

Original Publication

OPEN ACCESS

An Interactive Workshop Reviewing Basic Biostatistics and Applying Bayes' Theorem to Diagnostic Testing and Clinical Decision-Making

Adin Nelson, MD*

*Corresponding author: adin.nelson@rutgers.edu

Citation: Nelson A. An interactive workshop reviewing basic biostatistics and applying Bayes' theorem to diagnostic testing and clinical decision-making. *MedEdPORTAL*. 2018;14:10771. https://doi.org/10.15766/mep_2374-8265.10771

Copyright: © 2018 Nelson. This is an open-access publication distributed under the terms of the Creative Commons Attribution-NonCommercial-Share Alike license.

Abstract

Introduction: Sensitivity, specificity, and predictive values—the basic statistics behind using and interpreting screening and diagnostic tests—are taught in all medical schools, yet studies have shown that a majority of physicians cannot correctly define and apply these concepts. Previous work has not rigorously examined this disconnect and attempted to address it. **Methods:** We used adult learning theory to design a case-based interactive workshop to review biostatistics and apply them to clinical decision-making using Bayes' theorem. Participants took an anonymous multiple-choice pretest, posttest, and delayed posttest on definitions and application of the concepts, and we compared the scores between the three tests. Several experiences with early iterations provided feedback to improve the workshop but were not included for analysis. **Results:** We conducted the finalized workshop with 54 pediatrics students, residents, and faculty. All learners completed the immediate pre- and posttests, and eight completed the delayed posttest. Average scores rose from 4.5/8 (56%) on the pretest to 6.5/8 (81%) on the posttest and 6.4/8 (80%) on the delayed posttest. Two-tailed *t* tests showed $p < .001$ for the difference between the pretest and both posttests, and post hoc power analysis showed a power of 99% to detect the observed differences. There was no significant difference ($p = .8$) between the posttest and delayed posttest. **Discussion:** Our work demonstrates that an interactive workshop reviewing basic biostatistics and teaching rational diagnostic testing using Bayes' theorem can be effective in connecting theoretical knowledge of biostatistics to evidence-based decision-making in real clinical practice.

Keywords

Medical Education, Evidence-Based Medicine, Clinical Decision-Making, Epidemiology and Biostatistics

Educational Objectives

By the end of this workshop, learners will be able to:

1. Define sensitivity, specificity, and positive and negative predictive values both mathematically and in plain language.
2. Explain the similarities and differences between sensitivity, specificity, and positive and negative predictive values.
3. Explain how predictive values depend on the prevalence within the population.
4. Apply likelihood ratios to calculate posttest probability from pretest probability using Bayes' theorem.
5. Evaluate the utility of a particular test in a specific clinical scenario based on the pretest probability and the known characteristics of the test.

Introduction

Physicians depend on diagnostic tests to make clinical decisions, so mistakes in applying and interpreting these tests can lead to poor medical care. The theory behind diagnostic testing is well described and routinely taught in medical schools, yet studies have shown that a majority of physicians cannot correctly define and apply these concepts.¹ An unfortunately common clinical scenario easily illustrates this point: A patient presents with fever and sore throat but no tonsillar enlargement or erythema; the patient has a positive rapid strep test and receives antibiotics with little or no consideration of whether the test result is

Appendices

- A. Quiz.docx
- B. Quiz With Answers.docx
- C. Slide Presentation.pptx
- D. Handout.docx
- E. Instructor Guide.docx
- F. Train the Trainer Video .mp4

All appendices are peer reviewed as integral parts of the Original Publication.

true or whether it was appropriate to use that test in that scenario in the first place. (Similar examples from other disciplines might include mammography for breast cancer, PSA screening for prostate cancer, and head CT scans for intracranial hemorrhage in patients at low risk for these conditions.) In order to ensure high-quality clinical care, we must train physicians to understand the concepts behind diagnostic testing and apply them in their day-to-day medical decision-making.

Sensitivity, specificity, and predictive values are the basic statistics behind using and interpreting clinical tests. These ideas are taught in all medical schools and tested on the United States Medical Licensing Examination (USMLE).² First-time pass rates on the USMLE are generally high—suggesting that students do understand these concepts on a theoretical level. When it comes to clinical practice, however, medical students, residents, and experienced clinicians alike struggle to apply that knowledge to actual medical decision-making.¹ Board review programs routinely review these basic concepts but do not necessarily provide learners with the tools to apply the concepts in clinical practice. Board review programs also tend to focus on the mathematical formulas for sensitivity, specificity, and others, rather than on an intuitive natural-language understanding of the concepts. Furthermore, board review programs come into play towards the end of a given phase of a learner's clinical training at which point it may be difficult to change practice patterns that have already been established.

