

# The Vanderbilt Nigeria Biostatistics Training Program (VN-BioStat): Results From a Skills Workshop

Bryan E. Shepherd<sup>1,2</sup>, Nafiu Hussaini<sup>3</sup>, Alex Huang<sup>2</sup>, Chelsea Van Wyk<sup>2</sup>, Meira S. Kowalski<sup>2</sup>, Donna J. Ingles<sup>4</sup>, C. William Wester<sup>2,5</sup>, Chun Li<sup>6</sup>, Muktar H. Aliyu<sup>2,7</sup>

<sup>1</sup> Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN

<sup>2</sup> Vanderbilt Institute for Global Health, Vanderbilt University Medical Center, Nashville, TN

<sup>3</sup> Department of Mathematics, Bayero University, Kano, Nigeria

<sup>4</sup> Office of the Vice Provost for Research and Innovation, Vanderbilt University, Nashville, TN

<sup>5</sup> Department of Medicine, Division of Infectious Diseases, Vanderbilt University Medical Center, Nashville, TN

<sup>6</sup> Keck School of Medicine, University of Southern California, Los Angeles, CA

<sup>7</sup> Department of Health Policy, Vanderbilt University Medical Center, Nashville, TN

Correspondence: Muktar H. Aliyu, Vanderbilt Institute for Global Health, Vanderbilt University Medical Center, Nashville, TN

Received: October 29, 2023 Accepted: December 29, 2023 Online Published: January 10, 2024

doi:10.5539/ijsp.v12n6p66

URL: <https://doi.org/10.5539/ijsp.v12n6p66>

## Abstract

The Vanderbilt-Nigeria Biostatistics Training Program (VN-BioStat) aims to establish a research and training platform for biostatisticians doing HIV-related research in Nigeria, including enhancing mid-level biostatistics capacity through annual workshops. This paper describes findings from the inaugural workshop in Kano, Nigeria. Participants were surveyed before and after the workshop to assess their self-perceived familiarity with and confidence in their abilities to use statistical software and apply specific statistical techniques, as well as to gather feedback regarding the conduct of the workshop and future topic areas. Of the 23 participants enrolled in the workshop, 22 (96%) completed both pre- and post-workshop assessments. In both pre-workshop and post-workshop surveys, participants ranked their confidence in statistical skills using Likert scales. Scores were transformed to a 0-100 scale, and averages computed. Participants also shared open-ended feedback about the workshop and suggested future topic areas. Before the training, the average participant reported having either a "beginner" (30% of participants) or "moderate" (43%) level of familiarity with R. Many participants (65%) rated themselves as having "moderate" or "expert" familiarity with SPSS. Pre-workshop averages for confidence ranged from 26 to 64, with lowest confidence in "expanding continuous covariates in regression models and interpret results" and highest confidence in "fitting and interpreting results from a linear regression model". Post-workshop averages for confidence were all above 70. The lowest post-workshop score (74) was for "fit and interpret results from a semiparametric linear transformation model". The greatest increase in confidence was observed in "expanding continuous covariates in regression models using splines and interpreting results" and the lowest increase was in "fitting and interpreting results from a linear regression model." Participants offered positive feedback on instructor effectiveness (4.9/5) and overall course quality (4.9/5). While the overall course was rated on a 0-100 scale as "moderately difficult" (mean  $\pm$  SD: 40.5  $\pm$  17.5), the participants felt the course was highly organized (87.7  $\pm$  17.8), and the information was moderately easy to learn (81.9  $\pm$  15.9). Suggestions for future workshops included providing supplementary resources for out-of-classroom learning and releasing codes in advance to enhance participants' preparation. Among suggestions for future workshop topics, 80% of respondents listed survival analysis. Lessons learned provide insight into how short-term training opportunities can be leveraged to build biostatistics capacity in similar settings.

**Keywords:** biostatistics, capacity building, HIV, West Africa

## 1. Introduction

Biostatisticians play major roles designing, monitoring, and analyzing data from clinical trials aimed to develop the next generation of HIV therapies, reduce HIV incidence, and address HIV-associated comorbidities. Biostatistics expertise is also essential to making sense of large observational databases to better understand trends in the HIV continuum of care and HIV-related complications. In addition, biostatisticians play fundamental roles in laboratory studies aiming to

develop HIV vaccines or cures. In short, strong biostatistics support is critical for high quality HIV research.

