret the statistical output. In the SAS hands-on collaborative exercises, under the guidance of the instructors, participants write their own SAS programs using publicly available data sets and work in groups to interpret their statistical output and discuss the implications of their findings in the context of the research questions of interest.

Program Evaluation

The short-term program evaluation of ASIBS consists of 3 online components: a competencies-related, a curriculum-related, and a quality-related evaluation. The competencies-related evaluation consists of a pre- and post-assessment of the ASIBS participants' confidence in applying their topic-specific biostatistics knowledge and analyzing data using SAS. Anonymous IDs link participants' pre- and post-assessment responses. In the curriculum-related evaluation, conducted after the completion of ASIBS, participants are asked to evaluate the strengths and limitations of the curriculum. For each topic covered in the online and in-person sessions, we assess, on a scale of 1 to 5, the clarity of the lectures, quality of the teaching material, difficulty of the material, satisfaction with the topics covered, and efficient use of time and resources. Finally, for the quality-related evaluation, also administered upon completion of ASIBS, participants evaluate organizational aspects of the course such as scheduling and logistics, meals and accommodations, proficiency and availability of instructors and teaching assistants, program coordination, and recruitment strategies.

The long-term evaluation of ASIBS assesses the impact that the short course has on participants' research-related activities at their academic medical institutions through a short survey at 6 months and 24 months after the completion of ASIBS. In addition, the long-term evaluation facilitates the creation of a community of ASIBS alumni that allows us to follow their progress in the academic and research environment. All evaluations are conducted via REDCap [4].

Results

In 2016, which was the first year of ASIBS, there were a total of 43 eligible applicants (Table 3). Of these, 46.5% (n = 20) were accepted into the short course, of which 80% (n = 16) were female, 45% (n = 9) were URMs, and 60% (n = 12) were fellows. Accepted and rejected applicants were similar with respect to gender and race/ethnicity (URM vs. non-URM), while the proportion of junior faculty among the rejected applicants was more than double that of the accepted cohort (87% vs. 40%). Two major reasons for rejection were as follows:

Table 3. *Descriptive characteristics of applicants and selected participants for the first year of the Applied Statistical Independence in Biological Systems (ASIBS) Short Course*

	Total applicants (n = 43)	Accepted applicants (n = 20)	Rejected applicants (n = 23)
Academic rank			
Junior Faculty	28 (65%)	8 (40%)	20 (87%)
Fellows	15 (35%)	12 (60%)	3 (13%)
Gender			
Female	29 (67%)	16 (80%)	13 (57%)
Male	14 (33%)	4 (20%)	10 (43%)
URM status			
Yes	16 (37%)	9 (45%)	7 (30%)
No	27 (63%)	11 (55%)	16 (70%)

(1) personal statements in which applicants failed to articulate how participation in ASIBS and increasing their statistical competency could positively impact their long-term success as a clinical and translational researcher, and (2) applicants who were more advanced (i.e., very successful research career, more than 5 years at the assistant professor rank, etc.) than the earlier stage population ASIBS was designed to target.

Approximately 70% (n = 14) to 75% (n = 15) of participants completed the pre-post competencies-related evaluation. Among those, we observed substantial improvements in their confidence conducting and interpreting introductory to more advanced statistical analyses and SAS output (Table 4). The median confidence level (on a 5-point scale) was never more than 2 prior to their participation in ASIBS, with most participants not having any prior knowledge of SAS, whereas considerable gains ranging from a median of 2 to 4 were made in their ability to conduct and interpret statistical analyses and SAS output. We did not

observe improvement with respect to their confidence in conducting bivariate analyses for categorical data, but when asked about their ability to conduct bivariate analyses for categorical data in SAS and interpret their results from these analyses, improvements were observed.

A total of 14 (70%) participants responded to the curriculum-related and quality-related evaluations. Median (IQR) overall satisfaction with ASIBS, on a 1 to 10 scale, was 8 (8–9). In fact, they all indicated that they would recommend ASIBS to their colleagues. They also found the ASIBS Web site to be highly useful with a median (IQR) of 4 (4–5), on a scale of 1 to 5. The majority (56%, n = 9) thought that having the 8-week short course take place between January and March was best to ensure maximum engagement, with a quarter suggesting that ASIBS be offered between April and June.

Discussion

The first year of ASIBS was highly successful with regards to the statistical competency-related and SAS proficiency-related gains by participants along with their overall satisfaction with the quality and administration of the short course. Moreover, we were able to implement very effective strategies to ensure a racially/ethnically diverse cohort. This is of particular importance, given our primary goal to reduce the racial/ethnic gap in promotion among faculty in academic medicine. In addition, the fact that the majority of the first cohort was female is especially important since being a URM and female in academic medicine can decrease the chance of receiving NIH funding. Our effort to increase research capacities through statistical training certainly has the potential to improve funding outcomes for these important subpopulations of the clinical and translational investigator workforce.

