Benjamin Kinnear*, Philip A. Hagedorn, Matthew Kelleher, Courtney Ohlinger
and Jonathan Tolentino

Integrating Bayesian reasoning into medical education using smartphone apps

<https://doi.org/10.1515/dx-2018-0065>

Received August 8, 2018; accepted February 6, 2019; previously published online February 28, 2019

Abstract

Background: Diagnostic reasoning is an important topic in medical education, and diagnostic errors are increasingly recognized as large contributors to patient morbidity and mortality. One way to improve learner understanding of the diagnostic process is to teach the concepts of Bayesian reasoning and to make these concepts practical for clinical use. Many clinician educators do not fully understand Bayesian concepts and they lack the tools to incorporate Bayesian reasoning into clinical practice and teaching.

Methods: The authors developed an interactive workshop using visual models of probabilities and thresholds, clinical cases, and available smartphone apps to teach learners about Bayesian concepts.

Results: Evaluations from 3 years of workshops at a national internal medicine chief resident conference showed high satisfaction, with narrative comments suggesting learners found the visual and smartphone tools useful for applying the concepts with future learners.

Conclusions: Visual models, clinical cases, and smartphone apps were well received by chief residents as a way to learn and teach Bayesian reasoning. Further study will be needed to understand if these tools can improve diagnostic accuracy or patient outcomes.

Keywords: Bayesian; diagnostic reasoning; likelihood ratio; probabilities; smartphone.

*Corresponding author: Benjamin Kinnear, MD, MEd, Department of Pediatrics, University of Cincinnati, College of Medicine, 3333 Burnet Ave, MLC 5018, Cincinnati, OH 45229-3039, USA, Phone: +513-541-4667, E-mail: kinneabn@ucmail.uc.edu

Philip A. Hagedorn: Department of Pediatrics, University of Cincinnati, College of Medicine, Cincinnati, OH, USA, E-mail: Philip.Hagedorn@cchmc.org

Matthew Kelleher: University of Cincinnati, College of Medicine, Cincinnati, OH, USA, E-mail: kellehmv@ucmail.uc.edu

Courtney Ohlinger: Department of Internal Medicine, University of Cincinnati, College of Medicine, Cincinnati, OH, USA, E-mail: ohlingcl@ucmail.uc.edu

Jonathan Tolentino: Stony Brook University, School of Medicine, Stony Brook, NY, USA, E-mail: jonathan.tolentino@stonybrookmedicine.edu

Background

Diagnostic error is increasingly recognized as a leading cause of patient morbidity and mortality [1]. Understanding and teaching diagnostic reasoning is one strategy to mitigate this problem [2], with multiple approaches being suggested [3–5]. Using Bayesian reasoning, defined here as the process of using base rate (pre-test) probabilities and new clinical information (history, exam findings, or tests) to calculate a revised (post-test) probability, is one approach to reduce diagnostic error by improving deliberate, analytical (System 2) reasoning [4–6].

Physicians often have poor understanding of concepts germane to Bayesian reasoning such as likelihood ratios, which can lead to inaccurate post-test probability estimation [7]. Using natural frequencies (rather than probabilities) and visual aids has shown some promise in improving Bayesian reasoning [7]. Most physicians still do not use formal Bayesian methods because of perceived lack of practicality and discomfort with performing the calculations [8, 9]. This suggests physicians often lack understanding and tools (e.g. technology for real-time calculation) to integrate Bayesian reasoning into clinical practice and bedside teaching. Our aim was to create a workshop that equips clinician educators with the tools to effectively teach and use Bayesian reasoning in clinical settings.

Methods

We developed a workshop that incorporated smartphone apps, visual models, and case-based learning. Smartphone apps convey multiple advantages for teaching Bayesian reasoning. They quickly calculate post-test probabilities once pretest probabilities and test characteristics are entered, which mitigates principal discomfort with calculations [8, 9]. Variables can be easily manipulated in the apps for different clinical situations or alter scenarios, using “what ifs” to compare and contrast how post-test probabilities might change based on different clinical information. Many apps contain easily accessible repositories of clinical prediction rules (e.g. Wells’ criteria) and describe test characteristics. Several apps also provide references to primary literature, allowing physicians to review the source material and determine if the reported test characteristics apply to their patient population. Finally, smartphones are nearly ubiquitous with medical

trainees [10], making apps readily available, portable, and relatively affordable teaching tools. Using visual aids has been suggested as a method to understand Bayesian reasoning [9, 11, 12], so we created a visual model that helps demonstrate the concepts of probabilities and thresholds in clinical contexts (Figure 1). We employed case-based learning followed by reflective discussion to encourage practice with

the concepts and apps, and to facilitate experiential learning [13]. The authors iteratively developed clinical cases that involve uncertainty in diagnosis to promote discussion amongst learners.

Our 90-min workshop begins with a case presentation of a patient with a febrile respiratory illness. Learners anonymously enter how probable they think it is that the patient has influenza into Poll

Figure 1: Visual representation of probabilities and thresholds used in Bayesian reasoning.