Substantial HIV research is being conducted in Africa because of the high burden of HIV in the continent. Much of this work has been conducted collaboratively between institutions in high-income countries and medical centers in low- and middle-income countries (LMICs). As a result of these collaborations and targeted training programs, the number of biomedical researchers has grown significantly in Africa to the extent that in some places, including Nigeria, in-country research leadership has been established. However, as pointed out by others, “growth in biostatistics lags far behind” (Gezmu et al., 2011).

Both long-term and short-term training programs that build biostatistics capacity are especially needed in Africa (Machekano et al., 2015), where a growing portfolio of clinical research necessitates more in-country biostatistics leadership. Nigerian investigators have repeatedly expressed a need for more biostatistics support, particularly at the local level. The overarching goal of the Vanderbilt-Nigeria Biostatistics Training Program (VN-BioStat) is to develop such biostatistics leadership in Nigeria, thus advancing HIV research led by in-country investigators and providing a means to grow capacity across Africa through South-South partnerships (Shepherd et al., 2023). One of the objectives of VN-BioStat is to provide mid-level biostatistics training for HIV researchers/data scientists from West Africa through annual in-country workshops. The workshops are designed to provide HIV researchers/data scientists with training in contemporary biostatistics, to help them develop mid-level biostatistics skills and understanding that will improve their research, and to create a forum for gathering and creating a community of statisticians engaged in HIV and other biomedical research.

This paper describes the inaugural VN-BioStat week-long biostatistics workshop in Kano, Nigeria. As part of the workshop, participants were surveyed before and after to assess their self-perceived familiarity with and confidence in their abilities to use statistical software and apply specific statistical techniques. Participants also provided post-workshop feedback regarding the conduct of the workshop and future topic areas. This paper describes findings from these pre-/post-workshop surveys and offers insights that may inform the development of similar biostatistics workshops in the region.

## 2. Methods

### 2.1 Trainee Recruitment and Selection

The workshop was conducted at Aminu Kano Teaching Hospital (AKTH), a 750-bed tertiary care facility affiliated with Bayero University Kano (BUK) and located in Kano, Nigeria (population ~ 9.4 million). The workshop was widely advertised on the VN-BioStat webpage and throughout AKTH and relevant departments at BUK and a large nearby collaborating university in Zaria, Nigeria via emails, flyers, and posters. Faculty at AKTH/BUK identified potential students and invited them to attend. Potential students applied to participate in the workshop via a REDCap application (Harris et al., 2009). They were asked to provide background information to help tailor selection and guide instruction, a personal statement regarding their motivation to attend, and permission from their supervisor indicating approval for their attendance.

The workshop had a capacity of 25 students to ensure that each participant could receive individual attention and to allow all participants to fully engage in interactive/breakout sessions. VN-BioStat leadership examined all complete applications (total of 48) and selected the top 25; priority was given to female students and applicants from AKTH/BUK with a strong quantitative background who were currently engaged in HIV research. A total of 23 students (5 women) participated in the workshop; 5 had a PhD in statistics, 10 had an MS in statistics, 4 had a PhD in mathematics, 2 had an MS in mathematics, 1 had a PhD in computer science, and 1 had a BS in computer science. A limited number ( $n=10$ ) of travel scholarships were offered for students living outside metropolitan Kano. No registration fees were charged, and meals and breaks were provided during the workshop meeting times.

### 2.2 Specific Training Activities

The workshop was held between June 12 - 16, 2023, and included five full days of instruction and hands-on learning. The workshop was preceded by a pre-workshop introduction to R statistical software held June 9-10. Participants were asked to bring a laptop to the training workshop and computers were loaned to those who did not have laptops. The instruction was hands-on and utilized freely available R statistical software. Students were asked to install R on their computers prior to the workshop. A recent and successful NIH-funded workshop in Kano, Nigeria introduced AKTH/BUK junior faculty to basics of R analysis (Aliyu et al., 2022), although nearly all ( $n=22$ ) workshop attendees participated in the pre-workshop introduction to R training.