Three prior workshops have been published in *MedEdPORTAL* to address this gap, but all three present challenges to broad implementation. Two are designed as structured team-based learning (TBL) exercises for medical students in the preclinical phase of their education.^{3,4} These workshops require students, teachers, and a learning environment that are already familiar with the TBL system and comfortable using it, so they are difficult to generalize to a program without that structure in place. These workshops also target medical students early in their education, but based on the USMLE Content Outline and exam results, it is likely more experienced trainees who need more review of these topics. Teaching rational diagnostic testing early in medical school is also intrinsically difficult because the challenge is not the concepts themselves but applying them to clinical care, and first-year medical students do not yet have the clinical experience to make that connection. The third prior *MedEdPORTAL* publication in this area does focus on residents, fellows, and faculty, but it is a frontal lecture with no active or interactive elements and focuses on using the concepts of sensitivity, specificity, and others to interpret published medical literature, not on using those concepts in clinical practice.⁵ Additionally, none of those prior publications rigorously evaluate their workshops' effectiveness at higher levels of Kirkpatrick's pyramid.⁶ One includes no outcome data at all,⁴ another reports only the learners' impressions of the workshop,⁵ and the third reports the learners' average scores on final exam questions related to these topics but does not include a direct comparison to suggest that those scores are due to the workshop.³

Our resource improves upon those prior publications in four ways. First, it is designed as a stand-alone workshop with no outside requirements such as the structure and workflow of TBL. Second, our workshop is targeted primarily at residents, although it can also be used with more experienced medical students, fellows, or faculty. These learners are far enough away from their biostatistics classes that they likely need a review of the basics, and they also have enough clinical experience to be ready to apply the concepts to clinical care. Third, our workshop is highly interactive. This helps maintain learner engagement and provides ample opportunities for learners to apply and practice the material. Fourth, we used anonymous multiple-choice pre- and posttests to directly assess the workshop's effectiveness. This workshop seeks to address the disconnect between theoretical biostatistics taught early in medical education and later clinical practice by focusing both on definitions of basic biostatistics and on application of those concepts to medical decision-making through the framework of Bayes' theorem.

Methods

We designed and tested this workshop in a university-based general pediatrics residency program housed in an inner-city hospital. Third- and fourth-year medical students and all levels of pediatrics residents in our program participate in weekly academic half-day didactic sessions. The academic half-day

model makes it possible to conduct a longer workshop including the pretest, posttest, and small-group exercise. The final version of this workshop took 2 hours. By eliminating the pre- and posttests and converting the small-group exercise to a homework assignment, however, we were also able to conduct the workshop in under an hour. This project was approved by the Rutgers New Jersey Medical School Institutional Review Board.

Before the workshop, learners completed a short, anonymous, multiple-choice pretest (Appendix A; with answers, Appendix B). The workshop was guided by a set of PowerPoint slides (Appendix C). The instructional design was grounded in the principles of adult learning theory: We opened the workshop with a clinical scenario in which a diagnostic test was used inappropriately and interpreted incorrectly, leading to unnecessary medical treatment for a patient. This scenario resonated with the residents, so it made the workshop material relevant to their daily lives and earned their attention. In the first section of the workshop, we reviewed basic biostatistics; we quizzed the learners on definitions of sensitivity, specificity, and positive and negative predictive values in both mathematical formulas and plain language. We quizzed the learners instead of reviewing the concepts frontally to acknowledge their prior knowledge and build upon it rather than reteaching basic material that the learners had encountered before.

In the second section of the workshop, we took a constructivist approach, using interactive exercises to help the learners discover for themselves the relationships between sensitivity, specificity, predictive values, and prevalence. We built these exercises around analyzing and applying a fictitious lab test in different situations. We used a fictitious test instead of a real clinical test (such as a chest X-ray or a rapid strep test) in order to avoid confounding from learners' prior knowledge about the test in question. Specifically, our example was the dog test: a test in which an object is offered to a dog to see whether or not the dog will eat it as a test for whether or not the object is food. (If the dog eats the object, that is a positive test, and the item is presumed to be food. If the dog does not eat the object, then that is a negative test, and the item is presumed to not be food.) The learners applied the dog test in different scenarios to demonstrate how the predictive value of the test varied with the prevalence of disease—or food, in this case.

In the final section of the workshop, we introduced Bayes' theorem as a more practical mathematical model for the concepts that the learners had discovered in their work with the dog test. We finally replaced the dog test with a real clinical case to demonstrate how these lessons could be applied in actual clinical decision-making. The case presented a young adult with possible influenza, and we applied Bayes' theorem and the known test characteristics of the rapid influenza test to analyze whether or not the test should be performed, what different test results might mean, and whether or not the patient should be treated with antiviral medication. At the end of the workshop, the learners broke up into groups and completed a small-group exercise (Appendix D) in which they practiced applying Bayes' theorem to a real clinical scenario and diagnostic test of their own choosing.