There are 3 major strengths of the design and implementation of ASIBS. First, we were able to develop a comprehensive curriculum tailored to meet the research needs of early stage clinical investigators,

Table 4. *Short-term pre-post competencies-related evaluation for the first year of the Applied Statistical Independence in Biological Systems (ASIBS) Short Course*

	n	Pre-ASIBS evaluation [median (IQR)]	Post-ASIBS evaluation [median (IQR)]
Confidence conducting the following analyses			
Bivariate analyses for continuous data	15	2 (1–4)	3 (2–5)
Bivariate analyses for categorical data	15	2 (1–3)	2 (2–5)
Simple linear regression	14	2 (1–4)	3 (3–5)
Multiple linear regression	14	2 (1–3)	3 (3–4)
Simple logistic regression	14	1 (1–3)	3.5 (2–4)
Multiple logistic regression	14	1 (1-2)	3 (2–4)
Confidence conducting the following analyses using SAS			
Bivariate analyses for continuous data	15	1 (1–1)	2 (2–3)
Bivariate analyses for categorical data	15	1 (1–1)	2 (2–3)
Simple linear regression	14	1 (1–1)	3 (3–4)
Multiple linear regression	14	1 (1–1)	3.5 (3–4)
Simple logistic regression	14	1 (1–1)	3 (2–4)
Multiple logistic regression	14	1 (1–1)	3 (2–4)
Confidence interpreting results from the following analyses			
Simple linear regression	14	2 (1–4)	4 (3–5)
Multiple linear regression	14	2 (1–2)	4 (3–5)
Simple logistic regression	14	2 (1–3)	4 (3–4)
Multiple logistic regression	14	2 (1–2)	3.5 (3–4)
Confidence interpreting SAS output from the following analyses			
Simple linear regression	14	1 (1–1)	4 (3–5)
Multiple linear regression	14	1 (1–1)	4 (3–5)
Simple logistic regression	14	1 (1–1)	4 (3–4)
Multiple logistic regression	14	1 (1–1)	3.5 (3–4)

while ensuring that we could successfully meet our diversity-related objectives. Second, to our knowledge, ASIBS is the first federally-funded short course with a mixed online and in-person curriculum directly aimed at increasing the statistical competency of junior faculty and postdoctoral fellows in academic medical centers nationwide. This type of curriculum is ideal for most junior investigators who cannot take long lengths of time away from their clinical responsibilities. Third, and the most important, ASIBS provided the opportunity for cross-disciplinary and cross-institutional collaboration among the participants, therefore serving as an important means for the exchange of ideas and research innovation.

We recognize 2 major areas for quality improvement of ASIBS. First, while we conducted extensive pre- and post-course evaluations, the response rate ranging from 70% to 75% was less than expected. Given that the evaluations provide us with valuable information about the background statistical knowledge of participants, increased knowledge gained as a result of participating in ASIBS, and strengths of the curriculum along with essential areas for improvement, we intend to re-emphasize the importance of these evaluations and highly encourage participants to submit the precourse evaluations upon acceptance to ASIBS and the postcourse evaluations while on-site for the in-person part of the course. Second, while we observed improvements in the statistical competency and statistical computing proficiency of the participants, the gains were more substantial with regards to their confidence in interpreting results than their confidence in conducting statistical analyses independently. To address this limitation, we will encourage future participants to bring de-identified data sets from their current research projects to the in-person session so that they can discuss with the instructors the best analytic strategies to answer their research questions and start conducting appropriate preliminary analyses. Supplementing the in-person curriculum with relevant analytic and statistical computing instruction for participants' specific research questions and data sets may increase future participants' confidence with regards to conducting their own analyses.

Conclusions

ASIBS provides a unique educational model to ensure that junior clinical investigators and postdoctoral fellows in academic medicine have sufficient statistical and methodological knowledge to thrive as clinical and translational investigators. This is extremely important, given the overwhelming call for innovative and rigorous, yet reproducible, study designs and analytical strategies in the clinical research community [5]. Although we have presented the findings from the short-term evaluation of the first cohort of participants, future research will evaluate the long-term impact of ASIBS on obtaining research funding, number and quality of publications, and promotion, especially with regards to increasing the opportunities for and decreasing time to promotion of URMs in academic medical centers.

Acknowledgments

Disclosures

The authors have no conflicts of interest to declare.

References

1. **Jeffe DB, Yan Y, Andriole DA.** Do research activities during college, medical school, and residency mediate racial/ethnic disparities in full-time faculty appointments at U.S. Medical schools? *Academic Medicine: Journal of the Association of American Medical Colleges* 2012; **87**: 1582–1593.

2. **Ginther DK, et al.** Race, ethnicity, and NIH research awards. *Science* 2011; **333**: 1015–1019.

3. **Fang D, et al.** Racial and ethnic disparities in faculty promotion in academic medicine. *JAMA* 2000; **284**: 1085–1092.

4. **Harris PA, et al.** Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics* 2009; **42**: 377–381.

5. **National Institute of General Medical Sciences.** Clearinghouse for training modules to enhance data reproducibility [Internet], n.d. [cited Nov 16, 2016]. (https://www.nigms.nih.gov/training/pages/clearinghouse-for-training-modules-to-enhance-data-reproducibility.aspx)