The pre-workshop R training was taught by a biomathematics faculty member at AKTH. The main workshop was co-taught by biostatistics faculty from Vanderbilt University and the University of Southern California. A physician scientist with expertise in HIV also attended much of the workshop, helping to ground the instruction in practical

problems of importance to the field of HIV.

Workshop topics were decided through consultations between VN-BioStat program leaders and other Nigerian investigators. The workshop targeted trainees with some familiarity with statistics and biomedical research but was taught in such a manner to provide enough of the necessary foundation for participants with less experience and as a refresher for participants with more experience. The workshop was titled “*Linear, logistic, and ordinal regression models with R statistical software.*” Day 1 focused on logistic regression. Day 2 focused on linear regression and including covariates in regression models with interaction terms and natural splines. Day 3 focused on ordinal regression models. Day 4 focused on cumulative probability models (Liu et al., 2017). Day 5 was a catch-up/review day. Each day included presentations and then hands-on analysis/coding with R. Computing exercises included simulating and analyzing simulated data with the statistical technique discussed in lectures, with a particular emphasis on using simulations to do power and sample size calculations.

Trainees were given de-identified data on 2500 people living with HIV and receiving care at AKTH (Aliyu et al., 2019; Wudil et al., 2021) and were asked to apply the statistical techniques discussed in the lectures to investigate associations with hypertension (logistic regression), Joint National Committee blood pressure classification (normal, pre-hypertension, Stage 1 hypertension, Stage 2 hypertension) (ordinal regression), and blood pressure (linear regression and cumulative probability models). Instruction on the dataset and the scientific relevance of various predictor variables in the AKTH dataset was provided by the physician scientist with expertise in HIV who was the principal investigator for the study that collected the data.

As the students were working on computing exercises, both biostatistics faculty members moved around the room, answering questions. After the students had the opportunity to complete the computing exercise, one of the instructors would demonstrate to the entire classroom how they would do the coding/analyses. The other instructor would continue to circulate the room to ensure that students were following along. All didactic materials and codes were posted on a workshop website each evening so that students could continue to study or work on analyses.

### 2.3 Data Collection

Participants were surveyed using REDCap to assess their self-perceived familiarity with, and confidence in their abilities to use statistical software. REDCap was also used to gather feedback regarding the conduct of the workshop and future topic areas. In both pre-workshop and post-workshop surveys, participants ranked their confidence in statistical skills using 4 category Likert scales (e.g., none, beginner, moderate, expert; or not confident at all, a little confident, somewhat confident, very confident). Scores were transformed to a 0-100 scale (e.g., 0, 33, 67, 100 for none, beginner, moderate, and expert, respectively). Averages from these responses were then computed.

Participants also shared open-ended feedback about what they liked about the workshop, what they would change, and suggestion for future topic areas. We used a word cloud generator to identify the most frequently mentioned words and phrases, highlighting key aspects to consider for upcoming workshops. Responses that did not necessarily have the same phrasing but referred to the same topic were summed together.

## 3. Results

Of the 23 participants enrolled in the workshop, all (100%) completed the pre-workshop surveys and 22 (96%) completed the post-workshop assessments. Before the training, the average participant reported having either a “beginner” (30% of participants) or “moderate” (43%) level of familiarity with R; only three respondents reported having no prior experience with the software. Many participants (65%) rated themselves having “moderate” or “expert” familiarity with SPSS, 17% expressed “moderate” familiarity with Stata, and 17% expressed “moderate” familiarity with SAS.