The workshop PowerPoint slides made extensive use of animations in which elements of the slide were revealed one at a time. This helped pace the presentation as we discussed each element as it was revealed on the screen. This approach also helped keep the learners focused on the current point of discussion; they could not get distracted by upcoming points because those were not revealed until needed.

A single faculty member can easily teach this workshop to a group of up to 20 or 30 learners. For a larger group, a second facilitator would be helpful—especially during the small-group exercise at the end. The ideal instructor for this workshop should have both a strong background in the theoretical concepts of biostatistics and relevant clinical experience. A nonclinician, no matter how well he or she understands the mathematical concepts involved, is less likely to be able to bridge the understanding gap between the theoretical concepts and actual clinical practice. The instructor does not necessarily have to be a physician; a nurse, dentist, pharmacist, or any other health care professional should be able to conduct

this workshop effectively, provided he or she has the necessary understanding of the material. Regardless of their background, any instructors implementing this workshop should set aside 1-2 hours to review the materials in advance.

The following workshop materials are included as appendices:

- Diagnostic testing quiz (Appendix A; with answers, Appendix B): All participants took this anonymous multiple-choice test immediately before and after the workshop. We also created an online version of the test in Google Forms (similar to SurveyMonkey) to allow participants to take it again 1-2 weeks after the workshop. We used the test scores at these three time points as a research tool to evaluate the effectiveness of the workshop. However, given the increasingly recognized power of quizzing as a learning tool,⁷ using pre- and posttests (and even a delayed posttest) may be beneficial to learners and not just as a research tool. When using the pre- and posttests, two copies of this document should be printed per learner.
- PowerPoint slides (Appendix C): This set of 32 slides guides the group through the interactive review of basic statistical concepts, the constructivist exercises using the dog test, the mechanics of Bayes' theorem, and a clinical example applying Bayes' theorem. The slides also visually emphasize key teaching points. The slides make heavy use of animation, so they are not suitable for printing. Because of the animations, a handheld remote control to advance the slides, while not strictly necessary, is highly recommended.
- Handout (Appendix D): This handout includes the key formulas necessary to apply Bayes' theorem and the prompts for the small-group assignment. Whether the small-group exercise is used during the workshop or assigned as independent homework, copies should be printed for all participants.
- Instructor's guide (Appendix E): The instructor's guide lays out educational objectives and instructional methods for the workshop as a whole. It also identifies the specific content goals of each slide, emphasizes key teaching points, and provides instructions for the interactive exercises.
- Train the trainer video (Appendix F): This 15-minute video featuring the slides and a voice-over by the author is intended as an adjunct to the instructor's guide to help visual and auditory learners understand how to implement the workshop. The video incorporates both demonstration of the slides and explanatory comments emphasizing key teachings points and explaining instructional methods and interactive exercises. Most importantly, the video also demonstrates how the animations in the PowerPoint slides are used to pace the presentation by revealing teaching points one at a time.

Resources required to conduct the workshop include the following:

- Audiovisual equipment to display the PowerPoint slides.
- A whiteboard and markers (or similar) for learners to work out the interactive examples.
- Printed copies of the pretest and posttest (if used) and the handout.
- Since this is an interactive workshop, not a front lecture, it is better suited to a conference room than a lecture hall. Chairs clustered around small tables, chairs around a larger conference table, and individual chairs with writing arms would all work.

Results

We conducted the final version of the workshop with 28 pediatrics residents at a residency program in a neighboring institution and again with an additional five medical students, 12 residents, and nine faculty members at our own institution. (None of the residents or students who participated in that workshop had been exposed to any of the prior workshops.) We used the pooled scores on the anonymous multiple-choice pre- and posttests to evaluate the workshop. All 54 learners completed the immediate pre- and posttests given on paper during the workshop. Despite multiple email and in-person reminders, however, only 10 learners completed the online delayed posttest.

The average score rose from 4.5/8 (56%) on the pretest to 6.5/8 (81%) on the posttest. This difference was sustained with an average score of 6.4/8 (80%) on the delayed posttest. Two-tailed unpaired t tests showed $p < .001$ for the differences between the pretest and both posttests. Effect size calculations showed Cohen's d s of 1.0 between pre- and posttests and 1.8 between pre- and delayed tests. Post hoc power analysis showed a power of 99% to detect the observed differences between these samples, despite the fact that only 10 participants completed the delayed posttest. There was not a statistically significant difference ($p = .8$) between the posttest scores and the delayed posttest scores.