Panel (A) shows colored sections representing areas of probability with associated actions based on thresholds set by a provider. Each color next to the Fagan nomogram represents next steps based on thresholds set by the user. Green is below the test threshold, and no further testing is needed. Yellow is between the test and treat thresholds, and the clinician should consider getting further information via history, exam, or testing to move the probability into the red or green. The red is above the treat threshold, and the user should consider whatever management option they deem appropriate. The solid red line is an example case in which the pre-test probability is 50%, and a test with a LR of 10 is obtained, resulting in a post-test probability of 90%. Manipulation of this visual representation allows learners to see a map of how information affects their Bayesian reasoning. Panel (B) shows how groups of clinicians can recommend different actions, even when given the same information. For example, even if all of these clinicians agreed on a pretest probability of 70% (clinician 2's probability), they still disagree on next steps. Clinician 1 would recommend treatment, clinician 2 might be considering treatment but still unsure, and clinician 3 would want more information. Complicating things further, most clinicians do not share exactly the same pre-test probability (as illustrated), which can lead to even more confusion. Helping learners understand these concepts and how to verbalize them can facilitate deeper learning of the diagnostic process.

Everywhere (Poll Everywhere, San Francisco, CA, USA), an online audience response system. Next, they are given a negative rapid influenza test result with its sensitivity and specificity values, and are asked to enter a revised probability into Poll Everywhere. The broad range of submitted answers for both pre- and post-test probabilities are then reviewed and participants are asked to reflect on how they arrived at those numbers. The variation in answers leads to robust conversations about how clinicians set probabilities and interpret test results. We then review concepts about test characteristics such as sensitivity, specificity, and likelihood ratios, emphasizing the understanding of underlying concepts. Learners are introduced to the Fagan nomogram, [14] a tool that allows a user to draw a line from a pre-test probability (determined by the user) through a likelihood ratio value (determined by the “test”) to arrive at a post-test probability. The visual representation of these concepts (Figure 1) is displayed and manipulated through various scenarios to illustrate how changes in different variables affect each scenario, and to illustrate how multiple clinicians with different thresholds or probabilities can lead to provider disagreement and can create confusion for learners.

We then introduce multiple smartphone apps that allow for integration of pre-test probabilities, likelihood ratios, and clinical prediction rules at the bedside. This list of apps is not an exhaustive list for clinical practice, but rather a list of apps that we have found to be useful in teaching. The smartphone apps cost between \$0 and \$9.99; a few require subscription fees (Table 1). Participants work through the cases using the introduced smartphone apps, sharing with the group how they envision utilizing the concepts and apps for clinical practice and teaching learners.

For example, one case encourages participants to imagine themselves as supervisors on a ward team admitting a young adult patient with severe iron deficiency anemia and 3-year history of intermittent, severe abdominal pain, and diarrhea. After getting more case details, workshop participants discuss their pre-test probabilities of inflammatory bowel disease (IBD) and commit to a specific probability

between 0% and 100%. They are given the sensitivity and specificity for fecal calprotectin in IBD, and told that one of their imagined learners wants to send the test while another wants to refer the patient for endoscopy. Attendees role-play navigating this clinical decision with the use of a smartphone app and accompanying visual aids (Figure 1) to turn this disagreement between learners into a Bayesian learning opportunity. Using cases allows learners to practice applying the concepts and smartphone apps in a clinical scenario. Other cases integrate other concepts such as clinical prediction rules and risk/benefit considerations when ordering tests.

The Alliance for Academic Internal Medicine (AAIM) gave permission to share anonymous evaluation data. The University of Cincinnati Institutional Review Board deemed this analysis of aggregate evaluation data without individual identifiers as not human subject research.

Results

Our workshop has been delivered three times at the national AAIM chief resident meeting with an estimated 200 total learners (actual attendance was not recorded). Forty-nine evaluations were completed by attendees. The average weighted rating on a 5-point Likert scale over the 3 years for the prompt “Overall satisfaction with the session” was 4.32, with other workshops at the same conference averaging 3.96. Narrative comments from attendee free responses suggested that the visual model and smartphone apps would be potentially useful for teaching the concepts to their own learners:

Table 1: This shows the names of smartphone applications used during presentations and workshops for teaching Bayesian reasoning.

App name	Cost	Operating systems	Input pretest probability, sensitivity, specificity or LRs to get post-test probability	Includes clinical prediction rules	Repository of test characteristics (sensitivity, specificity, or LRs)	References supporting evidence
Bayes ^d	\$0.99	Apple	×			
Bayes' post-test probability calculator	Free	Apple	×			
EBM tools ^d	Free	Apple	×			
docLogica	Free	Apple	×	×	×	×
Docnomo	Free	Apple	×			
EBMcalc complete	\$9.99	Apple, Android	×	×		×
Medicine toolkit	\$4.99	Apple	×			×
Rx-Bayes ^d	\$0.99	Apple	×	×	×	×

The data are current as of July 2018. ^aLR – likelihood ratio. ^bClinical prediction rules – in-app access to rules used to establish probabilities of various diagnoses using combinations of epidemiologic information, clinical history, symptoms, exam findings, and previous test results. Examples include Wells' criteria for venous thromboembolism, Duke criteria for bacterial endocarditis, or hepatitis C cirrhosis probability. This can help users set pretest probabilities when available information allows. ^cRepository of test characteristics (sensitivity, specificity, or LRs) – in-app access to these characteristics for various tests, which prevents users from having to search external sources to find the appropriate values. ^dAs of submission of this manuscript, these are in need of updates from the app developer.