The pre- and post-workshop surveys captured information regarding trainees’ confidence in fitting and interpreting results from several different statistical methods that were covered during the workshop. Pre-workshop averages for confidence ranged from 26 to 64, with lowest confidence in “expanding continuous covariates in regression models using splines and interpret results” and highest confidence in “fitting and interpreting results from a linear regression model” (Table 1). Post-workshop averages for confidence were all above 70. The lowest post-workshop score (74) was for “fitting and interpreting results from a semiparametric linear transformation model,” which was taught on the fourth day of instruction and expected to be the most advanced topic. The greatest increase in confidence, moving from 26 pre-workshop to 80 post-workshop, was observed in “expanding continuous covariates in regression models using splines and interpreting results.” The lowest increase in confidence, moving from 64 to 87, was in “fitting and interpreting results from a linear regression model.” At the end of the workshop, the percentage of participants whose confidence improved or was already at the “very confident” level pre-workshop ranged from 86% to 96% across the various statistical methods (Table 1).

Table 1. Average confidence pre-workshop vs. post-workshop for abilities in R, along with the gap between the pre-and post-workshop averages as well as the percent increase in confidence, Kano, Nigeria

Topic	Pre-Workshop	Post-Workshop	Percentage of participants who were 'very confident' both pre- and post-workshop	Percentage of participants whose confidence improved	Percentage of participants whose confidence improved or were 'very confident' both pre- and post-workshop
Fit and interpret results from a logistic regression model	61.1	87.0	27.3	63.6	90.9
Fit and interpret results from a linear regression model	63.9	87.0	36.4	59.1	95.5
Fit and interpret results from an ordinal regression model	47.2	84.1	18.2	77.3	95.5
Fit and interpret results from a semiparametric linear transformation model	34.7	73.9	0	86.4	86.4
Expand continuous covariates in regression models using splines and interpret results	26.4	79.7	0	86.4	86.4
Include and interpret interactions in regression models	38.9	82.6	4.5	86.4	90.9
Perform power/sample size calculations	37.5	82.6	4.5	81.8	86.4
Simulate data to investigate power and properties of a statistical method	36.1	84.1	9.1	77.3	86.4
Apply statistics to biomedical applications	54.2	82.6	22.7	68.2	90.9

Table 2 shows a summary of findings from the post-workshop course and instructor evaluation. Participants offered positive feedback on instructor effectiveness (4.9/5) and overall course quality (4.9/5). While the overall course was rated on a 0-100 scale as "moderately difficult" (mean  $\pm$  SD: 40.5  $\pm$  17.5), the trainees felt the course was highly organized (87.7  $\pm$  17.8), and the information was moderately easy to learn (81.9  $\pm$  15.9). The overwhelming majority of respondents felt comfortable in putting the knowledge learned into practice (82.4  $\pm$  14.8). All respondents indicated that they would be "very likely" to recommend the course to fellow clinical researchers (100%).

Table 2. Post-workshop course and instructor evaluation, Kano, Nigeria

Item	Mean ±SD
<b>Difficulty of the course</b> (100 = “much too hard”, 1 = “much too easy”)	40.5 ± 17.5
<b>Organization of the course</b> (100 = “extremely well organized”, 1 = “not at all organized”)	87.7 ± 17.8
<b>Ease of learning information</b> (100 = “easy to learn”, 1 = “difficult to learn”)	81.9 ± 15.9
<b>Level of comfort in applying statistical knowledge</b> (100 = “extremely comfortable”, 1 = “not comfortable at all”)	82.4 ± 14.8
<b>Effectiveness of the instructors</b>	4.9 of 5
<b>Overall rating for the course</b>	4.9 of 5
<b>Likelihood of recommending course to peers</b> Very likely	100%

Suggestions to improve future workshops included providing supplementary resources for out-of-classroom learning and releasing codes in advance to enhance participants’ preparation. Among suggestions for future VN-BioStat workshop topics, 80% of respondents listed survival analysis as an area of interest. Other topics commonly listed by respondents included techniques for handling missing data, causal inference, data management, and disease modeling (Figure 1).



Figure 1. Word cloud representing the frequency of terms used by respondents when asked about topics they would like to see covered by future biostatistics workshops, Kano, Nigeria

**4. Discussion**

In this manuscript, we describe our first workshop as part of VN-BioStat to offer short-term biostatistical training for data scientists and HIV researchers in Nigeria. In general, the workshop was judged by participants to be effective, of high quality, moderately difficult, and well-organized. Trainees’ self-reported knowledge of various statistical topics

increased over the course of the training. Participants also reported feeling comfortable putting the knowledge learned into practice and were “very likely” to recommend the course to fellow clinical researchers.