Discussion

During the process of designing and testing this workshop, we conducted it five times. The first two times, we conducted the full 2-hour interactive workshop, initially with a group of 22 general pediatrics residents and third- and fourth-year medical students and then with a group of four preventive medicine residents. Those were earlier versions of the workshop (the slides, the tests, and the handout), so the test scores have not been included in the analysis, but the scores and feedback drove improvements in subsequent iterations. The initial workshops showed us that the test focused too much on basic definitions (which most of the learners got right even on the pretest) and not enough on understanding and applying Bayes' theorem. Resident feedback was generally positive, but several common points of difficulty emerged (particularly around applying Bayes' theorem), so we added slides to explicitly address those areas and explain those points. We then conducted the workshop for a group of ~40 medicine interns and for a group of ~35 pediatrics residents at a different institution; both of those used a shortened version of the workshop (excluding the pre- and posttests and assigning the small-group exercise as homework), so there were no test scores to include in the analysis. The workshops went well, and learner feedback was again generally positive. After those workshops, we again added material to the workshop to clarify confusion about the transition from working with a 2x2 table to working with Bayes' theorem and the homology between prevalence and positive predictive value and pre- and posttest proximities. Those experiences also showed that by eliminating the pre- and posttests and deferring the small-group exercise, the workshop could easily be done in under an hour.

Our work demonstrates that this workshop based on adult learning theory principles and built around analyzing and applying a fictitious diagnostic test is effective in reviewing basic biostatistics and teaching rational decision-making using Bayes' theorem. Understanding and applying this model can give trainees and clinicians the tools to make better-informed choices about using and interpreting clinical tests, and that should help decrease inappropriate utilization of tests and confusion due to false test results. In the process, we learned that most residents already knew (or at least remembered learning) the mathematical definitions of sensitivity, specificity, and predictive values, but they lacked an intuitive conceptual understanding of these test characteristics and did not know how to use Bayes' theorem to apply those concepts to answer clinical questions. The final version of the workshop is therefore designed to focus on those areas.

The major limitation of this work is the relatively small sample of learners included in the final analysis. A limited number of residents were available in our own and neighboring programs to participate in the project, and many of those learners had been exposed to earlier versions of the workshop and therefore could not be included in the evaluation of the final version. The difference in pre- and posttest scores shown in our analysis is highly statistically significant and well powered though, and the effect size is large. We also believe that the differences are educationally significant: An increase from 56% to 81% is, after all, the difference between a failing and a passing grade.

Moving forward, we plan to conduct this workshop with more learners and to increase the scope of learners included in data analysis to include third- and fourth-year medical students, fellows, and faculty. We hope that other institutions and residency programs will learn from our experience and employ this interactive workshop to review basic biostatistics and teach rational diagnostic testing based on Bayes' theorem to their own faculty and trainees.

Adin Nelson, MD: Assistant Professor, Department of Pediatrics, Rutgers New Jersey Medical School

Disclosures

None to report.

Funding/Support

None to report.

Informed Consent

All identifiable persons in this resource have granted their permission.

Ethical Approval

The Rutgers New Jersey Medical School Institutional Review Board approved this study.

References

1. Whiting PF, Davenport C, Jameson C, et al. How well do health professionals interpret diagnostic information? A systematic review. *BMJ Open*. 2015;5(7):e008155. <https://doi.org/10.1136/bmjopen-2015-008155>
 2. *USMLE Content Outline*. Philadelphia, PA: United States Medical Licensing Examination; 2017.
 3. Goedde M, Everse S, Wojewoda C. Diagnostic testing team-based learning. *MedEdPORTAL*. 2015;11:10155. https://doi.org/10.15766/mep_2374-8265.10155
 4. Bedinghaus J, Nelson D. Team-based learning of EBM: diagnostic testing. *MedEdPORTAL*. 2013;9:9519. https://doi.org/10.15766/mep_2374-8265.9519
 5. Mojica M. The Making Evidence Based Medicine Simple Series—diagnostic testing module. *MedEdPORTAL*. 2013;9:9475. https://doi.org/10.15766/mep_2374-8265.9475
 6. Kirkpatrick DL, Kirkpatrick JD. *Evaluating Training Programs: The Four Levels*. 3rd ed. San Francisco, CA: Berrett-Koehler Publishers; 2006.
 7. Brown PC, Roediger HL III, McDaniel MA. *Make It Stick: The Science of Successful Learning*. Cambridge, MA: Harvard University Press; 2014.
-

Received: July 6, 2018 | **Accepted:** October 4, 2018 | **Published:** November 9, 2018