"I feel much more confident in my understanding of EBM diagnostic reasoning, and I expect that I will be using much of what I gained from this lecture to teach residents in the coming academic year".

"I had heard this information in the past and never really bought into it because it was poorly delivered and provided no way to actually use it. This solved both of those problems by having an engaging lecture and also immediately providing tools that allow for this to be clinically incorporated".

Discussion

Our experiences suggest teaching Bayesian reasoning with smartphone apps, visual models, and case-based learning is well received by chief resident participants. In addition to the aforementioned conference, we have used these tools to teach Bayesian reasoning at local student and resident conferences, faculty development sessions, interdisciplinary regional meetings, and two grand rounds. In conversations with learners across levels and disciplines, a common theme that arises is difficulty assigning specific numbers to pre-test probabilities due to a feeling of subjectivity. Our case-based discussions emphasize that pre-test probabilities are often subjective but are not arbitrary, as clinicians can use epidemiology, clinical experience, clinical prediction rules, and illness scripts to determine them. Our workshop provides opportunities for practice and feedback with clinical scenarios to develop comfort with setting probabilities.

Our description of this innovation has several limitations. First, the evaluation was based on a small sample size with few narrative comments. Second, the quantitative outcomes presented are low on Kirkpatrick's hierarchy [15], and do not assess higher level learning outcomes such as knowledge, attitudes, or behaviors. Third, evaluation data from other venues where this workshop has been given were not included in the study. Fourth, due to the niche market that these apps occupy, they are often built and maintained by small groups or individual developers, and can be subject to delayed updates, bug fixes, and operating system compatibilities that limit their long-term viability.

Conclusions

A workshop using visual models, clinical cases, and smartphone apps was well received by chief residents as a way to learn and teach Bayesian reasoning. We anticipate

that our approach can be adapted to others' local education needs, and inform the development of Bayesian reasoning curricula. Knowledge and skills decay over time, so it is likely that a single workshop is not sufficient to teach Bayesian reasoning. Next steps include integrating Bayesian reasoning into our local residency and medical school curricula in a more longitudinal fashion by developing additional cases and giving learners more opportunities to practice with Bayesian concepts at regularly scheduled small-group learning sessions. Future outcome measures could include learner comfort with Bayesian reasoning and diagnostic accuracy with designed clinical cases such as measured by Vermeersch et al. [11] Outcomes higher on Kirkpatrick's hierarchy that could also be measured include clinician behaviors such as rate of app usage in true clinical care, test ordering rates, and resident teaching behaviors.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.

Employment or leadership: None declared.

Honorarium: None declared.

Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

References

1. Makary MA, Daniel M. Medical error – the third leading cause of death in the US. *Br Med J* 2016;353.
2. Cutrer WB, Sullivan WM, Fleming AE. Educational strategies for improving clinical reasoning. *Curr Probl Pediatr Adolesc Health Care* 2013;43:248–57.
3. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. *N Engl J Med* 2006;355:2217–25.
4. Graber ML, Kissam S, Payne VL, Meyer AN, Sorensen A, Lenfestey N, et al. Cognitive interventions to reduce diagnostic error: a narrative review. *BMJ Qual Saf* 2012;21:535–57.
5. Thammasitboon S, Cutrer WB. Diagnostic decision-making and strategies to improve diagnosis. *Curr Probl Pediatr Adolesc Health Care* 2013;43:232–41.
6. Brase GL, Hill WT. Good fences make for good neighbors but bad science: a review of what improves Bayesian reasoning and why. *Front Psychol* 2015;6:340.
7. Whiting PF, Davenport C, Jameson C, Burke M, Sterne JA, Hyde C, et al. How well do health professionals interpret diagnostic information? A systematic review. *BMJ Open* 2015;5:e008155.
8. Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy 1. *Am J Med* 1998;104:374–80.

9. Van den Ende J, Moreira J, Basinga P, Bisoffi Z. The trouble with likelihood ratios. *Lancet* 2005;366:548.
10. Gavali MY, Khismatrao DS, Gavali YV, Patil KB. Smartphone, the new learning aid amongst medical students. *J Clin Diagn Res* 2017;11:JC05–8.
11. Vermeersch P, Bossuyt X. Comparative analysis of different approaches to report diagnostic accuracy. *Arch Gen Intern Med* 2010;170:734–5.
12. Johnson KM. Using Bayes' rule in diagnostic testing: a graphical explanation. *Diagnosis (Berl)*. 2017;4:159–67.
13. Kolb D. *Experiential education: experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice Hall, 1984.
14. Fagan T. Nomogram for Bayes's theorem. *N Engl J Med* 1975;293:257.
15. Kirkpatrick DL. *Evaluating training programs: the four levels*. San Francisco, CA: Berrett-Koehler, 1994.