Several aspects of this workshop contributed to its success. First, the workshop format included interactive, hands-on training in specific statistical techniques using R, emphasizing their application within the context of HIV research. This approach fostered an environment conducive to active learner engagement and provided participants with a valuable opportunity to bridge the gap between theoretical concepts and practical applications. In addition, the workshop was tailored to individuals with backgrounds in quantitative sciences, including statistics, mathematics, and computer science. This focus allowed for the incorporation of more advanced quantitative concepts, which would have been challenging with a less quantitatively skilled group of trainees. Finally, the workshop served as a platform to showcase the exciting work biostatisticians can do in HIV research, and to introduce attendees to one-year biostatistics fellowship opportunities at Vanderbilt University Medical Center, also linked to the VN-BioStat grant. This approach builds a pipeline of competitive applicants for the fellowship program, thereby increasing its chances of long-term success.

This report has limitations. First, the relatively small sample size of learners restricts the extent to which we can generalize our findings. In addition, the absence of a control group of non-participants limits our ability to effectively assess potential pretest sensitization effects. However, the high survey response rates achieved in our study help reduce the risk of response bias, thereby enhancing the validity of our findings. Finally, expressing confidence in one’s ability to fit and interpret statistical models is different from effectively doing these things in practice. Although workshop trainees expressed high levels of confidence and improvement in their understanding of the methods taught, we did not formally measure their post-workshop abilities to perform these tasks in practice.

Future workshops will take into consideration the valuable feedback received from our trainees. This includes the provision of supplementary resources for self-directed learning and releasing codes in advance to facilitate better preparation. We will also place emphasis on requested topics such as survival analysis, techniques for handling missing data, and data management, among others. In addition, a future survey could help assess whether the post-workshop gain in confidence is long-term or temporary.

Our efforts at building biostatistics capacity in Nigeria are motivated by and modeled in part after other programs for biostatistics training in sub-Saharan Africa (Chirwa et al., 2020; Machezano et al. 2015). Annual workshops, such as the workshop described here, are one component of the VN-BioStat training program aimed at providing short-term training and awareness of biostatistics for a moderate number Nigerian researchers. The sustainability of our capacity building effort is supported by our longstanding and highly successful research partnership with the Aminu Kano Teaching Hospital (AKTH) and Bayero University Kano (BUK) in Kano, Nigeria. It has been shown by others that interactive workshops that review fundamental biostatistics concepts and theorems effectively bridge the gap between biostatistics theory and evidence-based decision-making for clinicians (Nelson 2018). We will consider expanding future workshop slots to include clinician scientists from our partner institutions who express an interest in biostatistics.

In summary, we demonstrated that an interactive workshop aimed at building biostatistics capacity in Nigeria was well-received and effective. The insights gained from this experience will inform the development of similar workshops in other parts of Africa.

## References

- Aliyu, M. H., Wudil, U. J., Ingles, D. J., Shepherd, B. E., Gong, W., Musa, B. M., ... Wester, C. W. (2019). Optimal management of HIV- positive adults at risk for kidney disease in Nigeria (Renal Risk Reduction "R3" Trial): protocol and study design. *Trials*, 20(1), 341. <https://doi.org/10.1186/s13063-019-3436-y>
- Aliyu, M. H., Sani, M. U., Ingles, D. J., Tsiga-Ahmed, F. I., Musa, B. M., Byers, M. S., ... Wester, C. W. (2022). Building Physician-Scientist Skills in R Programming: A Short Workshop Report. *International Journal of Translational Medical Research and Public Health*, 6(1), e418. <https://doi.org/10.21106/ijtmrph.418>
- Chirwa, T. F., Matsena Zingoni, Z., Munyewende, P., Manda, S. O., Mwambi, H., Kandala, N. B., ... Todd, J. (2020). Developing excellence in biostatistics leadership, training and science in Africa: How the Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB) training unites expertise to deliver excellence. *AAS Open Research*, 3, 51. <https://doi.org/10.12688/aasopenres.13144.2>
- Gezmu, M., DeGruttola, V., Dixon, D., Essex, M., Halloran, E., Hogan, J., ... Neaton, J. D. (2011). Strengthening biostatistics resources in sub-Saharan Africa: research collaborations through U.S. partnerships. *Statistics in Medicine*, 30(7), 695–708. <https://doi.org/10.1002/sim.4144>
- 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. *Journal of Biomedical Informatics*, 42(2), 377–381. <https://doi.org/10.1016/j.jbi.2008.08.010>

- Liu, Q., Shepherd, B. E., Li, C., & Harrell, F. E., Jr (2017). Modeling continuous response variables using ordinal regression. *Statistics in Medicine*, 36(27), 4316–4335. <https://doi.org/10.1002/sim.7433>
- Machekano, R., Young, T., Conradie, W., Rusakaniko, S., & Thabane, L. (2015). Workshop report: building biostatistics capacity in sub-Saharan Africa - taking action. *Pan African Medical Journal*, 21(1). <https://doi.org/10.11604/pamj.2015.21.167.5827>
- Machekano, R., Young, T., Rusakaniko, S., Musonda, P., Sartorius, B., Todd, J., ... Chikte, U. (2015). The Africa Center for Biostatistical Excellence: a proposal for enhancing biostatistics capacity for sub-Saharan Africa. *Statistics in Medicine*, 34(27), 3481–3489. <https://doi.org/10.1002/sim.6572>
- Nelson A. (2018). An Interactive Workshop Reviewing Basic Biostatistics and Applying Bayes' Theorem to Diagnostic Testing and Clinical Decision-Making. *MedEdPORTAL: the Journal of Teaching and Learning Resources*, 14, 10771. [https://doi.org/10.15766/mep\\_2374-8265.10771](https://doi.org/10.15766/mep_2374-8265.10771)
- Shepherd, B. E., Hussaini, N. N., Ingles, D. J., Van Wyk, C., Cassell, H. M., Wester, C. W., & Aliyu, M. H. (2023). Training the next generation of Biostatisticians in West Africa: The Vanderbilt Nigeria Biostatistics Training Program (VN-BioStat). *Journal of Global Health Reports*, 7, e2023067. <https://doi.org/10.29392/001c.88939>
- Wudil, U. J., Aliyu, M. H., Prigmore, H. L., Ingles, D. J., Ahonkhai, A. A., Musa, B. M., ... Wester, C. W. (2021). Apolipoprotein-1 risk variants and associated kidney phenotypes in an adult HIV cohort in Nigeria. *Kidney International*, 100(1), 146–154. <https://doi.org/10.1016/j.kint.2021.03.038>

### **Acknowledgements**

We would like to thank the members of our Training Advisory Committee. We also appreciate the secretarial support provided by Hauwa Sani and Bilya Musa at the Aminu Kano Teaching Hospital, Nigeria.

### **Funding**

We acknowledge the members of our Training Advisory Committee. This work is supported by the Fogarty International Center of the U.S. National Institutes of Health (NIH), award number 1D43TW012268. The workshop presented some new statistical methods whose development was funded by R01AI093234 and incorporated data funded by U01DK112271. The findings and conclusions are those of the authors and do not necessarily represent the official position of the NIH, the Department of Health and Human Services or the government of the United States of America.

### **Author contributions**

BES, NH, DI, and MA conceived the program; BES, NH, CL, CWW and MA designed the training program. All authors participated in drafting the manuscript and critical review of the article. BES, AH, CWW, and MA finalized the manuscript; BES is the guarantor of the paper; all authors approve the final version of the manuscript.

### **Informed consent**

Obtained.

### **Ethics/consent**

Approved as exempt by the Ethics Review Committee of Aminu Kano Teaching Hospital and the Vanderbilt University Institutional Review Board.

### **Provenance and peer review**

Not commissioned; externally double-blind peer reviewed.

### **Data availability statement**

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

### **Data sharing statement**

No additional data are available.

### **Open access**

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

### **Copyrights**

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